

Calculus, Applications and Theory

Kenneth Kuttler

April 14, 2010

Contents

1	Introduction	13
I	Functions Of One Variable	15
1	A Short Review Of Precalculus	17
1.0.1	Outcomes	17
1.1	Set Notation	17
1.2	Completeness Of The Real Numbers	18
1.3	A Few Algebraic Conventions And Techniques	19
1.4	The Circular Arc Subtended By An Angle	20
1.5	The Trigonometric Functions	24
1.6	Exercises	27
1.7	Parabolas, Ellipses, and Hyperbolas	30
1.7.1	The Parabola	30
1.7.2	The Ellipse	32
1.7.3	The Hyperbola	34
1.8	Exercises	36
1.9	The Complex Numbers	37
1.10	Exercises	41
2	Functions	43
2.0.1	Outcomes	43
2.1	Functions And Sequences	43
2.2	Exercises	48
2.3	Continuous Functions	49
2.4	Sufficient Conditions For Continuity	53
2.5	Continuity Of Circular Functions	54
2.6	Exercises	55
2.7	Properties Of Continuous Functions	56
2.8	Exercises	57
2.9	Limit Of A Function	58
2.10	Exercises	62
2.11	The Limit Of A Sequence	63
2.11.1	Sequences And Completeness	66
2.11.2	Decimals	67
2.11.3	Continuity And The Limit Of A Sequence	68
2.12	Exercises	68
2.13	Uniform Continuity*	69
2.14	Exercises	70
2.15	Fundamental Theory*	71

2.15.1	Combinations Of Functions And Sequences	71
2.15.2	The Intermediate Value Theorem	75
2.15.3	Extreme Value Theorem	77
2.15.4	Different Versions Of Completeness	79
2.15.5	Sequential Compactness Of Closed Intervals	80
2.16	Exercises	81
3	Derivatives	85
3.0.1	Outcomes	85
3.1	Velocity	85
3.2	The Derivative	86
3.3	Exercises	91
3.4	Local Extrema	92
3.5	Exercises	95
3.6	Mean Value Theorem	97
3.7	Exercises	99
3.8	Curve Sketching	100
3.9	Exercises	102
4	Some Important Special Functions	105
4.0.1	Outcomes	105
4.1	The Circular Functions	105
4.2	Exercises	108
4.3	The Exponential And Log Functions	108
4.3.1	The Rules Of Exponents	108
4.3.2	The Exponential Functions, A Wild Assumption	109
4.3.3	The Special Number, e	112
4.3.4	The Function $\ln x $	112
4.3.5	Logarithm Functions	112
4.4	Exercises	114
5	Properties And Applications Of Derivatives	117
5.0.1	Outcomes	117
5.1	The Chain Rule And Derivatives Of Inverse Functions	117
5.1.1	The Chain Rule	117
5.1.2	Implicit Differentiation	118
5.2	Exercises	121
5.3	The Function x^r For r A Real Number	121
5.3.1	Logarithmic Differentiation	122
5.4	Exercises	123
5.5	The Inverse Trigonometric Functions	125
5.6	The Hyperbolic And Inverse Hyperbolic Functions	127
5.7	Exercises	129
5.8	L'Hôpital's Rule	130
5.8.1	Interest Compounded Continuously	134
5.9	Exercises	135
5.10	Related Rates	136
5.11	Exercises	137
5.12	The Derivative And Optimization	139
5.13	Exercises	142
5.14	The Newton Raphson Method	144
5.15	Exercises	146

6	Antiderivatives	147
6.0.1	Outcomes	147
6.1	Initial Value Problems	147
6.2	The Method Of Substitution	150
6.3	Exercises	152
6.4	Integration By Parts	154
6.5	Exercises	155
6.6	Trig. Substitutions	156
6.7	Exercises	161
6.8	Partial Fractions	161
6.9	Rational Functions Of Trig. Functions	168
6.10	Exercises	169
6.11	Practice Problems For Antiderivatives	170
6.12	Computers And Antiderivatives	177
7	Applications Of Antiderivatives	179
7.1	Areas	179
7.2	Area Between Graphs	180
7.3	Exercises	184
7.4	Volumes	186
7.4.1	Volumes Using Cross Sections	186
7.4.2	Volumes Using Shells	189
7.5	Exercises	192
7.6	Lengths Of Curves And Areas Of Surfaces Of Revolution	194
7.6.1	Lengths	194
7.6.2	Surfaces Of Revolution	196
7.7	Exercises	198
7.8	Force On A Dam And Work	199
7.8.1	Force On A Dam	199
7.8.2	Work	200
7.9	Exercises	201
8	Other Differential Equations*	203
8.1	The Equation $y' + a(t)y = b(t)$	203
8.2	Separable Differential Equations	205
8.3	Exercises	207
8.4	The Equations Of Undamped And Damped Oscillation	209
8.5	Exercises	212
9	The Integral	215
9.0.1	Outcomes	215
9.1	Upper And Lower Sums	216
9.2	Exercises	219
9.3	Functions Of Riemann Integrable Functions	220
9.4	Properties Of The Integral	222
9.5	Fundamental Theorem Of Calculus	226
9.6	The Riemann Integral	231
9.7	Exercises	231
9.8	Return Of The Wild Assumption	234
9.9	Exercises	237
9.10	Techniques Of Integration	238
9.10.1	The Method Of Substitution	238
9.10.2	Integration By Parts	240

9.11 Exercises	241
9.12 Improper Integrals	244
9.13 Exercises	249
10 Infinite Series	253
10.0.1 Outcomes	253
10.1 Approximation By Taylor Polynomials	253
10.2 Exercises	255
10.3 Infinite Series Of Numbers	257
10.3.1 Basic Considerations	257
10.4 Exercises	263
10.5 More Tests For Convergence	265
10.5.1 Convergence Because Of Cancellation	265
10.5.2 Ratio And Root Tests	266
10.6 Double Series*	268
10.7 Exercises	273
10.8 Power Series	275
10.8.1 Functions Defined In Terms Of Series	275
10.8.2 Operations On Power Series	276
10.9 Exercises	283
10.10 Some Other Theorems	287
 II Basic Linear Algebra	 291
11 Fundamentals	293
11.0.1 Outcomes	293
11.1 $\mathbb{R}^n$	293
11.2 Algebra in $\mathbb{R}^n$	295
11.3 Geometric Meaning Of Vector Addition In $\mathbb{R}^3$	296
11.4 Lines	297
11.5 Distance in $\mathbb{R}^n$	300
11.6 Geometric Meaning Of Scalar Multiplication In $\mathbb{R}^3$	303
11.7 Exercises	303
12 Matrices And Linear Transformations	307
12.0.1 Outcomes	307
12.1 Matrix Arithmetic	307
12.1.1 Addition And Scalar Multiplication Of Matrices	307
12.1.2 Multiplication Of Matrices	310
12.1.3 The ij^{th} Entry Of A Product	312
12.1.4 Properties Of Matrix Multiplication	314
12.1.5 The Transpose	315
12.1.6 The Identity And Inverses	317
12.2 Linear Transformations	319
12.3 Constructing The Matrix Of A Linear Transformation	320
12.4 Exercises	322
13 Determinants	325
13.0.1 Outcomes	325
13.1 Basic Techniques And Properties	325
13.1.1 Cofactors And 2×2 Determinants	325
13.1.2 The Determinant Of A Triangular Matrix	328

13.1.3 Properties Of Determinants	330
13.1.4 Finding Determinants Using Row Operations	331
13.2 Applications	333
13.2.1 A Formula For The Inverse	333
13.2.2 Cramer's Rule	336
13.3 Exercises	338
14 Spectral Theory	345
14.0.1 Outcomes	345
14.1 Eigenvalues And Eigenvectors Of A Matrix	345
14.1.1 Definition Of Eigenvectors And Eigenvalues	345
14.1.2 Finding Eigenvectors And Eigenvalues	347
14.1.3 A Warning	349
14.1.4 Complex Eigenvalues	351
14.2 Exercises	352
III Vectors In $\mathbb{R}^n$	355
15 Vectors And Points In $\mathbb{R}^n$	357
15.0.1 Outcomes	357
15.1 Open And Closed Sets	357
15.2 Exercises	360
15.3 Physical Vectors	361
15.4 Exercises	365
16 Vector Products	367
16.0.1 Outcomes	367
16.1 The Dot Product	368
16.2 The Geometric Significance Of The Dot Product	370
16.2.1 The Angle Between Two Vectors	370
16.2.2 Work And Projections	372
16.2.3 The Parabolic Mirror, An Application	374
16.2.4 The Dot Product And Distance In $\mathbb{C}^n$	377
16.3 Exercises	379
16.4 The Cross Product	380
16.4.1 The Distributive Law For The Cross Product	384
16.4.2 Torque	385
16.4.3 Center Of Mass	386
16.4.4 Angular Velocity	387
16.4.5 The Box Product	388
16.5 Vector Identities And Notation	390
16.6 Volumes	393
16.7 Exercises	395
16.8 Planes	397
16.9 Quadric Surfaces	401
16.10 Exercises	404
IV Vector Calculus	407
17 Vector Valued Functions	409
17.0.1 Outcomes	409

17.1	Vector Valued Functions	409
17.2	Vector Fields	410
17.3	Exercises	412
17.4	Continuous Functions	413
17.4.1	Sufficient Conditions For Continuity	413
17.5	Limits Of A Function	414
17.6	Properties Of Continuous Functions	418
17.7	Exercises	418
17.8	Some Fundamentals*	420
17.8.1	The Nested Interval Lemma	423
17.8.2	The Extreme Value Theorem	424
17.8.3	Sequences And Completeness	425
17.8.4	Continuity And The Limit Of A Sequence	428
17.9	Exercises	429
18	Vector Valued Functions Of One Variable	431
18.0.1	Outcomes	431
18.1	Limits Of A Vector Valued Function Of One Variable	431
18.2	The Derivative And Integral	433
18.2.1	Geometric And Physical Significance Of The Derivative	434
18.2.2	Differentiation Rules	436
18.2.3	Leibniz's Notation	438
18.3	Exercises	438
18.4	Line Integrals	439
18.4.1	Arc Length And Orientations	439
18.4.2	Line Integrals And Work	442
18.4.3	Another Notation For Line Integrals	445
18.5	Exercises	445
18.6	Independence Of Parametrization*	447
18.6.1	Hard Calculus	447
18.6.2	Independence Of Parametrization	451
19	Motion On A Space Curve	453
19.0.3	Outcomes	453
19.1	Space Curves	453
19.1.1	Some Simple Techniques	456
19.2	Geometry Of Space Curves*	458
19.3	Exercises	461
20	Some Curvilinear Coordinate Systems	463
20.0.1	Outcomes	463
20.1	Polar Coordinates	463
20.1.1	Graphs In Polar Coordinates	464
20.2	The Area In Polar Coordinates	466
20.3	Exercises	467
20.4	The Acceleration In Polar Coordinates	468
20.5	Planetary Motion	470
20.5.1	The Equal Area Rule	470
20.5.2	Inverse Square Law Motion, Kepler's First Law	471
20.5.3	Kepler's Third Law	473
20.6	Exercises	474
20.7	Spherical And Cylindrical Coordinates	475
20.8	Exercises	476

V	Vector Calculus In Many Variables	479
21	Functions Of Many Variables	481
21.0.1	Outcomes	481
21.1	The Graph Of A Function Of Two Variables	481
21.2	Review Of Limits	483
21.3	Exercises	485
21.4	The Directional Derivative And Partial Derivatives	485
21.4.1	The Directional Derivative	485
21.4.2	Partial Derivatives	487
21.5	Exercises	489
21.6	Mixed Partial Derivatives	490
21.7	Partial Differential Equations	492
21.8	Exercises	492
22	The Derivative Of A Function Of Many Variables	495
22.0.1	Outcomes	495
22.1	The Derivative Of Functions Of One Variable, $\mathbf{o}(\mathbf{v})$	495
22.2	Exercises	497
22.3	The Derivative Of Functions Of Many Variables	497
22.4	Exercises	501
22.5	C^1 Functions	502
22.6	The Chain Rule	505
22.6.1	The Chain Rule For Functions Of One Variable	505
22.6.2	The Chain Rule For Functions Of Many Variables	506
22.7	Exercises	510
22.7.1	Related Rates Problems	511
22.7.2	The Derivative Of The Inverse Function	513
22.7.3	Proof Of The Chain Rule	513
22.8	Exercises	514
22.9	The Gradient	516
22.10	The Gradient And Tangent Planes	518
22.11	Exercises	520
23	Optimization	523
23.0.1	Outcomes	523
23.1	Local Extrema	523
23.2	Exercises	525
23.3	The Second Derivative Test	527
23.4	Exercises	530
23.5	Lagrange Multipliers	532
23.6	Exercises	537
23.7	Proof Of The Second Derivative Test*	539
24	The Riemann Integral On $\mathbb{R}^n$	541
24.0.1	Outcomes	541
24.1	Methods For Double Integrals	541
24.1.1	Density And Mass	548
24.2	Exercises	548
24.3	Methods For Triple Integrals	550
24.3.1	Definition Of The Integral	550
24.3.2	Iterated Integrals	552
24.4	Exercises	555

24.4.1	Mass And Density	556
24.5	Exercises	558
25	The Integral In Other Coordinates	559
25.0.1	Outcomes	559
25.1	Polar Coordinates	559
25.2	Exercises	561
25.3	Cylindrical And Spherical Coordinates	562
25.4	Exercises	570
25.5	The General Procedure	570
25.6	Exercises	573
25.7	The Moment Of Inertia And Center Of Mass	574
25.8	Exercises	576
26	The Integral On Two Dimensional Surfaces In $\mathbb{R}^3$	579
26.0.1	Outcomes	579
26.1	The Two Dimensional Area In $\mathbb{R}^3$	579
26.1.1	Surfaces Of The Form $z = f(x, y)$	583
26.2	Exercises	585
27	Calculus Of Vector Fields	587
27.0.1	Outcomes	587
27.1	Divergence And Curl Of A Vector Field	587
27.1.1	Vector Identities	588
27.1.2	Vector Potentials	590
27.1.3	The Weak Maximum Principle	590
27.2	Exercises	591
27.3	The Divergence Theorem	592
27.3.1	Coordinate Free Concept Of Divergence	596
27.4	Some Applications Of The Divergence Theorem	597
27.4.1	Hydrostatic Pressure	597
27.4.2	Archimedes Law Of Buoyancy	598
27.4.3	Equations Of Heat And Diffusion	598
27.4.4	Balance Of Mass	599
27.4.5	Balance Of Momentum	600
27.4.6	Frame Indifference	605
27.4.7	Bernoulli's Principle	606
27.4.8	The Wave Equation	607
27.4.9	A Negative Observation	607
27.4.10	Volumes Of Balls In $\mathbb{R}^n$	608
27.4.11	Electrostatics	610
27.5	Exercises	611
28	Stokes And Green's Theorems	615
28.0.1	Outcomes	615
28.1	Green's Theorem	615
28.2	Stoke's Theorem From Green's Theorem	620
28.2.1	The Normal And The Orientation	622
28.2.2	The Mobeus Band	624
28.2.3	Conservative Vector Fields	625
28.2.4	Some Terminology	628
28.3	Maxwell's Equations And The Wave Equation	629
28.4	Exercises	630

A Basic Plane Geometry	635
A.1 Similar Triangles And Parallel Lines	635
A.2 Distance Formula And Trigonometric Functions	637
B The Fundamental Theorem Of Algebra	641
C The Mathematical Theory Of Determinants*	645
C.1 The Function sgn_n	645
C.2 The Determinant	647
C.2.1 The Definition	647
C.2.2 Permuting Rows Or Columns	648
C.2.3 A Symmetric Definition	649
C.2.4 The Alternating Property Of The Determinant	649
C.2.5 Linear Combinations And Determinants	650
C.2.6 The Determinant Of A Product	650
C.2.7 Cofactor Expansions	651
C.2.8 Formula For The Inverse	653
C.2.9 Cramer's Rule	654
C.2.10 Upper Triangular Matrices	654
C.2.11 The Determinant Rank	655
C.2.12 Telling Whether A Is One To One Or Onto	656
C.2.13 Schur's Theorem	657
C.2.14 Symmetric Matrices	659
C.3 Exercises	660
D Implicit Function Theorem*	663
D.2 More Continuous Partial Derivatives	666
D.3 The Method Of Lagrange Multipliers	668
D.4 The Local Structure Of C^1 Mappings	669
E The Theory Of The Riemann Integral*	673
E.1 Basic Properties	676
E.2 Which Functions Are Integrable?	678
E.3 Iterated Integrals	686
E.4 The Change Of Variables Formula	690
E.5 Some Observations	696
F Volumes In Higher Dimensions*	699
F.1 p Dimensional Surfaces*	699
F.2 Spherical Coordinates In Many Dimensions*	703
G Some Physical Applications	707
G.1 Product Rule For Matrices	707
G.2 Moving Coordinate Systems	707
G.3 Newton's Laws Of Motion	710
G.4 Kinetic Energy	714
G.5 Kepler's First And Second Laws	715
G.6 Impulse And Momentum	717
G.6.1 The Spinning Top	718
G.6.2 Kinetic Energy	721
G.7 Acceleration With Respect To Moving Coordinate Systems*	722
G.7.1 The Coriolis Acceleration	722
G.7.2 The Coriolis Acceleration On The Rotating Earth	724

G.8 Exercises	729
G.9 Exercises With Answers	732
H Curvilinear Coordinates	733
H.1 Basis vectors	733
H.2 Exercises	736
H.3 Curvilinear Coordinates	737
H.4 Exercises	739
H.5 Transformation Of Coordinates.	741
H.6 Differentiation and Christoffel Symbols	742
H.7 Gradients And Divergence	744
H.8 Exercises	747
H.9 Curl And Cross Products	748
Copyright © 2004,	

Introduction

Calculus consists of the study of limits of various sorts and the systematic exploitation of the completeness axiom. It was developed over a period of several hundred years in order to solve problems from the physical sciences. It is the language by which precision and quantitative predictions for many complicated problems are obtained. It also has significant applications in pure mathematics. For example, it is used to define and find lengths of curves, areas and volumes of three dimensional shapes. It is essential in order to solve many maximization problems and it is prerequisite material in order to understand models based on differential equations. These and other applications are discussed to some extent in this book.

It is assumed the reader has a good understanding of algebra on the level of college algebra or what used to be called algebra II along with some exposure to geometry and trigonometry although the book does contain a review of these things.

If the optional sections and non standard sections are not included, this book is fairly short. However, there is a lot of non standard material, including the big theorems of advanced calculus.

I have also tried to give complete proofs of all theorems in one variable calculus and to at least give plausibility arguments for those in multiple dimensions with proofs given in appendices or optional sections. I have done this because I am sick and tired of books which do not bother to present proofs of the theorems stated and either pretend there is nothing to prove, state that the proof is “beyond the scope of this book” say they will omit the proof, or worse yet give a specious explanation. For a serious student, mathematics is not about accepting on faith unproved assertions presumably understood by someone else but “beyond the scope of this book”. Nevertheless, it has become fashionable to care nothing about such serious students and to write books for the convenience of those who care nothing for explanations, those people who are forced to take the course to get general education credit for example. It is my intent that this should not be one of those books.

I expect the reader to be able to use a calculator whenever it would be helpful to do so. The use of a calculator will sometimes be very helpful and you should be ready to use technology when it is appropriate. In addition, a minimal introduction to the use of computer algebra systems is presented. Having said this, calculus is not about using calculators or any other form of technology. Weierstrass could do calculus quite well without the benefit of modern gadgets. I believe that when the syntax and arcane notation associated with technology are presented too prominently, these things become the topic of study rather than the concepts of calculus and this is a real shame. This is a book on calculus.

Pictures are often helpful in seeing what is going on, and there are many pictures in this book for this reason. However, calculus is not art and ultimately rests on logic and definitions, like the rest of mathematics. Algebra plays a central role in gaining the sort of understanding which generalizes to higher dimensions where pictures are not

available. Therefore, I have emphasized the algebraic aspects of this subject far more than is usual, especially those parts of linear algebra which are essential to understand in order to do multi-variable calculus.

A star next to an exercise indicates it is either technically difficult or perhaps a little bit challenging to figure out. Don't be frightened by these exercises. They just might take a little longer to work but some are very worth while and ought to be attempted.

Part I

Functions Of One Variable

A Short Review Of Precalculus

1.0.1 Outcomes

1. Understand set notation.
2. Understand the radian measure of an angle.
3. Understand the definition of completeness of $\mathbb{R}$
4. Understand sup and inf
5. Understand and review basic geometry and the trig. functions and their properties.
6. Understand the basic facts about conic sections.
7. Understand the geometric significance of a complex number as a point in the plane.
8. Understand the algebra of complex numbers.
9. Understand the absolute value of a complex number and how to find it as well as its geometric significance.
10. Be able to do computations with complex numbers and understand the conjugate geometrically and algebraically.
11. Understand De'Moivre's theorem and be able to use it to find the roots of a complex number.

1.1 Set Notation

A set is just a collection of things called elements. Often these are also referred to as points in calculus. For example $\{1, 2, 3, 8\}$ would be a set consisting of the elements 1, 2, 3, and 8. To indicate that 3 is an element of $\{1, 2, 3, 8\}$, it is customary to write $3 \in \{1, 2, 3, 8\}$. $9 \notin \{1, 2, 3, 8\}$ means 9 is not an element of $\{1, 2, 3, 8\}$. Sometimes a rule specifies a set. For example you could specify a set as all integers larger than 2. This would be written as $S = \{x \in \mathbb{Z} : x > 2\}$. This notation says: the set of all integers, x , such that $x > 2$.

If A and B are sets with the property that every element of A is an element of B , then A is a subset of B . For example, $\{1, 2, 3, 8\}$ is a subset of $\{1, 2, 3, 4, 5, 8\}$, in

symbols, $\{1, 2, 3, 8\} \subseteq \{1, 2, 3, 4, 5, 8\}$. The same statement about the two sets may also be written as $\{1, 2, 3, 4, 5, 8\} \supseteq \{1, 2, 3, 8\}$.

The union of two sets is the set consisting of everything which is contained in at least one of the sets, A or B . As an example of the union of two sets, $\{1, 2, 3, 8\} \cup \{3, 4, 7, 8\} = \{1, 2, 3, 4, 7, 8\}$ because these numbers are those which are in at least one of the two sets. In general

$$A \cup B \equiv \{x : x \in A \text{ or } x \in B\}.$$

Be sure you understand that something which is in both A and B is in the union. It is not an exclusive or.

The intersection of two sets, A and B consists of everything which is in both of the sets. Thus $\{1, 2, 3, 8\} \cap \{3, 4, 7, 8\} = \{3, 8\}$ because 3 and 8 are those elements the two sets have in common. In general,

$$A \cap B \equiv \{x : x \in A \text{ and } x \in B\}.$$

When with real numbers, $[a, b]$ denotes the set of real numbers, x , such that $a \leq x \leq b$ and $[a, b)$ denotes the set of real numbers such that $a \leq x < b$. (a, b) consists of the set of real numbers, x such that $a < x < b$ and $(a, b]$ indicates the set of numbers, x such that $a < x \leq b$. $[a, \infty)$ means the set of all numbers, x such that $x \geq a$ and $(-\infty, a]$ means the set of all real numbers which are less than or equal to a . These sorts of sets of real numbers are called intervals. The two points, a and b are called endpoints of the interval. Other intervals such as $(-\infty, b)$ are defined by analogy to what was just explained. In general, the curved parenthesis indicates the end point it sits next to is not included while the square parenthesis indicates this end point is included. The reason that there will always be a curved parenthesis next to ∞ or $-\infty$ is that these are not real numbers. Therefore, they cannot be included in any set of real numbers.

A special set which needs to be given a name is the empty set also called the null set, denoted by $\emptyset$. Thus $\emptyset$ is defined as the set which has no elements in it. Mathematicians like to say the empty set is a subset of every set. The reason they say this is that if it were not so, there would have to exist a set, A , such that $\emptyset$ has something in it which is not in A . However, $\emptyset$ has nothing in it and so the least intellectual discomfort is achieved by saying $\emptyset \subseteq A$.

If A and B are two sets, $A \setminus B$ denotes the set of things which are in A but not in B . Thus

$$A \setminus B \equiv \{x \in A : x \notin B\}.$$

Set notation is used whenever convenient.

1.2 Completeness Of The Real Numbers

I assume the reader is familiar with the usual algebraic properties of the real numbers. However, they have another property known as completeness.

Definition 1.2.1 *A non empty set, $S \subseteq \mathbb{R}$ is bounded above (below) if there exists $x \in \mathbb{R}$ such that $x \geq (\leq) s$ for all $s \in S$. If S is a nonempty set in $\mathbb{R}$ which is bounded above, then a number, l which has the property that l is an upper bound and that every other upper bound is no smaller than l is called a least upper bound, l.u.b. (S) or often $\sup(S)$. If S is a nonempty set bounded below, define the greatest lower bound, g.l.b. (S) or $\inf(S)$ similarly. Thus g is the g.l.b. (S) means g is a lower bound for S and it is the largest of all lower bounds. If S is a nonempty subset of $\mathbb{R}$ which is not bounded above, this information is expressed by saying $\sup(S) = +\infty$ and if S is not bounded below, $\inf(S) = -\infty$.*

Every existence theorem in calculus depends on some form of the completeness axiom.

Axiom 1.2.2 (completeness) *Every nonempty set of real numbers which is bounded above has a least upper bound and every nonempty set of real numbers which is bounded below has a greatest lower bound.*

It is this axiom which distinguishes Calculus from Algebra. The notation $[a, b]$ means the set of all x such that $a \leq x \leq b$. The notation (a, b) means the set of all x such that $a < x < b$. Obvious modifications hold for $[a, b)$ and $(a, b]$.

A fundamental result about sup and inf is the following.

Proposition 1.2.3 *Let S be a nonempty set and suppose $\sup(S)$ exists. Then for every $\delta > 0$,*

$$S \cap (\sup(S) - \delta, \sup(S)] \neq \emptyset.$$

If $\inf(S)$ exists, then for every $\delta > 0$,

$$S \cap [\inf(S), \inf(S) + \delta) \neq \emptyset.$$

Proof: Consider the first claim. If the indicated set equals $\emptyset$, then $\sup(S) - \delta$ is an upper bound for S which is smaller than $\sup(S)$, contrary to the definition of $\sup(S)$ as the least upper bound. In the second claim, if the indicated set equals $\emptyset$, then $\inf(S) + \delta$ would be a lower bound which is larger than $\inf(S)$ contrary to the definition of $\inf(S)$.

1.3 A Few Algebraic Conventions And Techniques

Summation notation is a convenient way to specify a sum. This notation is defined in the following definition.

Definition 1.3.1 *For $i = m, \dots, n$ let a_i be specified. Then*

$$\sum_{i=m}^n a_i \equiv a_m + a_{m+1} + \dots + a_n$$

I will use this whenever convenient.

Example 1.3.2 *Find $\sum_{i=1}^3 2i - 1$.*

From the definition, it equals $(2 - 1) + (2 \times 2 - 1) + (2 \times 3 - 1) = 9$.

An important technique is the technique of proof by induction. I will illustrate with a simple example which is useful for its own sake.

Example 1.3.3 *For $n = 1, 2, 3, \dots$ and $\alpha > 0$ it is always the case that*

$$(1 + \alpha)^n \geq 1 + n\alpha + \frac{n(n-1)}{2}\alpha^2$$

Here is why. The statement is true if $n = 1$. Now suppose I can show that whenever the statement is true for some value of n it follows that it must be true for the next value of n . Then it must be the case that it is true for each value of n . Consider why this is. Since it is true for $n = 1$ and whenever it is true for some n it is true for $n + 1$ it follows that it must be true for 2. Since it is true for 2, it must be true for 3 and so forth. Thus it suffices to show that **if** it is true for n **then** it is true for $n + 1$. I haven't done this yet but I am about to do it.

Suppose then that the inequality is true for n . I need to verify it holds for $n + 1$. That is, the same formula needs to hold with n replaced everywhere with $n + 1$. Then using the assumption that it is true for n ,

$$\begin{aligned} (1 + \alpha)^{n+1} &= (1 + \alpha)(1 + \alpha)^n \\ &\geq (1 + \alpha) \left(1 + n\alpha + \frac{n(n-1)}{2}\alpha^2 \right) \\ &= 1 + n\alpha + \frac{n(n-1)}{2}\alpha^2 + \alpha + n\alpha^2 + \frac{n(n-1)}{2}\alpha^3 \\ &\geq 1 + (n+1)\alpha + \frac{n(n-1)}{2}\alpha^2 + n\alpha^2 \end{aligned}$$

where I simply threw out the last term in going to the last line. This equals

$$\begin{aligned} &1 + (n+1)\alpha + \frac{n(n-1) + 2n}{2}\alpha^2 \\ &= 1 + (n+1)\alpha + \frac{(n+1)n}{2}\alpha^2 \end{aligned}$$

Thus the inequality holds with n replaced with $n + 1$. This proves the desired inequality.

There are many other examples where induction is useful.

1.4 The Circular Arc Subtended By An Angle

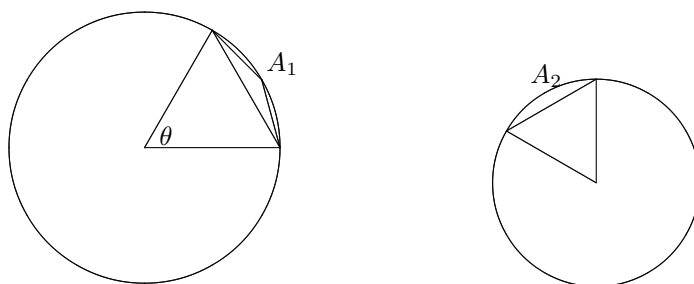
How can angles be measured? This will be done by considering arcs on a circle. To see how this will be done, let θ denote an angle and place the vertex of this angle at the center of the circle. Next, extend its two sides till they intersect the circle. Note the angle could be opening in any of infinitely many different directions. Thus this procedure could yield any of infinitely many different circular arcs. Each of these arcs is said to subtend the angle. In fact each of these arcs has the same length. When this has been shown, it will be easy to measure angles. Angles will be measured in terms of lengths of arcs subtended by the angle. Of course it is also necessary to define what is meant by the length of a circular arc in order to do any of this. First I will describe an intuitive way of thinking about this and then give a rigorous definition and proof. If the intuitive way of thinking about this satisfies you, no harm will be done by skipping the more technical discussion which follows.

Take an angle and place its vertex (the point) at the center of a circle of radius r . Then, extending the sides of the angle if necessary till they intersect the circle, this determines an arc on the circle. If r were changed to R , this really amounts to a change of units of length. Think, for example, of keeping the numbers the same but changing centimeters to meters in order to produce an enlarged version of the same picture. Thus the picture looks exactly the same, only larger. It is reasonable to suppose, based on this reasoning, that the way to measure the angle is to take the length of the arc subtended in whatever units being used and divide this length by the radius measured in the same units, thus obtaining a number which is independent of the units of length used, just as the angle itself is independent of units of length. After all, it is the same angle regardless of how far its sides are extended. This is in fact how to define the radian measure of an angle and the definition is well defined. Thus, in particular, the ratio between the circumference (length) of a circle and its radius is a constant which is independent of

the radius of the circle¹. Since the time of Euler in the 1700's, this constant has been denoted by 2π . In summary, if θ is the radian measure of an angle, the length of the arc subtended by the angle on a circle of radius r is $r\theta$.

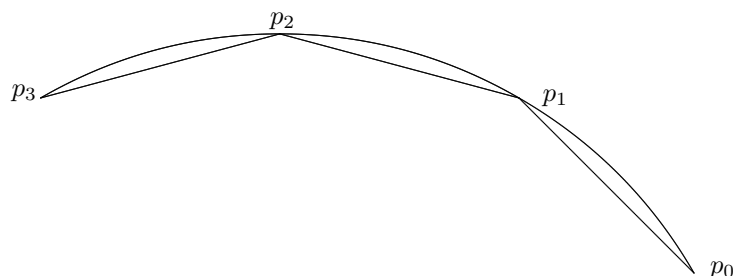
This is a little sloppy right now because no precise definition of the length of an arc of a circle has been given. For now, imagine taking a string, placing one end of it on one end of the circular arc and then wrapping the string till you reach the other end of the arc. Stretching this string out and measuring it would then give you the length of the arc. Such intuitive discussions involving string may or may not be enough to convey understanding. If you need to see more discussion, read on. Otherwise, skip to the next section.

To give a precise description of what is meant by the length of an arc, consider the following picture.



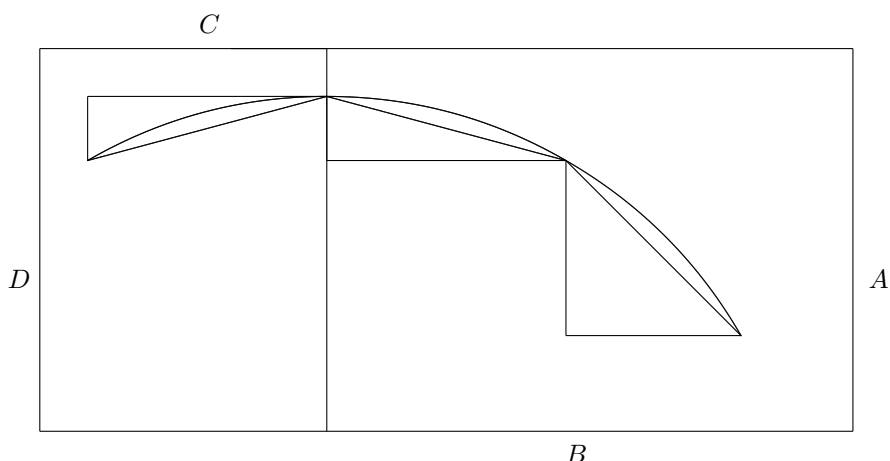
In this picture, there are two circles, a big one having radius, R and a little one having radius r . The angle, θ is situated in two different ways subtending the arcs A_1 and A_2 as shown.

Letting A be an arc of a circle, like those shown in the above picture, A subset of A , $\{p_0, \dots, p_n\}$ is a partition of A if p_0 is one endpoint, p_n is the other end point, and the points are encountered in the indicated order as one moves in the counter clockwise direction along the arc. To illustrate, see the following picture.



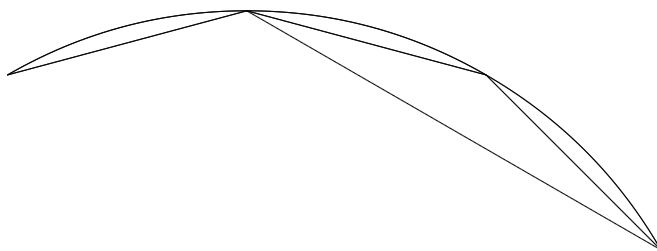
Also, denote by $\mathcal{P}(A)$ the set of all such partitions. For $P = \{p_0, \dots, p_n\}$, denote by $|p_i - p_{i-1}|$ the distance between p_i and p_{i-1} . Then for $P \in \mathcal{P}(A)$, define $|P| \equiv \sum_{i=1}^n |p_i - p_{i-1}|$. Thus $|P|$ consists of the sum of the lengths of the little lines joining successive points of P and appears to be an approximation to the length of the circular arc, A . By geometry, the length of any of the straight line segments joining successive points in a partition is smaller than the sum of the two sides of a right triangle having the given straight line segment as its hypotenuse. For example, see the following picture.

¹In 2 Chronicles 4:2 the "molten sea" found in Solomon's temple is described. It sat on 12 oxen, was round, 5 cubits high, 10 across and 30 around. Thus its radius was 5 and the the Bible, taken literally, gives the value of π as 3. This is not too far off but is not correct. Other incorrect values of π can be found in the Indiana pi bill of 1897. Later, methods will be given which allow one to calculate π more precisely. A better value is 3.1415926535 and presently this number is known to thousands of decimal places. It was proved by Lindeman in 1882 that π is transcendental which is the worst sort of irrational number.

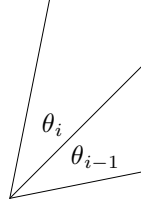


The sum of the lengths of the straight line segments in the part of the picture found in the right rectangle above is less than $A + B$ and the sum of the lengths of the straight line segments in the part of the picture found in the left rectangle above is less than $C + D$ and this would be so for any partition. Therefore, for any $P \in \mathcal{P}(A)$, $|P| \leq M$ where M is the perimeter of a rectangle containing the arc, A . To be a little sloppy, simply pick M to be the perimeter of a rectangle containing the whole circle of which A is a part. The only purpose for doing this is to obtain the existence of an upper bound. Therefore, $\{|P| : P \in \mathcal{P}(A)\}$ is a set of numbers which is bounded above by M and by completeness of $\mathbb{R}$ it is possible to define the length of A , $l(A)$, by $l(A) \equiv \sup \{|P| : P \in \mathcal{P}(A)\}$.

A fundamental observation following from the theorem in geometry which says the sum of the lengths of two sides of a triangle is always at least as large as the length of the third side, is that if $P, Q \in \mathcal{P}(A)$ and $P \subseteq Q$, then $|P| \leq |Q|$. To see this, add in one point at a time to P . This effect of adding in one point is illustrated in the following picture.



Also, letting $\{p_0, \dots, p_n\}$ be a partition of A , specify angles, θ_i as follows. The angle θ_i is formed by the two lines, one from the center of the circle to p_i and the other line from the center of the circle to p_{i-1} . Furthermore, a specification of these angles yields the partition of A in the following way. Place the vertex of θ_1 on the center of the circle, letting one side lie on the line from the center of the circle to p_0 and the other side extended resulting in a point further along the arc in the counter clockwise direction. When the angles, $\theta_1, \dots, \theta_{i-1}$ have produced points, $p_0, \dots, p_{i-1}$ on the arc, place the vertex of θ_i on the center of the circle and let one side of θ_i coincide with the side of the angle θ_{i-1} which is most counter clockwise, the other side of θ_i when extended, resulting in a point further along the arc, A in the counterclockwise direction as shown below.



Now let $\varepsilon > 0$ be given and pick $P_1 \in \mathcal{P}(A_1)$ such that $|P_1| + \varepsilon > l(A_1)$. Then determining the angles as just described, use these angles to produce a corresponding partition of A_2, P_2 . If $|P_2| + \varepsilon > l(A_2)$, then stop. Otherwise, pick $Q \in \mathcal{P}(A_2)$ such that $|Q| + \varepsilon > l(A_2)$ and let $P'_2 = P_2 \cup Q$. Then use the angles determined by P'_2 to obtain $P'_1 \in \mathcal{P}(A_1)$. Then $|P'_1| + \varepsilon > l(A_1)$, $|P'_2| + \varepsilon > l(A_2)$, and both P'_1 and P'_2 determine the same sequence of angles. Using plane geometry, the theorem which says that two isosceles triangles are similar if they have the same central angle,

$$\frac{|P'_1|}{|P'_2|} = \frac{R}{r}$$

and so

$$l(A_2) < |P'_2| + \varepsilon = \frac{r}{R} |P'_1| + \varepsilon \leq \frac{r}{R} l(A_1) + \varepsilon.$$

Since ε is arbitrary, this shows $Rl(A_2) \leq rl(A_1)$. But now reverse the argument and write

$$l(A_1) < |P'_1| + \varepsilon = \frac{R}{r} |P'_2| + \varepsilon \leq \frac{R}{r} l(A_2) + \varepsilon$$

which implies, since ε is arbitrary that $Rl(A_2) \geq rl(A_1)$ and this has proved the following theorem.

Theorem 1.4.1 *Let θ be an angle which subtends two arcs, A_R on a circle of radius R and A_r on a circle of radius r . Then denoting by $l(A)$ the length of a circular arc as described above, $Rl(A_r) = rl(A_R)$.*

Before proceeding further, note the proof of the above theorem involved showing $l(A_1) < \frac{R}{r}l(A_2) + \varepsilon$ where $\varepsilon > 0$ was arbitrary and from this, the conclusion that $l(A_1) \leq \frac{R}{r}l(A_2)$. This is a very typical way of showing one number is no larger than another. To show $a \leq b$ first show that for every $\varepsilon > 0$ it follows that $a < b + \varepsilon$. This implies $a - b < \varepsilon$ for all positive ε and so it must be the case that $a - b \leq 0$ since otherwise, you could take $\varepsilon = \frac{a-b}{2}$ and conclude $0 < a - b < \frac{a-b}{2}$, a contradiction.

With this preparation, here is the definition of the measure of an angle.

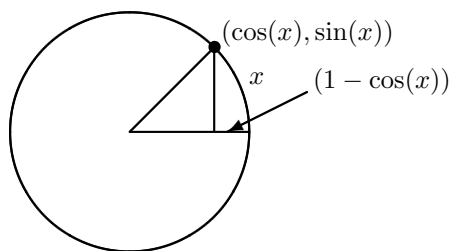
Definition 1.4.2 *Let θ be an angle. The measure of θ is defined to be the length of the circular arc subtended by θ on a circle of radius r divided by r . This is also called the radian measure of the angle.*

You should note the measure of θ is independent of dimension. This is because the units of length cancel when the division takes place.

Proposition 1.4.3 *The above definition is well defined and also, if A is an arc subtended by the angle θ on a circle of radius r then the length of A , denoted by $l(A)$ is given by $l(A) = r\theta$.*

Proof: That the definition is well defined follows from Theorem 1.4.1. The formula also follows from Theorem 1.4.1 and letting $R = 1$.

Now is a good time to present a useful inequality which may or may not be self evident. Here is a picture which illustrates the conclusion of this corollary.



The following corollary states that the length of the subtended arc shown in the picture is longer than the vertical side of the triangle and smaller than the sum of the vertical side with the segment having length $1 - \cos \theta$. To me, this seems abundantly clear but in case it is hard to believe, the following corollary gives a proof.

Corollary 1.4.4 *Let $0 < \text{radian measure of } \theta < \pi/4$. Then letting A be the arc on the unit circle resulting from situating the angle with one side on the positive x axis and the other side pointing up from the positive x axis,*

$$(1 - \cos \theta) + \sin \theta \geq l(A) \geq \sin \theta \quad (1.1)$$

Proof: Situate the angle, θ such that one side is on the positive x axis and extend the other side till it intersects the unit circle at the point, $(\cos \theta, \sin \theta)$. Then denoting the resulting arc on the circle by A , it follows that for all $P \in \mathcal{P}(A)$ the inequality $(1 - \cos \theta) + \sin \theta \geq |P| \geq \sin \theta$. It follows that $(1 - \cos \theta) + \sin \theta$ is an upper bound for all the $|P|$ where $P \in \mathcal{P}(A)$ and so $(1 - \cos \theta) + \sin \theta$ is at least as large as the sup or least upper bound of the $|P|$. This proves the top half of the inequality. The bottom half follows because $l(A) \geq L$ where L is the length of the line segment joining $(\cos \theta, \sin \theta)$ and $(1, 0)$ due to the definition of $l(A)$. However, $L \geq \sin \theta$ because L is the length of the hypotenuse of a right triangle having $\sin \theta$ as one of the sides.

1.5 The Trigonometric Functions

Now the Trigonometric functions will be defined as functions of an arbitrary real variable. Up till now they have been defined as functions of pointy things called angles. The following theorem will make possible the definition.

Theorem 1.5.1 *Let $b \in \mathbb{R}$. Then there exists a unique integer p and real number r such that $0 \leq r < 2\pi$ and $b = p2\pi + r$.*

Proof: Consider the equally spaced points $k(2\pi)$ where k is an integer. Now the disjoint half open intervals $[k(2\pi), (k+1)(2\pi))$ cover the whole real line and for different values of k these intervals are disjoint. Let p be the value of k such that $b \in [p(2\pi), (p+1)(2\pi))$. Then for some $r \in [0, 2\pi)$,

$$b = p2\pi + r.$$

This proves the theorem.

The following definition is for $\sin b$ and $\cos b$ for $b \in \mathbb{R}$.

Definition 1.5.2 *Let $b \in \mathbb{R}$. Then $\sin b \equiv \sin r$ and $\cos b \equiv \cos r$ where $b = 2\pi p + r$ for p an integer, and $r \in [0, 2\pi)$.*

Several observations are now obvious from this.

Observation 1.5.3 *Let $k \in \mathbb{Z}$, then the following formulas hold.*

$$\sin b = -\sin(-b), \cos b = \cos(-b), \quad (1.2)$$

$$\sin(b + 2k\pi) = \sin b, \cos(b + 2k\pi) = \cos b \quad (1.3)$$

$$\cos^2 b + \sin^2 b = 1 \quad (1.4)$$

The other trigonometric functions are defined in the usual way provided they make sense.

From the observation that the x and y axes intersect at right angles the four arcs on the unit circle subtended by these axes are all of equal length. Therefore, the measure of a right angle must be $2\pi/4 = \pi/2$. The measure of the angle which is determined by the arc from $(1, 0)$ to $(-1, 0)$ is seen to equal π by the same reasoning. From the definition of the trig functions, $\cos(\pi/2) = 0$ and $\sin(\pi/2) = 1$. You can easily find other values for $\cos$ and $\sin$ at all the other multiples of $\pi/2$.

The next topic is the important formulas for the trig. functions of sums and differences of numbers. For $b \in \mathbb{R}$, denote by r_b the element of $[0, 2\pi)$ having the property that $b = 2\pi p + r_b$ for p an integer.

Lemma 1.5.4 *Let $x, y \in \mathbb{R}$. Then $r_{x+y} = r_x + r_y + 2k\pi$ for some $k \in \mathbb{Z}$.*

Proof: By definition,

$$x + y = 2\pi p + r_{x+y}, \quad x = 2\pi p_1 + r_x, \quad y = 2\pi p_2 + r_y.$$

From this the result follows because

$$0 = ((x + y) - x) - y = 2\pi \overbrace{((p - p_1) - p_2)}^{\equiv -k} + r_{x+y} - (r_x + r_y).$$

Let $z \in \mathbb{R}$ and let $p(z)$ denote the point on the unit circle determined by the length r_z whose coordinates are $\cos z$ and $\sin z$. Thus, starting at $(1, 0)$ and moving counter clockwise a distance of r_z on the unit circle yields $p(z)$. Note also that $p(z) = p(r_z)$.

Lemma 1.5.5 *Let $x, y \in \mathbb{R}$. Then the length of the arc between $p(x + y)$ and $p(x)$ is equal to the length of the arc between $p(y)$ and $(1, 0)$.*

Proof: The length of the arc between $p(x + y)$ and $p(x)$ is $|r_{x+y} - r_x|$. There are two cases to consider here.

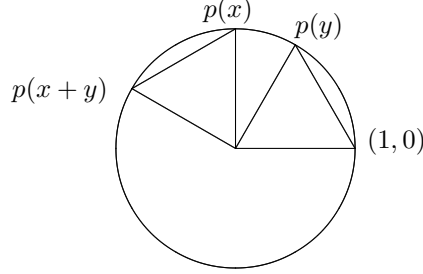
First assume $r_{x+y} \geq r_x$. Then $|r_{x+y} - r_x| = r_{x+y} - r_x = r_y + 2k\pi$. Since both r_{x+y} and r_x are in $[0, 2\pi)$, their difference is also in $[0, 2\pi)$ and so $k = 0$. Therefore, the arc joining $p(x)$ and $p(x + y)$ is of the same length as the arc joining $p(y)$ and $(1, 0)$. In the other case, $r_{x+y} < r_x$ and in this case $|r_{x+y} - r_x| = r_x - r_{x+y} = -r_y - 2k\pi$. Since r_x and r_{x+y} are both in $[0, 2\pi)$ their difference is also in $[0, 2\pi)$ and so in this case $k = -1$. Therefore, in this case, $|r_{x+y} - r_x| = 2\pi - r_y$. Now since the circumference of the unit circle is 2π , the length of the arc joining $p(2\pi - r_y)$ to $(1, 0)$ is the same as the length of the arc joining $p(r_y) = p(y)$ to $(1, 0)$. This proves the lemma.

The following theorem is the fundamental identity from which all the major trig. identities involving sums and differences of angles are derived.

Theorem 1.5.6 *Let $x, y \in \mathbb{R}$. Then*

$$\cos(x + y) \cos y + \sin(x + y) \sin y = \cos x. \quad (1.5)$$

Proof: Recall that for a real number, z , there is a unique point, $p(z)$ on the unit circle and the coordinates of this point are $\cos z$ and $\sin z$. Now from the above lemma, the length of the arc between $p(x+y)$ and $p(x)$ has the same length as the arc between $p(y)$ and $p(0)$. As an illustration see the following picture.



It follows from the definition of the radian measure of an angle that the two angles determined by these arcs are equal and so the distance between the points $p(x+y)$ and $p(x)$ must be the same as the distance from $p(y)$ to $p(0)$. Writing this in terms of the definition of the trig functions and the distance formula,

$$(\cos(x+y) - \cos x)^2 + (\sin(x+y) - \sin x)^2 = (\cos y - 1)^2 + \sin^2 x.$$

$$\begin{aligned} \cos^2(x+y) + \cos^2 x - 2\cos(x+y)\cos x + \sin^2(x+y) + \sin^2 x - 2\sin(x+y)\sin x \\ = \cos^2 y - 2\cos y + 1 + \sin^2 y \end{aligned}$$

From Observation 1.5.3 this implies 1.5. This proves the theorem.

Letting $y = \pi/2$, this shows that

$$\sin(x + \pi/2) = \cos x. \quad (1.6)$$

Now let $u = x + y$ and $v = y$. Then 1.5 implies

$$\cos u \cos v + \sin u \sin v = \cos(u - v) \quad (1.7)$$

Also, from this and 1.2,

$$\begin{aligned} \cos(u + v) &= \cos(u - (-v)) \\ &= \cos u \cos(-v) + \sin u \sin(-v) \\ &= \cos u \cos v - \sin u \sin v \end{aligned} \quad (1.8)$$

Thus, letting $v = \pi/2$,

$$\cos\left(u + \frac{\pi}{2}\right) = -\sin u. \quad (1.9)$$

It follows

$$\begin{aligned} \sin(x + y) &= -\cos\left(x + \frac{\pi}{2} + y\right) \\ &= -\left[\cos\left(x + \frac{\pi}{2}\right)\cos y - \sin\left(x + \frac{\pi}{2}\right)\sin y\right] \\ &= \sin x \cos y + \sin y \cos x \end{aligned} \quad (1.10)$$

Then using Observation 1.5.3 again, this implies

$$\sin(x - y) = \sin x \cos y - \cos x \sin y. \quad (1.11)$$

In addition to this, Observation 1.5.3 implies

$$\cos 2x = \cos^2 x - \sin^2 x \quad (1.12)$$

$$= 2 \cos^2 x - 1 \quad (1.13)$$

$$= 1 - 2 \sin^2 x \quad (1.14)$$

Therefore, making use of the above identities, and Observation 1.5.3,

$$\begin{aligned} \cos(3x) &= \cos 2x \cos x - \sin 2x \sin x \\ &= (2 \cos^2 x - 1) \cos x - 2 \cos x \sin^2 x \\ &= 4 \cos^3 x - 3 \cos x \end{aligned} \quad (1.15)$$

With these fundamental identities, it is easy to obtain the cosine and sine of many special angles, called reference angles. First, $\cos(\frac{\pi}{4})$.

$$0 = \cos\left(\frac{\pi}{2}\right) = \cos\left(\frac{\pi}{4} + \frac{\pi}{4}\right) = 2 \cos^2\left(\frac{\pi}{4}\right) - 1$$

and so $\cos(\frac{\pi}{4}) = \sqrt{2}/2$. (Why do isn't it equal to $-\sqrt{2}/2$? **Hint:** Draw a picture.) Thus $\sin(\frac{\pi}{4}) = \sqrt{2}/2$ also. (Why?) Here is another one. From 1.15,

$$\begin{aligned} 0 &= \cos\left(\frac{\pi}{2}\right) = \cos 3\left(\frac{\pi}{6}\right) \\ &= 4 \cos^3\left(\frac{\pi}{6}\right) - 3 \cos\left(\frac{\pi}{6}\right). \end{aligned}$$

Therefore, $\cos(\frac{\pi}{6}) = \frac{\sqrt{3}}{2}$ and consequently, $\sin(\frac{\pi}{6}) = \frac{1}{2}$. Here is a short table for some standard angles, sometimes called reference angles. In the table, θ refers to the radian measure of the angle. In calculus, angles are considered as real numbers, not as pointy things.

θ	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\cos \theta$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
$\sin \theta$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1

1.6 Exercises

1. The factorial symbol is defined as $n! \equiv n(n-1) \cdots 1$ whenever n is a positive integer. $0! \equiv 1$. Now the binomial coefficients are of the form

$$\frac{n!}{k!(n-k)!} \equiv \binom{n}{k}, 0 \leq k \leq n$$

Show that

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}$$

2. The binomial theorem says that

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$$

Using the identity of Problem 1 prove the binomial theorem by induction.

3. Find $\cos \theta$ and $\sin \theta$ for $\theta \in \{\frac{2\pi}{3}, \frac{3\pi}{4}, \frac{5\pi}{6}, \pi, \frac{7\pi}{6}, \frac{5\pi}{4}, \frac{4\pi}{3}, \frac{3\pi}{2}, \frac{5\pi}{3}, \frac{7\pi}{4}, \frac{11\pi}{6}, 2\pi\}$.

4. Prove $\cos^2 \theta = \frac{1+\cos 2\theta}{2}$ and $\sin^2 \theta = \frac{1-\cos 2\theta}{2}$.
5. $\pi/12 = \pi/3 - \pi/4$. Therefore, from Problem 4, $\cos(\pi/12) = \sqrt{\frac{1+(\sqrt{3}/2)}{2}}$. On the other hand,

$$\cos(\pi/12) = \cos(\pi/3 - \pi/4) = \cos \pi/3 \cos \pi/4 + \sin \pi/3 \sin \pi/4$$

and so $\cos(\pi/12) = \sqrt{2}/4 + \sqrt{6}/4$. Is there a problem here? Please explain.

6. Prove $1 + \tan^2 \theta = \sec^2 \theta$ and $1 + \cot^2 \theta = \csc^2 \theta$.
7. Prove that $\sin x \cos y = \frac{1}{2} (\sin(x+y) + \sin(x-y))$.
8. Prove that $\sin x \sin y = \frac{1}{2} (\cos(x-y) - \cos(x+y))$.
9. Prove that $\cos x \cos y = \frac{1}{2} (\cos(x+y) + \cos(x-y))$.
10. Using Problem 7, find an identity for $\sin x - \sin y$.
11. Suppose $\sin x = a$ where $0 < a < 1$. Find all possible values for
- (a) $\tan x$
 - (b) $\cot x$
 - (c) $\sec x$
 - (d) $\csc x$
 - (e) $\cos x$

12. Solve the equations and give all solutions.

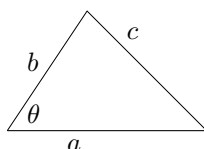
- (a) $\sin(3x) = \frac{1}{2}$
- (b) $\cos(5x) = \frac{\sqrt{3}}{2}$
- (c) $\tan(x) = \sqrt{3}$
- (d) $\sec(x) = 2$
- (e) $\sin(x+7) = \frac{\sqrt{2}}{2}$
- (f) $\cos^2(x) = \frac{1}{2}$
- (g) $\sin^4(x) = 4$

13. Sketch a graph of $y = \sin x$.
14. Sketch a graph of $y = \cos x$.
15. Sketch a graph of $y = \sin 2x$.
16. Sketch a graph of $y = \tan x$.
17. Find a formula for $\sin x \cos y$ in terms of sines and cosines of $x+y$ and $x-y$.
18. Using Problem 4 graph $y = \cos^2 x$.
19. If $f(x) = A \cos \alpha x + B \sin \alpha x$, show there exists ϕ such that

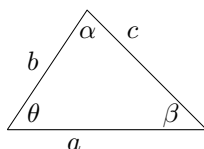
$$f(x) = \sqrt{A^2 + B^2} \sin(\alpha x + \phi).$$

Show that there also exists ψ such that $f(x) = \sqrt{A^2 + B^2} \cos(\alpha x + \psi)$. This is a very important result, enough that some of these quantities are given names. $\sqrt{A^2 + B^2}$ is called the amplitude and ϕ or ψ are called phase shifts.

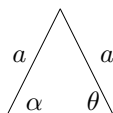
20. Using Problem 19 graph $y = \sin x + \sqrt{3} \cos x$.
21. Give all solutions to $\sin x + \sqrt{3} \cos x = \sqrt{3}$. **Hint:** Use Problem 20.
22. If ABC is a triangle where the capitol letters denote vertices of the triangle and the angle at the vertex. Let a be the length of the side opposite A and b is the length of the side opposite B and c is the length of the side opposite the vertex, C . The law of sines says $\sin(A)/a = \sin(B)/b = \sin(C)/c$. Prove the law of sines from the definition of the trigonometric functions.
23. In the picture, $a = 5, b = 3$, and $\theta = \frac{2}{3}\pi$. Find c .



24. In the picture, $\theta = \frac{1}{4}\pi$, $\alpha = \frac{2}{3}\pi$ and $c = 3$. Find a .



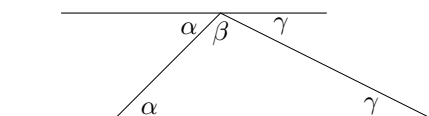
25. An isosceles triangle is one which has two equal sides. For example the following picture is of an isosceles triangle



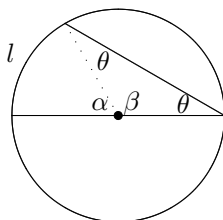
the two equal sides having length a . Show the “base angles” θ and α are equal.

Hint: You might want to use the law of sines.

26. Find a formula for $\tan(\theta + \beta)$ in terms of $\tan \theta$ and $\tan \beta$
27. Find a formula for $\tan(2\theta)$ in terms of $\tan \theta$.
28. Find a formula for $\tan(\frac{\theta}{2})$ in terms of $\tan \theta$.
29. Explain why through any point in the plane there exists a line parallel to a given line in the plane.
30. Explain why the sum of the radian measures of the angles in any triangle equals π . **Hint:** Consider the following picture and use the result of Problem 29.



31. The following picture is of an “inscribed angle”, denoted by θ in a circle of radius a . Drawing a line from the center as shown in the picture, it follows since this forms an isosceles triangle, the two base angles are equal. These are denoted as θ in the picture.



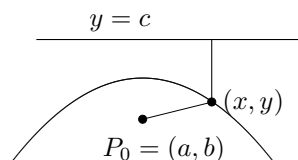
Now the radian measure of α is l/a . Using the result of Problem 30 show that the radian measure of θ equals $l/2a$.

32. The inscribed angle in Problem 31 has the special property that one side is a diameter of the circle. A general inscribed angle is just like the one shown in this problem but without the requirement that either of the sides of the angle are a diameter. Show that for a completely arbitrary inscribed angle a similar result holds to the one in Problem 31.

1.7 Parabolas, Ellipses, and Hyperbolas

1.7.1 The Parabola

A parabola is a collection of points, P in the plane such that the distance from P to a fixed line, called the directrix, is the same as the distance from P to a given point, P_0 , called the focus, as shown in the following picture.



From this geometric description, it is possible to obtain a formula for a parabola. For simplicity, assume the top (vertex) of the parabola is $(0,0)$ and the line is $y = c$. It follows the focus must be $(0, -c)$. Then if (x, y) is a point on the parabola, the geometric description requires

$$\left(x^2 + (y + c)^2\right)^{1/2} = |y - c|.$$

Squaring both sides yields

$$x^2 + y^2 + 2yc + c^2 = y^2 - 2yc + c^2$$

and so

$$x^2 = -4yc \tag{1.16}$$

If $c > 0$ then the largest y can be is 0. This follows from solving the above equation for y . Thus $y = -x^2/4c$ which is in this case is always less than or equal to 0. Therefore, the largest y can be and still satisfy the equation for the parabola is 0 when $x = 0$. Thus the vertex is $(0,0)$. Similarly, if $c < 0$, the smallest y can be is 0 and the vertex is still $(0,0)$.

This yields the following equation for a parabola.

Equation Of A Parabola: The equation for a parabola which has directrix $y = c$, focus $(0, -c)$ and vertex $(0, 0)$ is $x^2 = -4cy$.

This simple description can be used to consider more general situations in which, for example the vertex is not at $(0, 0)$.

Example 1.7.1 Find the focus and directrix of the parabola $2x^2 - 3x + 1 = 5y$.

First complete the square on the left. Thus $2(x^2 - \frac{3}{2}x + \frac{9}{16}) - \frac{9}{8} + 1 = 5y$. this yields

$$2\left(x - \frac{3}{4}\right)^2 = 5y + \frac{1}{8} = 5\left(y + \frac{1}{40}\right).$$

Dividing both sides by 2,

$$\left(x - \frac{3}{4}\right)^2 = \frac{5}{2}\left(y + \frac{1}{40}\right).$$

Now if this were just $x^2 = -4cy$, you would know the directrix is $y = c$ and the focus is $(0, -c)$. It doesn't look like this, however. Therefore, change the variables, letting $u = x - \frac{3}{4}$ and $v = y + \frac{1}{40}$. Then the equation in terms of u and v is of the form

$$u^2 = -4\left(\frac{-5}{8}\right)v \tag{1.17}$$

and so the focus for this parabola in the uv plane is $(0, \frac{5}{8})$ and its directrix is $v = -\frac{5}{8}$. Since $x = u + \frac{3}{4}$ and $y = v - \frac{1}{40}$, it follows the focus of the parabola in the xy plane is $(\frac{3}{4}, \frac{5}{8} - \frac{1}{40}) = (\frac{3}{4}, \frac{3}{5})$. Similarly, the directrix is $y = \frac{-5}{8} - \frac{1}{40} = -\frac{13}{20}$.

Sometimes the y variable is squared and the x variable is not squared. This corresponds to having a directrix of the form $x = c$ and to the parabola opening either to the right or the left. The same reasoning given above yields the following equation of a sideways parabola.

Equation Of A Parabola: The equation for a parabola which has directrix $x = c$, focus $(-c, 0)$ and vertex $(0, 0)$ is $y^2 = -4cx$.

Example 1.7.2 Find the focus and directrix of the parabola $x = 3y^2 + 2y + 1$.

Complete the square on the right to get

$$\begin{aligned} x &= 3\left(y^2 + \frac{2}{3}y + \frac{1}{9}\right) - \frac{1}{3} + 1 \\ &= 3\left(y + \frac{1}{3}\right)^2 + \frac{2}{3} \end{aligned}$$

and so

$$-4\left(\frac{-1}{12}\right)\left(x - \frac{2}{3}\right) = \left(y + \frac{1}{3}\right)^2. \tag{1.18}$$

changing variables letting $u = x - \frac{2}{3}$ and $v = y + \frac{1}{3}$, this gives

$$-4\left(\frac{-1}{12}\right)u = v^2.$$

Therefore, in the uv plane, the directrix is $u = \frac{-1}{12}$ and the focus is $(\frac{1}{12}, 0)$ while the vertex is $(0, 0)$. In terms of the original variables the directrix is $x = \frac{2}{3} - \frac{1}{12} = \frac{7}{12}$, the focus is $(\frac{1}{12} + \frac{2}{3}, 0 - \frac{1}{3}) = (\frac{3}{4}, -\frac{1}{3})$ while the vertex is $(\frac{2}{3}, \frac{-1}{3})$.

This illustrates the following procedure.

Procedure 1.7.3 To find the focus, directrix and vertex of a parabola of the form $x = ay^2 + by + c$, you complete the square and otherwise massage things to get it in the form

$$-4c(x - p) = (y - q)^2$$

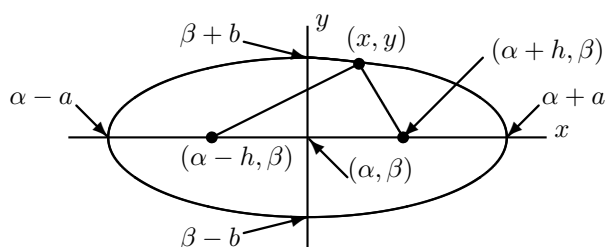
and then the vertex is at (p, q) and the focus is at $(p - c, q)$ while the directrix is $x = p + c$. To find the focus directrix and vertex of a parabola of the form $y = ax^2 + bx + c$, you complete the square and otherwise massage things to obtain an equation of the form

$$(x - p)^2 = -4c(y - q).$$

then the vertex is (p, q) the focus is $(p, q - c)$ and the directrix is $y = q + c$.

1.7.2 The Ellipse

With an ellipse, there are two points, P_1 and P_2 which are fixed and the ellipse consists of the set of points, P such that $d(P, P_1) + d(P, P_2) = d$, where d is a fixed positive number. These two points are called the foci of the ellipse. Each is called a focus point by itself. Here is a picture in case the two points are of the form $(\alpha - h, \beta)$ and $(\alpha + h, \beta)$.



Note in the picture $2a$ is the length of this ellipse and $2b$ is its height. The major axis is thus $2a$ and the minor axis is $2b$. You can think of a string stretched tight and the pencil being at the point on the graph of the ellipse. Thus you can see from the picture that $d = 2a$. Just imagine the point is at $(a + \alpha, \beta)$, the right vertex of the ellipse.

For simplicity assume the foci are at $(-c, 0)$ and $(c, 0)$ where $c \geq 0$. Then from the description and letting $2a$ be the length of the ellipse between the two farthest vertices, the equation of the ellipse is given by

$$\sqrt{(x + c)^2 + y^2} + \sqrt{(x - c)^2 + y^2} = 2a$$

Therefore,

$$\sqrt{(x + c)^2 + y^2} = 2a - \sqrt{(x - c)^2 + y^2}$$

and so squaring both sides,

$$x^2 + 2xc + c^2 + y^2 = 4a^2 + x^2 - 2xc + c^2 + y^2 - 4a\sqrt{(x - c)^2 + y^2}$$

Subtracting like terms from both sides and then dividing by 4

$$xc = a^2 - a\sqrt{(x - c)^2 + y^2}$$

It follows

$$xc - a^2 = -a\sqrt{(x - c)^2 + y^2}.$$

Now square both sides of this to obtain

$$x^2c^2 - 2xca^2 + a^4 = a^2(x^2 - 2cx + c^2 + y^2).$$

Next simplify this some more to obtain

$$a^2(a^2 - c^2) = (a^2 - c^2)x^2 + a^2y^2.$$

Therefore,

$$1 = \frac{x^2}{a^2} + \frac{y^2}{b^2}$$

where $b^2 = a^2 - c^2$. This verifies the following equation of an ellipse.

Equation Of An Ellipse: Suppose the ellipse is centered at $(0, 0)$ has major axis parallel to the x axis and minor axis parallel to the y axis and that the major axis is $2a$ while the minor axis is $2b$. Then the equation of the ellipse is

$$1 = \frac{x^2}{a^2} + \frac{y^2}{b^2}$$

and the focus points are at $(-\sqrt{a^2 - b^2}, 0)$ and $(\sqrt{a^2 - b^2}, 0)$. Switching the variables, if the ellipse is centered at $(0, 0)$ and the major axis of length $2a$ is parallel to the y axis while the minor axis of length $2b$ is parallel to the x axis. Then the equation of the ellipse is

$$1 = \frac{x^2}{b^2} + \frac{y^2}{a^2}$$

and the focus points are at $(0, -\sqrt{a^2 - b^2})$ and $(0, \sqrt{a^2 - b^2})$.

As in the case of the parabola, you can use this to find the foci of ellipses which are not centered at $(0, 0)$.

Example 1.7.4 Find the focus points for the ellipse $\frac{(x-1)^2}{9} + \frac{(y-2)^2}{4} = 1$.

Let $u = x - 1$ and $v = y - 2$. Then the focus points of the identical ellipse in the uv plane would be $(\sqrt{5}, 0)$ and $(-\sqrt{5}, 0)$. Now $x = u + 1, y = v + 2$ so the focus points of the original ellipse in the xy plane are $(\sqrt{5} + 1, 2)$ and $(-\sqrt{5} + 1, 2)$.

Example 1.7.5 Show the following is the equation of an ellipse and find its major and minor axes along with the focus points. $2x^2 + 4x + y^2 + 2y + 2 = 0$.

As usual you have to complete the square in this. Thus $2(x^2 + 2x + 1) + y^2 + 2y + 1 = 1$. Therefore,

$$\frac{(x+1)^2}{(1/\sqrt{2})^2} + (y+1)^2 = 1$$

The major axis has length 2 and is parallel to the y axis and the minor axis has length $\sqrt{2}$ and is parallel to the x axis. Let $x + 1 = u$ and $y + 1 = v$. Then in terms of u, v this equation is of the form

$$\frac{u^2}{(1/\sqrt{2})^2} + v^2 = 1.$$

The focus points in the uv plane are $(0, \pm\sqrt{1 - (1/2)}) = (0, \pm\frac{1}{\sqrt{2}})$. Therefore, the focus points of the ellipse in the xy plane are $(-1, -1 \pm \frac{1}{\sqrt{2}})$.

This illustrates the following procedure.

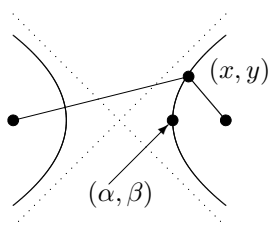
Procedure 1.7.6 For $a, b > 0$ and an ellipse in the form $Ax^2 + By^2 + Cx + Dy + E = 0$ you first complete the square and then massage to obtain something of the form

$$\frac{(x-p)^2}{a^2} + \frac{(y-q)^2}{b^2} = 1$$

where a, b are positive numbers. If $a > b$ then the major axis has length $2a$ and is parallel to the x axis. In this case the focus points are at $(p - \sqrt{a^2 - b^2}, q)$ and $(p + \sqrt{a^2 - b^2}, q)$ and the minor axis has length $2b$ and is parallel to the y axis. If $a < b$ then the major axis is parallel to the y axis and has length $2b$. In this case the focus points are $(p, q + \sqrt{b^2 - a^2})$ and $(p, q - \sqrt{b^2 - a^2})$ and the minor axis is parallel to the x axis and has length $2a$. If $a = b$, the ellipse is a circle and it has center at (p, q) with radius equal to $a = b$.

1.7.3 The Hyperbola

With a hyperbola, there are two points, P_1 and P_2 which are fixed and the hyperbola consists of the set of points, P such that $d(P, P_1) - d(P, P_2) = d$, where d is a fixed positive number. These two points are called the foci of the hyperbola. Each is called a focus point by itself. The following picture is descriptive of the above situation.



Now one can obtain an equation which will describe a hyperbola. For simplicity, consider the case where the focus points are $(-c, 0)$ and $(c, 0)$ and let d denote the difference between the two distances. Then from the above description of the hyperbola,

$$\sqrt{(x+c)^2 + y^2} - \sqrt{(x-c)^2 + y^2} = d.$$

Then

$$x^2 + 2cx + y^2 = d^2 + 2d\sqrt{(x-c)^2 + y^2} + x^2 - 2cx + y^2.$$

Subtracting like terms,

$$4cx - d^2 = 2d\sqrt{(x-c)^2 + y^2}.$$

Therefore, squaring both sides again,

$$16c^2x^2 + d^4 - 8xcd^2 = 4d^2(x^2 - 2cx + c^2 + y^2)$$

Another simplification gives

$$d^2(d^2 - 4c^2) = 4(d^2 - 4c^2)x^2 + 4d^2y^2.$$

Therefore,

$$1 = \frac{x^2}{(d^2/4)} - \frac{y^2}{(4c^2 - d^2)/4}$$

Therefore, the vertices of this hyperbola, one of which is labeled (α, β) in the above picture are in this case at $(\frac{d}{2}, 0)$ and $(-\frac{d}{2}, 0)$ so d is also the distance between the two vertices which is less than the distance between the two focus points, $2c$. Therefore,

$4c^2 - d^2 > 0$. Let $a^2 = d^2/4$ and let $b^2 = (4c^2 - d^2)/4$ so the equation of the hyperbola is of the form

$$1 = \frac{x^2}{a^2} - \frac{y^2}{b^2}.$$

Observe also the relation between a, b and c . $a^2 + b^2 = c^2$.

Another interesting thing about hyperbolas is the asymptotes. Solve the equation above for x .

$$x = \pm \left(a^2 + \frac{a^2 y^2}{b^2} \right)^{1/2}$$

Now for large $|y|$, this is essentially equal to $\pm \frac{ay}{b}$. The asymptotes of the hyperbola are $x = \frac{a}{b}y$ and $x = -\frac{a}{b}y$. These are the dotted lines in the above picture.

Equation For A Hyperbola: Suppose you have the relation,

$$1 = \frac{x^2}{a^2} - \frac{y^2}{b^2}.$$

Then this is a hyperbola having vertices at $(a, 0)$ and $(-a, 0)$ and foci at $(-\sqrt{a^2 + b^2}, 0)$ and $(\sqrt{a^2 + b^2}, 0)$. Its asymptotes are given by

$$x = \pm \frac{a}{b}y.$$

Switching the variables, suppose now you have the relation

$$1 = \frac{y^2}{b^2} - \frac{x^2}{a^2}.$$

Then it is a hyperbola having vertices at $(0, b)$ and $(0, -b)$ and foci at $(0, -\sqrt{a^2 + b^2})$ and $(0, \sqrt{a^2 + b^2})$. Its asymptotes are

$$y = \pm \frac{bx}{a}.$$

Example 1.7.7 Find the focus points and vertices for the hyperbola $\frac{(x-1)^2}{4} - \frac{(y-2)^2}{9} = 1$.

Change the variables. Let $u = x - 1$ and $v = y - 2$. Then the focus points of the identical hyperbola in the uv plane would be $(\sqrt{13}, 0)$ and $(-\sqrt{13}, 0)$. Now $x = u + 1, y = v + 2$ so the focus points of the original hyperbola in the xy plane are $(\sqrt{13} + 1, 2)$ and $(-\sqrt{13} + 1, 2)$. The vertices are obtained by taking $y = 2$, $(3, 2)$ and $(-1, 2)$.

Example 1.7.8 Show that the relation, $2y^2 + 4y - 4x^2 + 8x = 3$ is a hyperbola and find its foci.

As usual, you complete the square. Thus

$$2(y^2 + 2y + 1) - 4(x^2 - 2x + 1) = 3 + 2 - 4 = 1.$$

Therefore,

$$\frac{(y+1)^2}{(1/\sqrt{2})^2} - \frac{(x-1)^2}{(1/2)^2} = 1.$$

Therefore, the vertices are $\left(1, -1 + \frac{1}{\sqrt{2}}\right)$ and $\left(1, -1 - \frac{1}{\sqrt{2}}\right)$. The foci are at

$$\left(1, -1 - \frac{\sqrt{3}}{2}\right)$$

and $\left(1, -1 + \frac{\sqrt{3}}{2}\right)$.

This illustrates the following procedure for a hyperbola.

Procedure 1.7.9 Suppose $Ax^2 + By^2 + Cx + Dy + E = 0$ and A and B have different signs. Complete the square and massage to obtain an expression of one of the following forms.

$$\begin{aligned}\frac{(x-p)^2}{a^2} - \frac{(y-q)^2}{b^2} &= 1 \text{ a.)} \\ \frac{(y-q)^2}{b^2} - \frac{(x-p)^2}{a^2} &= 1 \text{ b.)}\end{aligned}$$

where a, b are positive numbers. In the case of a.) the focus points are at $(p + \sqrt{a^2 + b^2}, q)$ and $(p - \sqrt{a^2 + b^2}, q)$, the vertices are at $(p + a, q)$, $(p - a, q)$, and the asymptotes are of the form

$$\frac{(x-p)}{a} \pm \frac{(y-q)}{b} = 1$$

In the case of b.) the focus points are at $(p, q + \sqrt{a^2 + b^2})$ and $(p, q - \sqrt{a^2 + b^2})$, the vertices are at $(p, q + b)$, $(p, q - b)$, and the asymptotes are of the form

$$\frac{(y-q)}{b} \pm \frac{(x-p)}{a} = 1.$$

1.8 Exercises

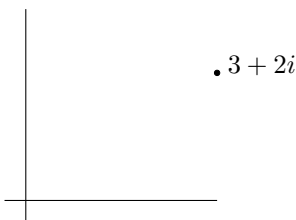
1. Consider $y = 2x^2 + 3x + 7$. Find the focus and the directrix of this parabola.
2. Sketch a graph of the ellipse whose equation is $\frac{(x-1)^2}{4} + \frac{(y-2)^2}{9} = 1$.
3. Sketch a graph of the ellipse whose equation is $\frac{(x-1)^2}{9} + \frac{(y-2)^2}{4} = 1$.
4. Sketch a graph of the hyperbola, $\frac{x^2}{4} - \frac{y^2}{9} = 1$.
5. Sketch a graph of the hyperbola, $\frac{y^2}{4} - \frac{x^2}{9} = 1$.
6. Find the focus points for the hyperbola $\frac{(x-1)^2}{4} - \frac{(y-2)^2}{9} = 1$.
7. Consider the hyperbola, $\frac{y^2}{4} - \frac{x^2}{9} = 1$. Show that $y = \pm \sqrt{b^2 + \frac{b^2 x^2}{a^2}}$. The straight lines $y = \frac{bx}{a}$ and $y = -\frac{bx}{a}$ are called the asymptotes of the hyperbola. Show that for large x , $\sqrt{b^2 + \frac{b^2 x^2}{a^2}} - \frac{bx}{a}$ is very small.
8. What is the diameter of the ellipse, $\frac{(x-1)^2}{9} + \frac{(y-2)^2}{4} = 1$? Remember the diameter of a set, S is defined as $\sup \{|\mathbf{x} - \mathbf{y}| : \mathbf{x}, \mathbf{y} \in S\}$.
9. Consider $9x^2 - 36x + 32 - 4y^2 - 8y = 36$. Identify this as either an ellipse, a hyperbola or a parabola. Then find its focus point(s) and its directrix if it is a parabola. **Hint:** First complete the square.
10. Consider $4x^2 - 8x + 68 + 16y^2 - 64y = 64$. Identify this as either an ellipse, a hyperbola or a parabola. Then find its focus point(s) and its directrix if it is a parabola. **Hint:** First complete the square.
11. Consider $-8x + 8 + 16y^2 - 64y = 64$. Identify this as either an ellipse, a hyperbola or a parabola. Then find its focus point(s) and its directrix if it is a parabola. **Hint:** First complete the square.
12. Consider $5x + 3y^2 + 2y = 7$. Identify this as either an ellipse, a hyperbola or a parabola. Then find its focus point(s) and its directrix if it is a parabola.

13. Consider the two points, $P_1 = (0, 0)$ and $P_2 = (1, 1)$. Find the equation of the ellipse defined by $d(P, P_1) + d(P, P_2) = 4$ which has these as focus points.
14. Find the equation of the parabola which has focus $(0, 0)$ and directrix $x + y = 1$. This is pretty hard. To do it you need to figure out how to find the distance between a point and the given line.
15. As explained earlier, $(\cos t, \sin t)$ for $t \in \mathbb{R}$ is a point on the circle of radius 1. Find a formula for the coordinates of a point on the ellipse, $\frac{(x-2)^2}{4} + \frac{(y+1)^2}{8} = 1$. **Hint:** This says $\left(\frac{x-2}{2}, \frac{y+1}{\sqrt{8}}\right)$ is a point on the unit circle.

1.9 The Complex Numbers

This is a brief treatment of the complex numbers. This will not be needed in Calculus but you will need it when you take differential equations and various other subjects so it is a good idea to consider the subject. These things used to be taught in precalculus classes and people were expected to know them before taking calculus. However, if you are in a hurry to get to calculus, you can skip this section.

Just as a real number should be considered as a point on the line, a complex number is considered a point in the plane which can be identified in the usual way using the Cartesian coordinates of the point. Thus (a, b) identifies a point whose x coordinate is a and whose y coordinate is b . In dealing with complex numbers, such a point is written as $a + ib$. For example, in the following picture, I have graphed the point $3 + 2i$. You see it corresponds to the point in the plane whose coordinates are $(3, 2)$.



Multiplication and addition are defined in the most obvious way subject to the convention that $i^2 = -1$. Thus,

$$(a + ib) + (c + id) = (a + c) + i(b + d)$$

and

$$\begin{aligned} (a + ib)(c + id) &= ac + iad + ibc + i^2bd \\ &= (ac - bd) + i(bc + ad). \end{aligned}$$

Every non zero complex number, $a + ib$, with $a^2 + b^2 \neq 0$, has a unique multiplicative inverse.

$$\frac{1}{a + ib} = \frac{a - ib}{a^2 + b^2} = \frac{a}{a^2 + b^2} - i \frac{b}{a^2 + b^2}.$$

You should prove the following theorem.

Theorem 1.9.1 *The complex numbers with multiplication and addition defined as above form a field satisfying all the field axioms,*

Axiom 1.9.2 $x + y = y + x$, (commutative law for addition)

Axiom 1.9.3 $x + 0 = x$, (additive identity).

Axiom 1.9.4 For each $x \in \mathbb{R}$, there exists $-x \in \mathbb{R}$ such that $x + (-x) = 0$, (existence of additive inverse).

Axiom 1.9.5 $(x + y) + z = x + (y + z)$, (associative law for addition).

Axiom 1.9.6 $xy = yx$, (commutative law for multiplication).

Axiom 1.9.7 $(xy)z = x(yz)$, (associative law for multiplication).

Axiom 1.9.8 $1x = x$, (multiplicative identity).

Axiom 1.9.9 For each $x \neq 0$, there exists x^{-1} such that $xx^{-1} = 1$. (existence of multiplicative inverse).

Axiom 1.9.10 $x(y + z) = xy + xz$. (distributive law).

The field of complex numbers is denoted as $\mathbb{C}$. An important construction regarding complex numbers is the complex conjugate denoted by a horizontal line above the number. It is defined as follows.

$$\overline{a + ib} \equiv a - ib.$$

What it does is reflect a given complex number across the x axis. Algebraically, the following formula is easy to obtain.

$$(\overline{a + ib})(a + ib) = a^2 + b^2.$$

Definition 1.9.11 Define the absolute value of a complex number as follows.

$$|a + ib| \equiv \sqrt{a^2 + b^2}.$$

Thus, denoting by z the complex number, $z = a + ib$,

$$|z| = (z\bar{z})^{1/2}.$$

With this definition, it is important to note the following. Be sure to verify this. It is not too hard but you need to do it.

Remark 1.9.12 : Let $z = a + ib$ and $w = c + id$. Then $|z - w| = \sqrt{(a - c)^2 + (b - d)^2}$. Thus the distance between the point in the plane determined by the ordered pair, (a, b) and the ordered pair (c, d) equals $|z - w|$ where z and w are as just described.

For example, consider the distance between $(2, 5)$ and $(1, 8)$. From the distance formula this distance equals $\sqrt{(2 - 1)^2 + (5 - 8)^2} = \sqrt{10}$. On the other hand, letting $z = 2 + i5$ and $w = 1 + i8$, $z - w = 1 - i3$ and so $(z - w)(\overline{z - w}) = (1 - i3)(1 + i3) = 10$ so $|z - w| = \sqrt{10}$, the same thing obtained with the distance formula.

Proposition 1.9.13 Let z, w be complex numbers. Then the triangle inequality holds.

$$|z + w| \leq |z| + |w|, \quad ||z| - |w|| \leq |z - w|.$$

Proof: From the definition,

$$\begin{aligned} |z + w|^2 &= (z + w)(\bar{z} + \bar{w}) \\ &= z\bar{z} + w\bar{w} + z\bar{w} + \bar{z}w = |z|^2 + 2\operatorname{Re}(z\bar{w}) + |w|^2 \\ &\leq |z|^2 + 2|z||w| + |w|^2 = (|z| + |w|)^2 \end{aligned}$$

and taking the square root, this gives the first inequality above. To get the second,

$$z = z - w + w, \quad w = w - z + z$$

and so by the first form of the inequality

$$|z| \leq |z - w| + |w|, \quad |w| \leq |z - w| + |z|$$

and so both $|z| - |w|$ and $|w| - |z|$ are no larger than $|z - w|$ and this proves the second version because $||z| - |w||$ is one of $|z| - |w|$ or $|w| - |z|$. This proves the proposition.

Complex numbers, are often written in the so called polar form which is described next. Suppose $x + iy$ is a complex number. Then

$$x + iy = \sqrt{x^2 + y^2} \left(\frac{x}{\sqrt{x^2 + y^2}} + i \frac{y}{\sqrt{x^2 + y^2}} \right).$$

Now note that

$$\left(\frac{x}{\sqrt{x^2 + y^2}} \right)^2 + \left(\frac{y}{\sqrt{x^2 + y^2}} \right)^2 = 1$$

and so

$$\left(\frac{x}{\sqrt{x^2 + y^2}}, \frac{y}{\sqrt{x^2 + y^2}} \right)$$

is a point on the unit circle. Therefore, there exists a unique angle, $\theta \in [0, 2\pi)$ such that

$$\cos \theta = \frac{x}{\sqrt{x^2 + y^2}}, \quad \sin \theta = \frac{y}{\sqrt{x^2 + y^2}}.$$

The polar form of the complex number is then

$$r(\cos \theta + i \sin \theta)$$

where θ is this angle just described and $r = \sqrt{x^2 + y^2}$.

A fundamental identity is the formula of De Moivre which follows.

Theorem 1.9.14 *Let $r > 0$ be given. Then if n is a positive integer,*

$$[r(\cos t + i \sin t)]^n = r^n (\cos nt + i \sin nt).$$

Proof: It is clear the formula holds if $n = 1$. Suppose it is true for n .

$$[r(\cos t + i \sin t)]^{n+1} = [r(\cos t + i \sin t)]^n [r(\cos t + i \sin t)]$$

which by induction equals

$$\begin{aligned} &= r^{n+1} (\cos nt + i \sin nt) (\cos t + i \sin t) \\ &= r^{n+1} ((\cos nt \cos t - \sin nt \sin t) + i (\sin nt \cos t + \cos nt \sin t)) \\ &= r^{n+1} (\cos(n+1)t + i \sin(n+1)t) \end{aligned}$$

by the formulas for the cosine and sine of the sum of two angles.

Corollary 1.9.15 *Let z be a non zero complex number. Then there are always exactly k k^{th} roots of z in $\mathbb{C}$.*

Proof: Let $z = x + iy$ and let $z = |z|(\cos t + i \sin t)$ be the polar form of the complex number. By De Moivre's theorem, a complex number,

$$r(\cos \alpha + i \sin \alpha),$$

is a k^{th} root of z if and only if

$$r^k(\cos k\alpha + i \sin k\alpha) = |z|(\cos t + i \sin t).$$

This requires $r^k = |z|$ and so $r = |z|^{1/k}$ and also both $\cos(k\alpha) = \cos t$ and $\sin(k\alpha) = \sin t$. This can only happen if

$$k\alpha = t + 2l\pi$$

for l an integer. Thus

$$\alpha = \frac{t + 2l\pi}{k}, l \in \mathbb{Z}$$

and so the k^{th} roots of z are of the form

$$|z|^{1/k} \left(\cos \left(\frac{t + 2l\pi}{k} \right) + i \sin \left(\frac{t + 2l\pi}{k} \right) \right), l \in \mathbb{Z}.$$

Since the cosine and sine are periodic of period 2π , there are exactly k distinct numbers which result from this formula.

Example 1.9.16 *Find the three cube roots of i .*

First note that $i = 1(\cos(\frac{\pi}{2}) + i \sin(\frac{\pi}{2}))$. Using the formula in the proof of the above corollary, the cube roots of i are

$$1 \left(\cos \left(\frac{(\pi/2) + 2l\pi}{3} \right) + i \sin \left(\frac{(\pi/2) + 2l\pi}{3} \right) \right)$$

where $l = 0, 1, 2$. Therefore, the roots are

$$\cos \left(\frac{\pi}{6} \right) + i \sin \left(\frac{\pi}{6} \right), \cos \left(\frac{5}{6}\pi \right) + i \sin \left(\frac{5}{6}\pi \right),$$

and

$$\cos \left(\frac{3}{2}\pi \right) + i \sin \left(\frac{3}{2}\pi \right).$$

Thus the cube roots of i are $\frac{\sqrt{3}}{2} + i(\frac{1}{2})$, $\frac{-\sqrt{3}}{2} + i(\frac{1}{2})$, and $-i$.

The ability to find k^{th} roots can also be used to factor some polynomials.

Example 1.9.17 *Factor the polynomial $x^3 - 27$.*

First find the cube roots of 27. By the above procedure using De Moivre's theorem, these cube roots are $3, 3\left(\frac{-1}{2} + i\frac{\sqrt{3}}{2}\right)$, and $3\left(\frac{-1}{2} - i\frac{\sqrt{3}}{2}\right)$. Therefore, $x^3 - 27 =$

$$(x - 3) \left(x - 3 \left(\frac{-1}{2} + i\frac{\sqrt{3}}{2} \right) \right) \left(x - 3 \left(\frac{-1}{2} - i\frac{\sqrt{3}}{2} \right) \right).$$

Note also $\left(x - 3 \left(\frac{-1}{2} + i\frac{\sqrt{3}}{2} \right) \right) \left(x - 3 \left(\frac{-1}{2} - i\frac{\sqrt{3}}{2} \right) \right) = x^2 + 3x + 9$ and so

$$x^3 - 27 = (x - 3)(x^2 + 3x + 9)$$

where the quadratic polynomial, $x^2 + 3x + 9$ cannot be factored without using complex numbers.

1.10 Exercises

1. Let $z = 5 + i9$. Find z^{-1} .
2. Let $z = 2 + i7$ and let $w = 3 - i8$. Find $zw, z + w, z^2$, and w/z .
3. Give the complete solution to $x^4 + 16 = 0$.
4. Graph the complex cube roots of 8 in the complex plane. Do the same for the four fourth roots of 16.
5. If z is a complex number, show there exists ω a complex number with $|\omega| = 1$ and $\omega z = |z|$.
6. De Moivre's theorem says $[r(\cos t + i \sin t)]^n = r^n(\cos nt + i \sin nt)$ for n a positive integer. Does this formula continue to hold for all integers, n , even negative integers? Explain.
7. You already know formulas for $\cos(x + y)$ and $\sin(x + y)$ and these were used to prove De Moivre's theorem. Now using De Moivre's theorem, derive a formula for $\sin(5x)$ and one for $\cos(5x)$.
8. If z and w are two complex numbers and the polar form of z involves the angle θ while the polar form of w involves the angle ϕ , show that in the polar form for zw the angle involved is $\theta + \phi$. Also, show that in the polar form of a complex number, z , $r = |z|$.
9. Factor $x^3 + 8$ as a product of linear factors.
10. Write $x^3 + 27$ in the form $(x + 3)(x^2 + ax + b)$ where $x^2 + ax + b$ cannot be factored any more using only real numbers.
11. Completely factor $x^4 + 16$ as a product of linear factors.
12. Factor $x^4 + 16$ as the product of two quadratic polynomials each of which cannot be factored further without using complex numbers.
13. If z, w are complex numbers prove $\overline{zw} = \overline{z}\overline{w}$ and then show by induction that $\overline{z_1 \cdots z_m} = \overline{z_1} \cdots \overline{z_m}$. Also verify that $\sum_{k=1}^m z_k = \sum_{k=1}^m \overline{z_k}$. In words this says the conjugate of a product equals the product of the conjugates and the conjugate of a sum equals the sum of the conjugates.
14. Suppose $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ where all the a_k are real numbers. Suppose also that $p(z) = 0$ for some $z \in \mathbb{C}$. Show it follows that $p(\overline{z}) = 0$ also.
15. I claim that $1 = -1$. Here is why.

$$-1 = i^2 = \sqrt{-1}\sqrt{-1} = \sqrt{(-1)^2} = \sqrt{1} = 1.$$

This is clearly a remarkable result but is there something wrong with it? If so, what is wrong?

16. De Moivre's theorem is really a grand thing. I plan to use it now for rational exponents, not just integers.

$$1 = 1^{(1/4)} = (\cos 2\pi + i \sin 2\pi)^{1/4} = \cos(\pi/2) + i \sin(\pi/2) = i.$$

Therefore, squaring both sides it follows $1 = -1$ as in the previous problem. What does this tell you about De Moivre's theorem? Is there a profound difference between raising numbers to integer powers and raising numbers to non integer powers?

17. Review Problem 6 at this point. Now here is another question: If n is an integer, is it always true that $(\cos \theta - i \sin \theta)^n = \cos(n\theta) - i \sin(n\theta)$? Explain.
18. Suppose you have any polynomial in $\cos \theta$ and $\sin \theta$. By this I mean an expression of the form $\sum_{\alpha=0}^m \sum_{\beta=0}^n a_{\alpha\beta} \cos^\alpha \theta \sin^\beta \theta$ where $a_{\alpha\beta} \in \mathbb{C}$. Can this always be written in the form $\sum_{\gamma=-(n+m)}^{m+n} b_\gamma \cos \gamma\theta + \sum_{\tau=-(n+m)}^{n+m} c_\tau \sin \tau\theta$? Explain.

Functions

2.0.1 Outcomes

1. Understand the definition of a function and be able to tell whether a given relation defines a function. Define and understand the domain of a function and be able to identify the domain when the function comes as a formula.
2. Understand the definition of a continuous function. Know what it means for a function to be continuous at a point. Be able to prove a given function is continuous or is not continuous.
3. Explain why $\lim_{x \rightarrow 0} \frac{\sin(x)}{x} = 1$
4. Be able to explain why the circular functions are continuous.
5. Understand the extreme value theorem and the intermediate value theorem for continuous functions.
6. Explain the concept of limit of a function and be able to find the limit of a function from the definition.
7. Describe the limit of a sequence. Understand the relation between the limit of a sequence and continuity.
8. Understand uniform continuity and the proofs of the theorems about continuous functions.

2.1 Functions And Sequences

By this time, you have seen several examples of functions such as the trig. functions. It is a good idea to formalize this concept before proceeding further. The concept of a function is that of something which gives a unique output for a given input.

Definition 2.1.1 Consider two sets, D and R along with a rule which assigns a unique element of R to every element of D . This rule is called a function and it is denoted by a letter such as f . The symbol, $D(f) = D$ is called the domain of f . The set R , also written $R(f)$, is called the range of f . The set of all elements of R which are of the form $f(x)$ for some $x \in D$ is often denoted by $f(D)$. When $R = f(D)$, the function, f , is said to be onto. It is common notation to write $f : D(f) \rightarrow R$ to denote the situation just described in this definition where f is a function defined on D having values in R .

Example 2.1.2 Consider the list of numbers, $\{1, 2, 3, 4, 5, 6, 7\} \equiv D$. Define a function which assigns an element of D to $R \equiv \{2, 3, 4, 5, 6, 7, 8\}$ by $f(x) \equiv x + 1$ for each $x \in D$.

In this example there was a clearly defined procedure which determined the function. However, sometimes there is no discernible procedure which yields a particular function.

Example 2.1.3 Consider the ordered pairs, $(1, 2), (2, -2), (8, 3), (7, 6)$ and let

$$D \equiv \{1, 2, 8, 7\},$$

the set of first entries in the given set of ordered pairs, $R \equiv \{2, -2, 3, 6\}$, the set of second entries, and let $f(1) = 2, f(2) = -2, f(8) = 3$, and $f(7) = 6$.

Sometimes functions are not given in terms of a formula. For example, consider the following function defined on the positive real numbers having the following definition.

Example 2.1.4 For $x \in \mathbb{R}$ define

$$f(x) = \begin{cases} \frac{1}{n} & \text{if } x = \frac{m}{n} \text{ in lowest terms for } m, n \in \mathbb{Z} \\ 0 & \text{if } x \text{ is not rational} \end{cases} \quad (2.1)$$

This is a very interesting function called the Dirichlet function. Note that it is not defined in a simple way from a formula.

Example 2.1.5 Let D consist of the set of people who have lived on the earth except for Adam and for $d \in D$, let $f(d) \equiv$ the biological father of d . Then f is a function.

This function is not the sort of thing studied in calculus but it is a function just the same. The next functions are studied in calculus.

Example 2.1.6 Consider a weight which is suspended at one end of a spring which is attached at the other end to the ceiling. Suppose the weight has extended the spring so that the force exerted by the spring exactly balances the force resulting from the weight on the spring. Measure the displacement of the mass, x , from this point with the positive direction being up, and define a function as follows: $x(t)$ will equal the displacement of the spring at time t given knowledge of the velocity of the weight and the displacement of the weight at some particular time.

Example 2.1.7 Certain chemicals decay with time. Suppose A_0 is the amount of chemical at some given time. Then you could let $A(t)$ denote the amount of the chemical at time t .

These last two examples show how physical problems can result in functions. Examples 2.1.6 and 2.1.7 are considered later in the book and techniques for finding $x(t)$ and $A(t)$ from the given conditions are presented.

In this chapter the functions are defined on some subset of $\mathbb{R}$ having values in $\mathbb{R}$. Later this will be generalized. When $D(f)$ is not specified, it is understood to consist of everything for which f makes sense. The following definition gives several ways to make new functions from old ones.

Definition 2.1.8 Let f, g be functions with values in $\mathbb{R}$. Let a, b be points of $\mathbb{R}$. Then $af + bg$ is the name of a function whose domain is $D(f) \cap D(g)$ which is defined as

$$(af + bg)(x) = af(x) + bg(x).$$

The function, fg is the name of a function which is defined on $D(f) \cap D(g)$ given by

$$(fg)(x) = f(x)g(x).$$

Similarly for k an integer, f^k is the name of a function defined as

$$f^k(x) = (f(x))^k$$

The function, f/g is the name of a function whose domain is

$$D(f) \cap \{x \in D(g) : g(x) \neq 0\}$$

defined as

$$(f/g)(x) = f(x)/g(x).$$

If $f : D(f) \rightarrow X$ and $g : D(g) \rightarrow Y$, then $g \circ f$ is the name of a function whose domain is

$$\{x \in D(f) : f(x) \in D(g)\}$$

which is defined as

$$g \circ f(x) \equiv g(f(x)).$$

This is called the composition of the two functions.

You should note that $f(x)$ is not a function. It is the value of the function at the point, x . The name of the function is f . Nevertheless, people often write $f(x)$ to denote a function and it doesn't cause too many problems in beginning courses. When this is done, the variable, x should be considered as a generic variable free to be anything in $D(f)$. I will use this slightly sloppy abuse of notation whenever convenient. Thus, $x^2 + 4$ may mean the function, f , given by $f(x) = x^2 + 4$.

Sometimes people get hung up on formulas and think that the only functions of importance are those which are given by some simple formula. It is a mistake to think this way. Functions involve a domain and a range and a function is determined by what it does. Functions are well described by a well known scripture, Matthew 7:20, Wherefore by their fruits ye shall know them. When you have specified what it does to something in its domain, you have told what the function is.

Example 2.1.9 Let $f(t) = t$ and $g(t) = 1 + t$. Then $fg : \mathbb{R} \rightarrow \mathbb{R}$ is given by

$$fg(t) = t(1 + t) = t + t^2.$$

Example 2.1.10 Let $f(t) = 2t + 1$ and $g(t) = \sqrt{1 + t}$. Then

$$g \circ f(t) = \sqrt{1 + (2t + 1)} = \sqrt{2t + 2}$$

for $t \geq -1$. If $t < -1$ the inside of the square root sign is negative so makes no sense. Therefore, $g \circ f : \{t \in \mathbb{R} : t \geq -1\} \rightarrow \mathbb{R}$.

Note that in this last example, it was necessary to fuss about the domain of $g \circ f$ because g is only defined for certain values of t .

The concept of a one to one function is very important. This is discussed in the following definition.

Definition 2.1.11 For any function, $f : D(f) \subseteq X \rightarrow Y$, define the following set known as the inverse image of y .

$$f^{-1}(y) \equiv \{x \in D(f) : f(x) = y\}.$$

There may be many elements in this set, but when there is always at most one element in this set for all $y \in f(D(f))$, the function f is one to one sometimes written, $1 - 1$. Thus f is one to one, $1 - 1$, if whenever $f(x) = f(x_1)$, then $x = x_1$. If f is one to one, the inverse function, f^{-1} is defined on $f(D(f))$ and $f^{-1}(y) = x$ where $f(x) = y$. Thus from the definition, $f^{-1}(f(x)) = x$ for all $x \in D(f)$ and $f(f^{-1}(y)) = y$ for all $y \in f(D(f))$. Defining id by $\text{id}(z) \equiv z$ this says $f \circ f^{-1} = \text{id}$ and $f^{-1} \circ f = \text{id}$.

Polynomials and rational functions are particularly easy functions to understand because they do come from a simple formula.

Definition 2.1.12 *A function f is a polynomial if*

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

where the a_i are real numbers and n is a nonnegative integer. In this case the degree of the polynomial, $f(x)$ is n . Thus the degree of a polynomial is the largest exponent appearing on the variable.

f is a rational function if

$$f(x) = \frac{h(x)}{g(x)}$$

where h and g are polynomials.

For example, $f(x) = 3x^5 + 9x^2 + 7x + 5$ is a polynomial of degree 5 and

$$\frac{3x^5 + 9x^2 + 7x + 5}{x^4 + 3x + x + 1}$$

is a rational function.

Note that in the case of a rational function, the domain of the function might not be all of $\mathbb{R}$. For example, if

$$f(x) = \frac{x^2 + 8}{x + 1},$$

the domain of f would be all real numbers not equal to -1 .

Closely related to the definition of a function is the concept of the graph of a function.

Definition 2.1.13 *Given two sets, X and Y , the Cartesian product of the two sets, written as $X \times Y$, is assumed to be a set described as follows.*

$$X \times Y = \{(x, y) : x \in X \text{ and } y \in Y\}.$$

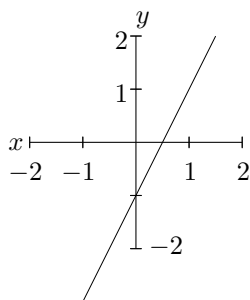
$\mathbb{R}^2$ denotes the Cartesian product of $\mathbb{R}$ with $\mathbb{R}$.

The notion of Cartesian product is just an abstraction of the concept of identifying a point in the plane with an ordered pair of numbers.

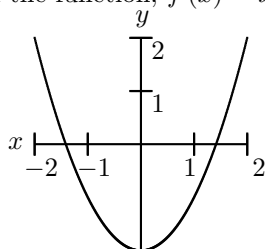
Definition 2.1.14 *Let $f : D(f) \rightarrow R(f)$ be a function. The graph of f consists of the set,*

$$\{(x, y) : y = f(x) \text{ for } x \in D(f)\}.$$

Note that knowledge of the graph of a function is equivalent to knowledge of the function. To find $f(x)$, simply observe the ordered pair which has x as its first element and the value of y equals $f(x)$. The graph of f can be represented by drawing a picture. For example, consider the picture of a part of the graph of the function $f(x) = 2x - 1$.



Here is the graph of the function, $f(x) = x^2 - 2$



Definition 2.1.15 A function whose domain is defined as a set of the form

$$\{k, k+1, k+2, \dots\}$$

for k an integer is known as a sequence. Thus you can consider $f(k)$, $f(k+1)$, $f(k+2)$, etc. Usually the domain of the sequence is either $\mathbb{N}$, the natural numbers consisting of $\{1, 2, 3, \dots\}$ or the nonnegative integers, $\{0, 1, 2, 3, \dots\}$. Also, it is traditional to write f_1, f_2 , etc. instead of $f(1)$, $f(2)$, $f(3)$ etc. when referring to sequences. In the above context, f_k is called the first term, f_{k+1} the second and so forth. It is also common to write the sequence, not as f but as $\{f_i\}_{i=k}^{\infty}$ or just $\{f_i\}$ for short.

Example 2.1.16 Let $\{a_k\}_{k=1}^{\infty}$ be defined by $a_k \equiv k^2 + 1$.

This gives a sequence. In fact, $a_7 = a(7) = 7^2 + 1 = 50$ just from using the formula for the k^{th} term of the sequence.

It is nice when sequences come in this way from a formula for the k^{th} term. However, this is often not the case. Sometimes sequences are defined recursively. This happens, when the first several terms of the sequence are given and then a rule is specified which determines a_{n+1} from knowledge of $a_1, \dots, a_n$. This rule which specifies a_{n+1} from knowledge of a_k for $k \leq n$ is known as a recurrence relation.

Example 2.1.17 Let $a_1 = 1$ and $a_2 = 1$. Assuming $a_1, \dots, a_{n+1}$ are known, $a_{n+2} \equiv a_n + a_{n+1}$.

Thus the first several terms of this sequence, listed in order, are 1, 1, 2, 3, 5, 8, $\dots$. This particular sequence is called the Fibonacci sequence and is important in the study of reproducing rabbits. Note this defines a function without giving a formula for it. Such sequences occur naturally in the solution of differential equations using power series methods and in many other situations of great importance.

For sequences, it is very important to consider something called a subsequence.

Definition 2.1.18 Let $\{a_n\}$ be a sequence and let $n_1 < n_2 < n_3, \dots$ be any strictly increasing list of integers such that n_1 is at least as large as the first number in the domain of the function defining the sequence. Then if $b_k \equiv a_{n_k}$, $\{b_k\}$ is called a subsequence of $\{a_n\}$.

For example, suppose $a_n = (n^2 + 1)$. Thus $a_1 = 2$, $a_3 = 10$, etc. If

$$n_1 = 1, n_2 = 3, n_3 = 5, \dots, n_k = 2k - 1,$$

then letting $b_k = a_{n_k}$, it follows

$$b_k = ((2k - 1)^2 + 1) = 4k^2 - 4k + 2.$$

2.2 Exercises

1. Let $g(t) \equiv \sqrt{2-t}$ and let $f(t) = \frac{1}{t}$. Find $g \circ f$. Include the domain of $g \circ f$.

2. Give the domains of the following functions.

(a) $f(x) = \frac{x+3}{3x-2}$

(b) $f(x) = \sqrt{x^2 - 4}$

(c) $f(x) = \sqrt{4 - x^2}$

(d) $f(x) = \sqrt{\frac{x-4}{3x+5}}$

(e) $f(x) = \sqrt{\frac{x^2-4}{x+1}}$

3. Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be defined by $f(t) \equiv t^3 + 1$. Is f one to one? Can you find a formula for f^{-1} ?

4. Suppose $a_1 = 1$, $a_2 = 3$, and $a_3 = -1$. Suppose also that for $n \geq 4$ it is known that $a_n = a_{n-1} + 2a_{n-2} + 3a_{n-3}$. Find a_4 . Are you able to guess a formula for the k^{th} term of this sequence?

5. Let $f: \{t \in \mathbb{R} : t \neq -1\} \rightarrow \mathbb{R}$ be defined by $f(t) \equiv \frac{t}{t+1}$. Find f^{-1} if possible.

6. A function, $f: \mathbb{R} \rightarrow \mathbb{R}$ is a strictly increasing function if whenever $x < y$, it follows that $f(x) < f(y)$. If f is a strictly increasing function, does f^{-1} always exist? Explain your answer.

7. Let $f(t)$ be defined by

$$f(t) = \begin{cases} 2t+1 & \text{if } t \leq 1 \\ t & \text{if } t > 1 \end{cases}.$$

Find f^{-1} if possible.

8. Suppose $f: D(f) \rightarrow R(f)$ is one to one, $R(f) \subseteq D(g)$, and $g: D(g) \rightarrow R(g)$ is one to one. Does it follow that $g \circ f$ is one to one?

9. If $f: \mathbb{R} \rightarrow \mathbb{R}$ and $g: \mathbb{R} \rightarrow \mathbb{R}$ are two one to one functions, which of the following are necessarily one to one on their domains? Explain why or why not by giving a proof or an example.

(a) $f + g$

(b) fg

(c) f^3

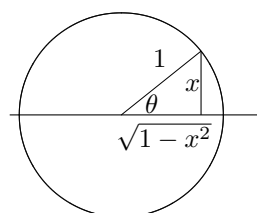
(d) f/g

10. Draw the graph of the function $f(x) = x^3 + 1$.

11. Draw the graph of the function $f(x) = x^2 + 2x + 2$.

12. Draw the graph of the function $f(x) = \frac{x}{1+x}$.

13. The function, $\sin$ has domain equal to $\mathbb{R}$ and range $[-1, 1]$. However, this function is not one to one because $\sin(x + 2\pi) = \sin x$. Show that if the domain of the function is restricted to be $[-\frac{\pi}{2}, \frac{\pi}{2}]$, then $\sin$ still maps onto $[-1, 1]$ but is now also one to one on this restricted domain. Therefore, there is an inverse function, called $\arcsin$ which is defined by $\arcsin(x) \equiv$ the angle whose $\sin$ is x which is in the interval, $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Thus $\arcsin(\frac{1}{2})$ is the angle whose $\sin$ is $\frac{1}{2}$ which is in $[-\frac{\pi}{2}, \frac{\pi}{2}]$. This angle is $\frac{\pi}{6}$. Suppose you wanted to find $\tan(\arcsin(x))$. How would you do it? Consider the following picture which corresponds to the case where $x > 0$.



Then letting $\theta = \arcsin(x)$, the thing which is wanted is $\tan \theta$. Now from the picture, you see this is $\frac{x}{\sqrt{1-x^2}}$. If x were negative, you would have the little triangle pointing down rather than up as in the picture. The result would be the same for $\tan \theta$. Find the following:

(a) $\cot(\arcsin(x))$

(b) $\sec(\arcsin(x))$

(c) $\csc(\arcsin(x))$

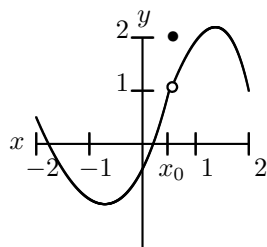
(d) $\cos(\arcsin(x))$

14. Using Problem 13 and the formulas for the trig functions of a sum of angles, find the following.
 - (a) $\cot(\arcsin(2x))$
 - (b) $\sec(\arcsin(x+y))$
 - (c) $\csc(\arcsin(x^2))$
 - (d) $\cos(2\arcsin(x))$
 - (e) $\tan(\arcsin(x) + \arcsin(y))$
15. The function, $\cos$, is onto $[-1, 1]$ but fails to be one to one. Show that if the domain of $\cos$ is restricted to be $[0, \pi]$, then $\cos$ is one to one on this restricted domain and still is onto $[-1, 1]$. Define $\arccos(x) \equiv$ the angle whose cosine is x which is in $[0, \pi]$. Find the following.
 - (a) $\tan(\arccos(x))$
 - (b) $\cot(\arccos(x))$
 - (c) $\sin(\arccos(x))$
 - (d) $\csc(\arccos(x))$
 - (e) $\sec(\arccos(x))$
16. Using Problem 15 and the formulas for the trig functions of a sum of angles, find the following.
 - (a) $\cot(\arccos(2x))$
 - (b) $\sec(\arccos(x+y))$
 - (c) $\csc(\arccos(x^2))$
 - (d) $\cos(2\arcsin(x) + \arcsin(y))$
17. The function, $\arctan$ is defined as $\arctan(x) \equiv$ the angle whose tangent is x which is in $(-\frac{\pi}{2}, \frac{\pi}{2})$. Show this is well defined and is the inverse function for $\tan$ if the domain of $\tan$ is restricted to be $(-\frac{\pi}{2}, \frac{\pi}{2})$. Find
 - (a) $\cos(\arctan(x))$
 - (b) $\cot(\arctan(x))$
 - (c) $\sin(\arctan(x))$
 - (d) $\csc(\arctan(x))$
 - (e) $\sec(\arctan(x))$
18. Using Problem 17 and the formulas for the trig functions of a sum of angles, find the following.
 - (a) $\cot(\arctan(2x))$
 - (b) $\sec(\arctan(x+y))$
 - (c) $\csc(\arccos(x^2))$
 - (d) $\cos(2\arctan(x) + \arcsin(y))$
19. Suppose $a_n = \frac{1}{n}$ and let $n_k = 2^k$. Find b_k where $b_k = a_{n_k}$.

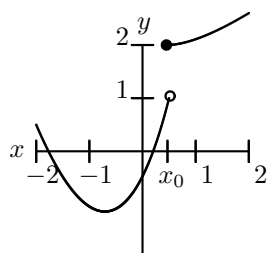
2.3 Continuous Functions

The concept of function is far too general to be useful in calculus. There are various ways to restrict the concept in order to study something interesting and the types of restrictions considered depend very much on what you find interesting. In Calculus, the most fundamental restriction made is to assume the functions are continuous. Continuous functions are those in which a sufficiently small change in x results in a small change in $f(x)$. They rule out things which could never happen physically. For example, it is not possible for a car to jump from one point to another instantly. Making this restriction precise turns out to be surprisingly difficult although many of the most important theorems about continuous functions seem intuitively clear.

Before giving the careful mathematical definitions, here are examples of graphs of functions which are not continuous at the point x_0 .



You see, there is a hole in the picture of the graph of this function and instead of filling in the hole with the appropriate value, $f(x_0)$ is too large. This is called a removable discontinuity because the problem can be fixed by redefining the function at the point x_0 . Here is another example.



You see from this picture that there is no way to get rid of the jump in the graph of this function by simply redefining the value of the function at x_0 . That is why it is called a nonremovable discontinuity or jump discontinuity. Now that pictures have been given of what it is desired to eliminate, it is time to give the precise definition.

The definition which follows, due to Cauchy¹ and Weierstrass² is the precise way to exclude the sort of behavior described above and all statements about continuous functions must ultimately rest on this definition from now on.

Definition 2.3.1 A function $f : D(f) \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is continuous at $x \in D(f)$ if for each $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $y \in D(f)$ and

$$|y - x| < \delta$$

it follows that

$$|f(x) - f(y)| < \varepsilon.$$

¹Augustin Louis Cauchy 1789-1857 was the son of a lawyer who was married to an aristocrat. He was born in France just after the fall of the Bastille and his family fled the reign of terror and hid in the countryside till it was over. Cauchy was educated at first by his father who taught him Greek and Latin. Eventually Cauchy learned many languages. He was also a good Catholic.

After the reign of terror, the family returned to Paris and Cauchy studied at the university to be an engineer but became a mathematician although he made fundamental contributions to physics and engineering. Cauchy was one of the most prolific mathematicians who ever lived. He wrote several hundred papers which fill 24 volumes. He also did research on many topics in mechanics and physics including elasticity, optics and astronomy. More than anyone else, Cauchy invented the subject of complex analysis. He is also credited with giving the first rigorous definition of continuity.

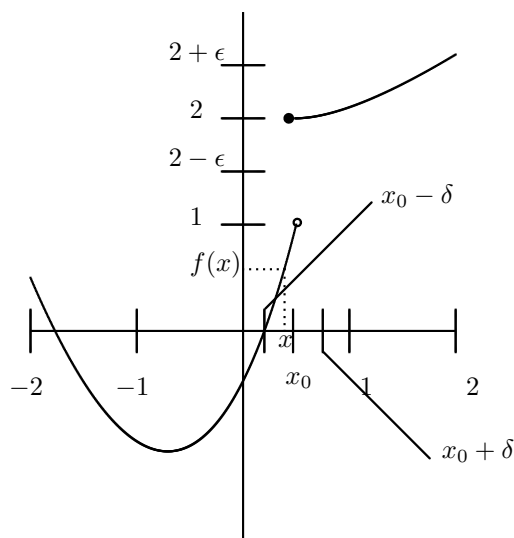
He married in 1818 and lived for 12 years with his wife and two daughters in Paris till the revolution of 1830. Cauchy refused to take the oath of allegiance to the new ruler and ended up leaving his family and going into exile for 8 years.

Notwithstanding his great achievements he was not known as a popular teacher.

²Wilhelm Theodor Weierstrass 1815-1897 brought calculus to essentially the state it is in now. When he was a secondary school teacher, he wrote a paper which was so profound that he was granted a doctor's degree. He made fundamental contributions to partial differential equations, complex analysis, calculus of variations, and many other topics. He also discovered some pathological examples such as space filling curves. Cauchy gave the definition in words and Weierstrass, somewhat later produced the totally rigorous $\varepsilon \delta$ definition presented here. The need for rigor in the subject of calculus was only realized over a long period of time.

A function, f is continuous if it is continuous at every point of $D(f)$.

In sloppy English this definition says roughly the following: A function, f is continuous at x when it is possible to make $f(y)$ as close as desired to $f(x)$ provided y is taken close enough to x . In fact this statement in words is pretty much the way Cauchy described it. The completely rigorous definition above is due to Weierstrass. This definition does indeed rule out the sorts of graphs drawn above. Consider the second nonremovable discontinuity. The removable discontinuity case is similar.



For the ε shown you can see from the picture that no matter how small you take δ , there will be points, x , between $x_0 - \delta$ and x_0 where $f(x) < 2 - \varepsilon$. In particular, for these values of x , $|f(x) - f(x_0)| > \varepsilon$. Therefore, the definition of continuity given above excludes the situation in which there is a jump in the function. Similar reasoning shows it excludes the removable discontinuity case as well. There are many ways a function can fail to be continuous and it is impossible to list them all by drawing pictures. This is why it is so important to use the definition. The other thing to notice is that the concept of continuity as described in the definition is a point property. That is to say it is a property which a function may or may not have at a single point. Here is an example.

Example 2.3.2 Let

$$f(x) = \begin{cases} x & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}.$$

This function is continuous at $x = 0$ and nowhere else.

To verify the assertion about the above function, first show it is not continuous at x if $x \neq 0$. Take such an x and let $\varepsilon = |x|/2$. Now let $\delta > 0$ be completely arbitrary. In the interval, $(x - \delta, x + \delta)$ there are rational numbers, y_1 such that $|y_1| > |x|$ and irrational numbers, y_2 . Thus $|f(y_1) - f(y_2)| = |y_1| > |x|$. If f were continuous at x , there would exist $\delta > 0$ such that for every point, $y \in (x - \delta, x + \delta)$, $|f(y) - f(x)| < \varepsilon$. But then, letting y_1 and y_2 be as just described,

$$\begin{aligned} |x| &< |y_1| = |f(y_1) - f(y_2)| \\ &\leq |f(y_1) - f(x)| + |f(x) - f(y_2)| < 2\varepsilon = |x|, \end{aligned}$$

which is a contradiction. Since a contradiction is obtained by assuming that f is continuous at x , it must be concluded that f is not continuous there. To see f is continuous at 0, let $\varepsilon > 0$ be given and let $\delta = \varepsilon$. Then if $|y - 0| < \delta = \varepsilon$, Then

$$\begin{aligned} |f(y) - f(0)| &= 0 \text{ if } y \text{ is irrational} \\ |f(y) - f(0)| &= |y| < \varepsilon \text{ if } y \text{ is rational.} \end{aligned}$$

either way, whenever $|y - 0| < \delta$, it follows $|f(y) - f(0)| < \varepsilon$ and so f is continuous at $x = 0$. How did I know to let $\delta = \varepsilon$? That is a very good question. The choice of δ for a particular ε is usually arrived at by using intuition, the actual $\varepsilon \delta$ argument reduces to a verification that the intuition was correct. Here is another example.

Example 2.3.3 Show the function, $f(x) = -5x + 10$ is continuous at $x = -3$.

To do this, note first that $f(-3) = 25$ and it is desired to verify the conditions for continuity. Consider the following.

$$|-5x + 10 - (25)| = 5|x - (-3)|.$$

This allows one to find a suitable δ . If $\varepsilon > 0$ is given, let $0 < \delta \leq \frac{1}{5}\varepsilon$. Then if $0 < |x - (-3)| < \delta$, it follows from this inequality that

$$|-5x + 10 - (25)| = 5|x - (-3)| < 5\frac{1}{5}\varepsilon = \varepsilon.$$

Sometimes the determination of δ in the verification of continuity can be a little more involved. Here is another example.

Example 2.3.4 Show the function, $f(x) = \sqrt{2x + 12}$ is continuous at $x = 5$.

First note $f(5) = \sqrt{22}$. Now consider:

$$\begin{aligned} \left| \sqrt{2x + 12} - \sqrt{22} \right| &= \left| \frac{2x + 12 - 22}{\sqrt{2x + 12} + \sqrt{22}} \right| \\ &= \frac{2}{\sqrt{2x + 12} + \sqrt{22}} |x - 5| \leq \frac{1}{11} \sqrt{22} |x - 5| \end{aligned}$$

whenever $|x - 5| < 1$ because for such x , $\sqrt{2x + 12} > 0$. Now let $\varepsilon > 0$ be given. Choose δ such that $0 < \delta \leq \min\left(1, \frac{\varepsilon\sqrt{22}}{2}\right)$. Then if $|x - 5| < \delta$, all the inequalities above hold and

$$\left| \sqrt{2x + 12} - \sqrt{22} \right| \leq \frac{2}{\sqrt{22}} |x - 5| < \frac{2}{\sqrt{22}} \frac{\varepsilon\sqrt{22}}{2} = \varepsilon.$$

Exercise 2.3.5 Show $f(x) = -3x^2 + 7$ is continuous at $x = 7$.

First observe $f(7) = -140$. Now

$$|-3x^2 + 7 - (-140)| = 3|x + 7||x - 7| \leq 3(|x| + 7)|x - 7|.$$

If $|x - 7| < 1$, it follows from the version of the triangle inequality which states $||s| - |t|| \leq |s - t|$ that $|x| < 1 + 7$. Therefore, if $|x - 7| < 1$, it follows that

$$\begin{aligned} |-3x^2 + 7 - (-140)| &\leq 3((1 + 7) + 7)|x - 7| \\ &= 3(1 + 27)|x - 7| = 84|x - 7|. \end{aligned}$$

Now let $\varepsilon > 0$ be given. Choose δ such that $0 < \delta \leq \min\left(1, \frac{\varepsilon}{84}\right)$. Then for $|x - 7| < \delta$, it follows

$$|-3x^2 + 7 - (-140)| \leq 84|x - 7| < 84\left(\frac{\varepsilon}{84}\right) = \varepsilon.$$

These ε δ proofs will not be emphasized any more than necessary. However, you should try a few of them because until you master this concept, you will not really understand calculus as it has been understood for approximately 150 years. The best you can do without this definition is to gain an understanding of the subject as it was understood by people in the 1700's, before the need for rigor was realized. Don't be discouraged by these historical observations. If you are able to master calculus as understood by Lagrange or Laplace³, you will have learned some very profound ideas even if they did originate in the eighteenth century, but it is better if you can understand the subject as Weierstrass did.

2.4 Sufficient Conditions For Continuity

The next theorem is a fundamental result which is convenient for avoiding the ε δ definition of continuity.

Theorem 2.4.1 *The following assertions are valid for f, g functions and a, b numbers.*

1. *The function, $af + bg$ is continuous at x when f, g are continuous at $x \in D(f) \cap D(g)$ and $a, b \in \mathbb{R}$.*
2. *If f and g are each real valued functions continuous at x , then fg is continuous at x . If, in addition to this, $g(x) \neq 0$, then f/g is continuous at x .*
3. *If f is continuous at x , $f(x) \in D(g) \subseteq \mathbb{R}$, and g is continuous at $f(x)$, then $g \circ f$ is continuous at x .*
4. *The function $f: \mathbb{R} \rightarrow \mathbb{R}$, given by $f(x) = |x|$ is continuous.*

The proof of this theorem is in the last section of this chapter but its conclusions are not surprising. For example the first claim says that $(af + bg)(y)$ is close to $(af + bg)(x)$ when y is close to x provided the same can be said about f and g . For the third claim, continuity of f indicates that if y is close enough to x then $f(x)$ is close to $f(y)$ and so by continuity of g at $f(x)$, $g(f(y))$ is close to $g(f(x))$. The fourth claim is verified as follows.

$$|x| = |x - y + y| \leq |x - y| + |y|$$

and so

$$|x| - |y| \leq |x - y|.$$

Similarly,

$$|y| - |x| \leq |x - y|.$$

Therefore,

$$||x| - |y|| \leq |x - y|.$$

³Lagrange and Laplace were two great physicists of the 1700's. They made fundamental contributions to the calculus of variations and to mechanics and astronomy. Lagrange is considered one of the inventors of calculus of variations. He also did work in many areas of mathematics and at one time was interested in botany.

It follows that if $\varepsilon > 0$ is given, one can take $\delta = \varepsilon$ and obtain that for $|x - y| < \delta = \varepsilon$,

$$||x| - |y|| \leq |x - y| < \delta = \varepsilon$$

which shows continuity of the function, $f(x) = |x|$.

2.5 Continuity Of Circular Functions

The functions $\sin x$ and $\cos x$ are often called the circular functions. This is because for each $x \in \mathbb{R}$, $(\cos x, \sin x)$ is a point on the unit circle. You can skip the proof of the following theorem if you wish. This continuity is easily seen without the proof by simply referring to the geometric description of these functions as the coordinates of a point on the unit circle. The graphs of these functions are as shown below. The first is the graph of $\sin$ and the second is the graph of $\cos$



Theorem 2.5.1 *The functions, $\cos$ and $\sin$ are continuous.*

Proof: First it will be shown that $\cos$ and $\sin$ are continuous at 0. By Corollary 1.4.4 on Page 24 the following inequality is valid for small positive values of θ .

$$1 - \cos \theta + \sin \theta \geq \theta \geq \sin \theta.$$

It follows that for θ small and positive, $|\theta| \geq |\sin \theta| = \sin \theta$. If $\theta < 0$, then $-\theta = |\theta| > 0$ and $-\theta \geq \sin(-\theta)$. But then this means $|\sin \theta| = -\sin \theta = \sin(-\theta) \leq -\theta = |\theta|$ in this case also. Therefore, whenever $|\theta|$ is small enough,

$$|\theta| \geq |\sin \theta|.$$

Now let $\varepsilon > 0$ be given and take $\delta = \varepsilon$. Then if $|\theta| < \delta$, it follows

$$|\sin \theta - 0| = |\sin \theta - \sin 0| = |\sin \theta| \leq |\theta| < \delta = \varepsilon,$$

showing $\sin$ is continuous at 0.

Next, note that for $|\theta| < \pi/2$, $\cos \theta \geq 0$ and so for such θ ,

$$\sin^2 \theta \geq \frac{\sin^2 \theta}{1 + \cos \theta} = \frac{1 - \cos^2 \theta}{1 + \cos \theta} = 1 - \cos \theta \geq 0. \quad (2.2)$$

From the first part of this argument for $\sin$, given $\varepsilon > 0$ there exists $\delta > 0$ such that if $|\theta| < \delta$, then $|\sin \theta| < \sqrt{\varepsilon}$. It follows from 2.2 that if $|\theta| < \delta$, then $\varepsilon > 1 - \cos \theta \geq 0$. This proves these functions are continuous at 0. Now $y = (y - x) + x$ and so

$$\cos y = \cos(y - x) \cos x - \sin(y - x) \sin x.$$

Therefore,

$$\cos y - \cos x = \cos(y - x) \cos x - \sin(y - x) \sin x - \cos x$$

and so, since $|\cos x|, |\sin x| \leq 1$,

$$\begin{aligned} |\cos y - \cos x| &\leq |\cos x (\cos(y - x) - 1)| + |\sin x| |\sin(y - x)| \\ &\leq |\cos(y - x) - 1| + |\sin(y - x)|. \end{aligned}$$

From the first part of this theorem, if $|y - x|$ is sufficiently small, both of these last two terms are less than $\varepsilon/2$ and this proves $\cos$ is continuous at x . The proof that $\sin$ is continuous is left for you to verify.

2.6 Exercises

1. Let $f(x) = 2x + 7$. Show f is continuous at every point x . **Hint:** You need to let $\varepsilon > 0$ be given. In this case, you should try $\delta \leq \varepsilon/2$. Note that if one δ works in the definition, then so does any smaller δ .

2. Let $f(x) = x^2 + 1$. Show f is continuous at $x = 3$. **Hint:**

$$\begin{aligned} |f(x) - f(3)| &= |x^2 + 1 - (9 + 1)| \\ &= |x + 3||x - 3|. \end{aligned}$$

Thus if $|x - 3| < 1$, it follows from the triangle inequality, $|x| < 1 + 3 = 4$ and so

$$|f(x) - f(3)| < 4|x - 3|.$$

Now try to complete the argument by letting $\delta \leq \min(1, \varepsilon/4)$. The symbol, $\min$ means to take the minimum of the two numbers in the parenthesis.

3. Let $f(x) = x^2 + 1$. Show f is continuous at $x = 4$.
4. Let $f(x) = 2x^2 + 1$. Show f is continuous at $x = 1$.
5. Let $f(x) = x^2 + 2x$. Show f is continuous at $x = 2$. Then show it is continuous at every point.
6. Let $f(x) = |2x + 3|$. Show f is continuous at every point. **Hint:** Review the two versions of the triangle inequality for absolute values.
7. Let $f(x) = \frac{1}{x^2 + 1}$. Show f is continuous at every value of x .
8. Show $\sin$ is continuous.
9. Let $f(x) = \sqrt{x}$ show f is continuous at every value of x in its domain. **Hint:** You might want to make use of the identity,

$$\sqrt{x} - \sqrt{y} = \frac{x - y}{\sqrt{x} + \sqrt{y}}$$

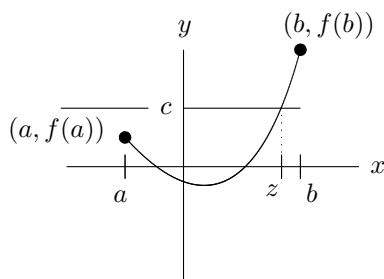
at some point in your argument.

10. Using Theorem 2.4.1, show all polynomials are continuous and that a rational function is continuous at every point of its domain. **Hint:** First show the function given as $f(x) = x$ is continuous and then use the Theorem 2.4.1.
11. Let $f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$ and consider $g(x) = f(x) \sin x$. Determine where g is continuous and explain your answer.
12. Suppose f is any function whose domain is the integers. Thus $D(f) = \mathbb{Z}$, the set of whole numbers, $\dots, -3, -2, -1, 0, 1, 2, 3, \dots$. Then f is continuous. Why? **Hint:** In the definition of continuity, what if you let $\delta = \frac{1}{4}$? Would this δ work for a given $\varepsilon > 0$? This shows that the idea that a continuous function is one for which you can draw the graph without taking the pencil off the paper is a lot of nonsense.
13. Describe the points where $\tan$, $\cot$, $\sec$, and $\csc$ are continuous.
14. Give an example of a function, f which is not continuous at some point but $|f|$ is continuous at that point.

15. Find two functions which fail to be continuous but whose product is continuous.
16. Find two functions which fail to be continuous but whose sum is continuous.
17. Find two functions which fail to be continuous but whose quotient is continuous.
18. Where is the function, $\sin(\tan(x))$ continuous? What is its domain?
19. Where is the function, $\tan(\sin(x))$ continuous? What is its domain?

2.7 Properties Of Continuous Functions

Continuous functions have many important properties which are consequences of the completeness axiom. Proofs of these theorems are in the last section at the end of this chapter. The next theorem is called the intermediate value theorem and the following picture illustrates its conclusion. It gives the existence of a certain point.



You see in the picture there is a horizontal line, $y = c$ and a continuous function which starts off less than c at the point a and ends up greater than c at point b . The intermediate value theorem says there is some point between a and b shown in the picture as z such that the value of the function at this point equals c . It may seem this is obvious but without completeness the conclusion of the theorem cannot be drawn. Nevertheless, the above picture makes this theorem very easy to believe.

Theorem 2.7.1 Suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous and suppose $f(a) < c < f(b)$. Then there exists $z \in (a, b)$ such that $f(z) = c$.

Example 2.7.2 Does there exist a solution to the equation $\sqrt{x^4 + 7} - x^3 \sin x = 0$?

By Theorem 2.4.1 and Problem 9 on Page 55 it follows easily that the function, f , given by $f(x) = \sqrt{x^4 + 7} - x^3 \sin x$ is continuous. Also, $f(0) = \sqrt{7} > 0$ while

$$f\left(\frac{5\pi}{2}\right) = \sqrt{\left(\frac{5\pi}{2}\right)^4 + 7} - \left(\frac{5\pi}{2}\right)^3 \sin\left(\frac{5\pi}{2}\right)$$

which is approximately equal to $-422.731\,331\,831\,631\,6 < 0$. Therefore, by the intermediate value theorem there must exist $x \in (0, \frac{5\pi}{2})$ such that $f(x) = 0$.

This example illustrates the use of this major theorem very well. It says something exists but it does not tell how to find it.

Definition 2.7.3 A function, f , defined on some interval is strictly increasing if whenever $x < y$, it follows $f(x) < f(y)$. The function is strictly decreasing if whenever $x < y$, it follows $f(x) > f(y)$.

You should draw a picture of the graph of a strictly increasing or decreasing function from the definition.

Lemma 2.7.4 *Let $\phi : [a, b] \rightarrow \mathbb{R}$ be a one to one continuous function. Then ϕ is either strictly increasing or strictly decreasing.*

This lemma is not real easy to prove but it is one of those things which seems obvious. To say a function is one to one is to say that every horizontal line intersects the graph of the function in no more than one point. (This is called the horizontal line test.) Now if your function is continuous (having no jumps) and is one to one, try to imagine how this could happen without it being either strictly increasing or decreasing and you will soon see this is highly believable and in fact, for it to fail would be incredible. The proof of this lemma is in the last section of this chapter in case you are interested. See Lemma 2.15.6.

Corollary 2.7.5 *Let $\phi : (a, b) \rightarrow \mathbb{R}$ be a one to one continuous function. Then ϕ is either strictly increasing or strictly decreasing.*

The proof of this corollary is the same as the proof of the lemma. The next corollary follows from the above.

Corollary 2.7.6 *Let $f : (a, b) \rightarrow \mathbb{R}$ be one to one and continuous. Then $f(a, b)$ is an open interval, (c, d) and $f^{-1} : (c, d) \rightarrow (a, b)$ is continuous. Also, if $f : [a, b] \rightarrow \mathbb{R}$ is one to one and continuous, then $f([a, b])$ is a closed interval, $[c, d]$ and $f^{-1} : [c, d] \rightarrow [a, b]$ is continuous.*

This corollary is not too surprising either. To view the graph of the inverse function, simply turn things on the side and switch x and y . If the original graph has no jumps in it, neither will the new graph. Of course, the concept of continuity is tied to a rigorous definition, not to the drawing of pictures. This is why there is a proof in the last section of this chapter.

The next theorem is known as the max min theorem or extreme value theorem. An easy proof is on Page 79.

Theorem 2.7.7 *Let $I = [a, b]$ and let $f : I \rightarrow \mathbb{R}$ be continuous. Then f achieves its maximum and its minimum on I . This means there exist, $x_1, x_2 \in I$ such that for all $x \in I$,*

$$f(x_1) \leq f(x) \leq f(x_2).$$

The proof of this theorem is based on the following lemma called the nested interval lemma which depends on the completeness axiom.

Lemma 2.7.8 *Let $I_k = [a^k, b^k]$ and suppose that for all $k = 1, 2, \dots$,*

$$I_k \supseteq I_{k+1}.$$

Then there exists a point which is in each of these intervals.

2.8 Exercises

1. Give an example of a continuous function defined on $(0, 1)$ which does not achieve its maximum on $(0, 1)$. It suffices to draw a graph of this function.
2. Give an example of a continuous function defined on $(0, 1)$ which is bounded but which does not achieve either its maximum or its minimum. It suffices to draw a graph of this function.

3. Give an example of a discontinuous function defined on $[0, 1]$ which is bounded but does not achieve either its maximum or its minimum. It suffices to draw a graph of this function.
4. Give an example of a continuous function defined on $[0, 1) \cup (1, 2]$ which is positive at 2, negative at 0 but is not equal to zero for any value of x . It suffices to draw a graph of this function.
5. Let $f(x) = x^5 + ax^4 + bx^3 + cx^2 + dx + e$ where a, b, c, d , and e are numbers. Show there exists x such that $f(x) = 0$.
6. Give an example of a function which is one to one but neither strictly increasing nor strictly decreasing. **Hint:** Look for discontinuous functions satisfying the horizontal line test. It suffices to draw a graph of this function.
7. Do you believe in $\sqrt[7]{8}$? That is, does there exist a number which multiplied by itself seven times yields 8? Before you jump to any conclusions, the number you get on your calculator is wrong. In fact, your calculator does not even know about $\sqrt[7]{8}$. All it can do is try to approximate it and what it gives you is this approximation. Why does it exist? **Hint:** Use the intermediate value theorem on the function, $f(x) = x^7 - 8$.
8. It has been known since the time of Pythagoras that $\sqrt{2}$ is irrational. If you throw out all the irrational numbers, show that the conclusion of the intermediate value theorem could no longer be obtained. That is, show there exists a function which starts off less than zero and ends up larger than zero and yet there is no number where the function equals zero. **Hint:** Try $f(x) = x^2 - 2$. You supply the details.
9. Let f be a continuous function defined on a closed interval, $I_1 \equiv [a, b]$ such that $f(a) < 0$ and $f(b) > 0$. Consider $\frac{a+b}{2}$. If $f(\frac{a+b}{2}) \geq 0$, let $I_2 = [a, \frac{a+b}{2}]$ and if $f(\frac{a+b}{2}) < 0$, let $I_2 \equiv [\frac{a+b}{2}, b]$. Thus $I_1 \supset I_2$ and the interval I_2 has exactly the same property that I_1 had in terms of f being negative at the left endpoint and nonnegative at the right endpoint. Continue this way obtaining a sequence of nested closed and bounded intervals, $\{I_k\}$. From the nested interval lemma on Page 57, there exists a unique point x in all these intervals. Show $f(x) = 0$. This is called the method of bisection and can be used to find a solution to the equation, $f(x) = 0$.
10. Apply the method of bisection described in Problem 9 to find $\sqrt[7]{8}$. Use a calculator to raise things to the seventh power. It will be much easier than doing it by hand.
11. A circular hula hoop lies partly in the shade and partly in the hot sun. Show there exist two points on the hula hoop which are at opposite sides of the hoop which have the same temperature. **Hint:** Imagine this is a circle and points are located by specifying their angle, θ from a fixed diameter. Then letting $T(\theta)$ be the temperature in the hoop, $T(\theta + 2\pi) = T(\theta)$. You need to have $T(\theta) = T(\theta + \pi)$ for some θ . Assume T is a continuous function of θ .
12. A car starts off on a long trip with a full tank of gas. The driver intends to drive the car till it runs out of gas. Show that at some time the number of miles the car has gone exactly equals the number of gallons of gas in the tank.

2.9 Limit Of A Function

A concept closely related to continuity is that of the limit of a function.

Definition 2.9.1 Let $f : D(f) \subseteq \mathbb{R} \rightarrow \mathbb{R}$ be a function where $D(f) \supseteq (x-r, x) \cup (x, x+r)$ for some $r > 0$. Note that f is not necessarily defined at x . Then

$$\lim_{y \rightarrow x} f(y) = L$$

if and only if the following condition holds. For all $\varepsilon > 0$ there exists $\delta > 0$ such that if

$$0 < |y - x| < \delta,$$

then,

$$|L - f(y)| < \varepsilon.$$

If everything is the same as the above, except y is required to be larger than x and f is only required to be defined on $(x, x+r)$, then the notation is

$$\lim_{y \rightarrow x+} f(y) = L.$$

If f is only required to be defined on $(x-r, x)$ and y is required to be less than x , with the same conditions above, we write

$$\lim_{y \rightarrow x-} f(y) = L.$$

Limits are also taken as a variable “approaches” infinity. Of course nothing is “close” to infinity and so this requires a slightly different definition.

$$\lim_{x \rightarrow \infty} f(x) = L$$

if for every $\varepsilon > 0$ there exists l such that whenever $x > l$,

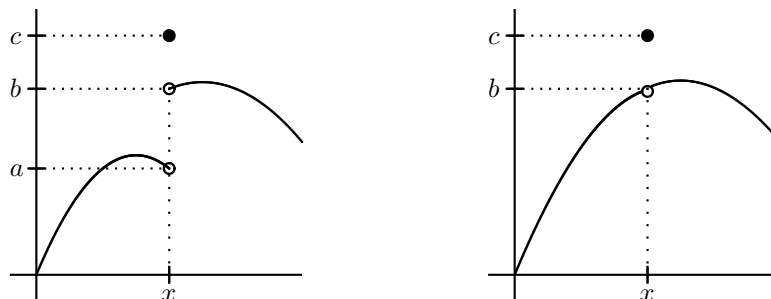
$$|f(x) - L| < \varepsilon \tag{2.3}$$

and

$$\lim_{x \rightarrow -\infty} f(x) = L$$

if for every $\varepsilon > 0$ there exists l such that whenever $x < l$, 2.3 holds.

The following pictures illustrate some of these definitions.



In the left picture is shown the graph of a function. Note the value of the function at x equals c while $\lim_{y \rightarrow x+} f(y) = b$ and $\lim_{y \rightarrow x-} f(y) = a$. In the second picture, $\lim_{y \rightarrow x} f(y) = b$. Note that the value of the function at the point x has nothing to do with the limit of the function in any of these cases. **The value of a function at x has nothing to do with the value of the limit at x !** This must always be kept in mind. You do not evaluate interesting limits by computing $f(x)$! In the above picture, $f(x)$ is always wrong! It may be the case that $f(x)$ is right but this is merely a happy coincidence when it occurs and as explained below in Theorem 2.9.6, this is often equivalent to f being continuous at x .

Theorem 2.9.2 *If $\lim_{y \rightarrow x} f(y) = L$ and $\lim_{y \rightarrow x} f(y) = L_1$, then $L = L_1$.*

Proof: Let $\varepsilon > 0$ be given. There exists $\delta > 0$ such that if $0 < |y - x| < \delta$, then

$$|f(y) - L| < \varepsilon, \quad |f(y) - L_1| < \varepsilon.$$

Therefore, for such y ,

$$|L - L_1| \leq |L - f(y)| + |f(y) - L_1| < \varepsilon + \varepsilon = 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this shows $L = L_1$.

The above theorem holds for any of the kinds of limits presented in the above definition.

Another concept is that of a function having either ∞ or $-\infty$ as a limit. In this case, the values of the function do not ever get close to their target because nothing can be close to $\pm\infty$. Roughly speaking, the limit of the function equals ∞ if the values of the function are ultimately larger than any given number. More precisely:

Definition 2.9.3 *If $f(x) \in \mathbb{R}$, then $\lim_{y \rightarrow x} f(y) = \infty$ if for every number l , there exists $\delta > 0$ such that whenever $|y - x| < \delta$, then $f(y) > l$. $\lim_{x \rightarrow \infty} f(x) = \infty$ if for all k , there exists l such that $f(x) > k$ whenever $x > l$. One sided limits and limits as the variable approaches $-\infty$, are defined similarly.*

It may seem there is a lot to memorize here. In fact, this is not so because all the definitions are intuitive when you understand them.

The proof of the following theorem is on Page 2.15.2. This is one of those messy theorems which is intuitively fairly easy to believe.

Theorem 2.9.4 *In this theorem, the symbol, $\lim_{y \rightarrow x}$ denotes any of the limits described above. Suppose $\lim_{y \rightarrow x} f(y) = L$ and $\lim_{y \rightarrow x} g(y) = K$ where K and L are real numbers in $\mathbb{R}$. Then if $a, b \in \mathbb{R}$,*

$$\lim_{y \rightarrow x} (af(y) + bg(y)) = aL + bK, \quad (2.4)$$

$$\lim_{y \rightarrow x} fg(y) = LK \quad (2.5)$$

and if $K \neq 0$,

$$\lim_{y \rightarrow x} \frac{f(y)}{g(y)} = \frac{L}{K}. \quad (2.6)$$

Also, if h is a continuous function defined near L , then

$$\lim_{y \rightarrow x} h \circ f(y) = h(L). \quad (2.7)$$

Suppose $\lim_{y \rightarrow x} f(y) = L$. If $f(y) \leq a$ all y of interest, then $L \leq a$ and if $f(y) \geq a$ then $L \geq a$.

A very useful theorem for finding limits is called the squeezing theorem. Imagine two men are walking toward a light post and a third man is right between them. Then the third man is also walking toward the light post. That is the idea of this theorem.

Theorem 2.9.5 *Suppose $\lim_{x \rightarrow a} f(x) = L = \lim_{x \rightarrow a} g(x)$ and for all x near a ,*

$$f(x) \leq h(x) \leq g(x).$$

Then

$$\lim_{x \rightarrow a} h(x) = L.$$

Proof: If $L \geq h(x)$, then

$$|h(x) - L| \leq |f(x) - L|.$$

If $L < h(x)$, then

$$|h(x) - L| \leq |g(x) - L|.$$

Therefore,

$$|h(x) - L| \leq |f(x) - L| + |g(x) - L|.$$

Now let $\varepsilon > 0$ be given. There exists δ_1 such that if $0 < |x - a| < \delta_1$,

$$|f(x) - L| < \varepsilon/2$$

and there exists δ_2 such that if $0 < |x - a| < \delta_2$, then

$$|g(x) - L| < \varepsilon/2.$$

Letting $0 < \delta \leq \min(\delta_1, \delta_2)$, if $0 < |x - a| < \delta$, then

$$\begin{aligned} |h(x) - L| &\leq |f(x) - L| + |g(x) - L| \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

This proves the theorem.

Theorem 2.9.6 For $f : I \rightarrow \mathbb{R}$, and I is an interval of the form (a, b) , $[a, b)$, $(a, b]$, or $[a, b]$, then f is continuous at $x \in I$ if and only if $\lim_{y \rightarrow x} f(y) = f(x)$.

Proof: You fill in the details. Compare the definition of continuous and the definition of the limit just given.

Example 2.9.7 Find $\lim_{x \rightarrow 3} \frac{x^2 - 9}{x - 3}$.

Note that $\frac{x^2 - 9}{x - 3} = x + 3$ whenever $x \neq 3$. Therefore, if $0 < |x - 3| < \varepsilon$,

$$\left| \frac{x^2 - 9}{x - 3} - 6 \right| = |x + 3 - 6| = |x - 3| < \varepsilon.$$

It follows from the definition that this limit equals 6.

You should be careful to note that in the definition of limit, the variable **never equals the thing it is getting close to**. In this example, x is never equal to 3. This is very significant because, in interesting limits, the function whose limit is being taken will usually not be defined at the point of interest.

Example 2.9.8 Let

$$f(x) = \frac{x^2 - 9}{x - 3} \text{ if } x \neq 3.$$

How should f be defined at $x = 3$ so that the resulting function will be continuous there?

In the previous example, the limit of this function equals 6. Therefore, by Theorem 2.9.6 it is necessary to define $f(3) \equiv 6$.

Example 2.9.9 Find $\lim_{x \rightarrow \infty} \frac{x}{1+x}$.

Write $\frac{x}{1+x} = \frac{1}{1+(1/x)}$. Now it seems clear that $\lim_{x \rightarrow \infty} 1 + (1/x) = 1 \neq 0$. Therefore, Theorem 2.9.4 implies

$$\lim_{x \rightarrow \infty} \frac{x}{1+x} = \lim_{x \rightarrow \infty} \frac{1}{1+(1/x)} = \frac{1}{1} = 1.$$

Example 2.9.10 Show $\lim_{x \rightarrow a} \sqrt{x} = \sqrt{a}$ whenever $a \geq 0$. In the case that $a = 0$, take the limit from the right.

There are two cases. First consider the case when $a > 0$. Let $\varepsilon > 0$ be given. Multiply and divide by $\sqrt{x} + \sqrt{a}$. This yields

$$|\sqrt{x} - \sqrt{a}| = \left| \frac{x - a}{\sqrt{x} + \sqrt{a}} \right|.$$

Now let $0 < \delta_1 < a/2$. Then if $|x - a| < \delta_1$, $x > a/2$ and so

$$\begin{aligned} |\sqrt{x} - \sqrt{a}| &= \left| \frac{x - a}{\sqrt{x} + \sqrt{a}} \right| \leq \frac{|x - a|}{(\sqrt{a}/\sqrt{2}) + \sqrt{a}} \\ &\leq \frac{2\sqrt{2}}{\sqrt{a}} |x - a|. \end{aligned}$$

Now let $0 < \delta \leq \min\left(\delta_1, \frac{\varepsilon\sqrt{a}}{2\sqrt{2}}\right)$. Then for $0 < |x - a| < \delta$,

$$|\sqrt{x} - \sqrt{a}| \leq \frac{2\sqrt{2}}{\sqrt{a}} |x - a| < \frac{2\sqrt{2}}{\sqrt{a}} \frac{\varepsilon\sqrt{a}}{2\sqrt{2}} = \varepsilon.$$

Next consider the case where $a = 0$. In this case, let $\varepsilon > 0$ and let $\delta = \varepsilon^2$. Then if $0 < x - 0 < \delta = \varepsilon^2$, it follows that $0 \leq \sqrt{x} < (\varepsilon^2)^{1/2} = \varepsilon$.

2.10 Exercises

1. Find the following limits if possible

(a) $\lim_{x \rightarrow 0+} \frac{|x|}{x}$

(b) $\lim_{x \rightarrow 0+} \frac{x}{|x|}$

(c) $\lim_{x \rightarrow 0-} \frac{|x|}{x}$

(d) $\lim_{x \rightarrow 4} \frac{x^2 - 16}{x + 4}$

(e) $\lim_{x \rightarrow 3} \frac{x^2 - 9}{x + 3}$

(f) $\lim_{x \rightarrow -2} \frac{x^2 - 4}{x - 2}$

(g) $\lim_{x \rightarrow \infty} \frac{x}{1 + x^2}$

(h) $\lim_{x \rightarrow \infty} -2 \frac{x}{1 + x^2}$

2. Find $\lim_{h \rightarrow 0} \frac{\frac{1}{(x+h)^3} - \frac{1}{x^3}}{h}$.

3. Find $\lim_{x \rightarrow 4} \frac{\sqrt[4]{x} - \sqrt{2}}{\sqrt{x} - 2}$.

4. Find $\lim_{x \rightarrow \infty} \frac{\sqrt[5]{3x} + \sqrt[4]{x} + 7\sqrt{x}}{\sqrt{3x+1}}$.

5. Find $\lim_{x \rightarrow \infty} \frac{(x-3)^{20}(2x+1)^{30}}{(2x^2+7)^{25}}$.

6. Find $\lim_{x \rightarrow 2} \frac{x^2 - 4}{x^3 + 3x^2 - 9x - 2}$.

7. Find

$$\lim_{x \rightarrow \infty} \left(\sqrt{1 - 7x + x^2} - \sqrt{1 + 7x + x^2} \right).$$

8. Prove Theorem 2.9.2 for right, left and limits as $y \rightarrow \infty$.

9. Prove from the definition that $\lim_{x \rightarrow a} \sqrt[3]{x} = \sqrt[3]{a}$ for all $a \in \mathbb{R}$.

Hint: You might want to use the formula for the difference of two cubes,

$$a^3 - b^3 = (a - b)(a^2 + ab + b^2).$$

10. Find $\lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h}$.

11. Prove Theorem 2.9.6 from the definitions of limit and continuity.

12. Find $\lim_{h \rightarrow 0} \frac{(x+h)^3 - x^3}{h}$.

13. Find $\lim_{h \rightarrow 0} \frac{\frac{1}{x+h} - \frac{1}{x}}{h}$.

14. Find $\lim_{x \rightarrow -3} \frac{x^3 + 27}{x + 3}$.

15. Find $\lim_{h \rightarrow 0} \frac{\sqrt{(3+h)^2} - 3}{h}$ if it exists.

16. Find the values of x for which $\lim_{h \rightarrow 0} \frac{\sqrt{(x+h)^2} - x}{h}$ exists and find the limit.

17. Find $\lim_{h \rightarrow 0} \frac{\sqrt[3]{(x+h)} - \sqrt[3]{x}}{h}$ if it exists. Here $x \neq 0$.
18. Suppose $\lim_{y \rightarrow x+} f(y) = L_1 \neq L_2 = \lim_{y \rightarrow x-} f(y)$. Show $\lim_{y \rightarrow x} f(x)$ does not exist. **Hint:** Roughly, the argument goes as follows: For $|y_1 - x|$ small and $y_1 > x$, $|f(y_1) - L_1|$ is small. Also, for $|y_2 - x|$ small and $y_2 < x$, $|f(y_2) - L_2|$ is small. However, if a limit existed, then $f(y_2)$ and $f(y_1)$ would both need to be close to some number and so both L_1 and L_2 would need to be close to some number. However, this is impossible because they are different.
19. Show $\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1$. **Hint:** You might consider Theorem 2.5.1 on Page 54 to write the inequality $|\sin x| + 1 - \cos x \geq |x| \geq |\sin x|$ whenever $|x|$ is small. Then divide both sides by $|\sin x|$ and use some trig. identities to write $\frac{\sin^2 x}{|\sin x|(1 + \cos x)} + 1 \geq \frac{|x|}{|\sin x|} \geq 1$ and then use squeezing theorem.
20. Let $f(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$. Find $\lim_{x \rightarrow 0} (\lim_{y \rightarrow 0} f(x, y))$ and $\lim_{y \rightarrow 0} (\lim_{x \rightarrow 0} f(x, y))$. If you did it right you got -1 for one answer and 1 for the other. What does this tell you about interchanging limits?

2.11 The Limit Of A Sequence

A closely related concept is the limit of a sequence. This was defined precisely by Bolzano⁴ a little before the definition of the limit. The following is the precise definition of what is meant by the limit of a sequence.

Definition 2.11.1 A sequence $\{a_n\}_{n=1}^{\infty}$ converges to a ,

$$\lim_{n \rightarrow \infty} a_n = a \text{ or } a_n \rightarrow a$$

if and only if for every $\varepsilon > 0$ there exists n_ε such that whenever $n \geq n_\varepsilon$,

$$|a_n - a| < \varepsilon.$$

In words the definition says that given any measure of closeness, ε , the terms of the sequence are eventually all this close to a . Note the similarity with the concept of limit. Here, the word “eventually” refers to n being sufficiently large. Earlier, it referred to y being sufficiently close to x on one side or another or else x being sufficiently large in either the positive or negative directions. The above definition is always the definition of what is meant by the limit of a sequence. If the a_n are complex numbers or later on, vectors the definition remains the same. If $a_n = x_n + iy_n$ and $a = x + iy$, $|a_n - a| = \sqrt{(x_n - x)^2 + (y_n - y)^2}$. Recall the way you measure distance between two complex numbers.

The next theorem says you can speak of **the** limit.

Theorem 2.11.2 If $\lim_{n \rightarrow \infty} a_n = a$ and $\lim_{n \rightarrow \infty} a_n = a_1$ then $a_1 = a$.

Proof: Suppose $a_1 \neq a$. Then let $0 < \varepsilon < |a_1 - a|/2$ in the definition of the limit. It follows there exists n_ε such that if $n \geq n_\varepsilon$, then $|a_n - a| < \varepsilon$ and $|a_n - a_1| < \varepsilon$. Therefore, for such n ,

$$\begin{aligned} |a_1 - a| &\leq |a_1 - a_n| + |a_n - a| \\ &< \varepsilon + \varepsilon < |a_1 - a|/2 + |a_1 - a|/2 = |a_1 - a|, \end{aligned}$$

⁴Bernhard Bolzano lived from 1781 to 1848. He was a Catholic priest and held a position in philosophy at the University of Prague. He had strong views about the absurdity of war, educational reform, and the need for individual conscience. His convictions got him in trouble with Franz I of Austria and when he refused to recant, was forced out of the university. He understood the need for absolute rigor in mathematics. He also did work on physics.

a contradiction.

Example 2.11.3 Let $a_n = \frac{1}{n^2+1}$.

Then it seems clear that

$$\lim_{n \rightarrow \infty} \frac{1}{n^2+1} = 0.$$

In fact, this is true from the definition. Let $\varepsilon > 0$ be given. Let $n_\varepsilon \geq \sqrt{\varepsilon^{-1}}$. Then if

$$n > n_\varepsilon \geq \sqrt{\varepsilon^{-1}},$$

it follows that $n^2 + 1 > \varepsilon^{-1}$ and so

$$0 < \frac{1}{n^2+1} = a_n < \varepsilon.$$

Thus $|a_n - 0| < \varepsilon$ whenever n is this large.

Note **the definition was of no use in finding a candidate for the limit**. This had to be produced based on other considerations. The definition is for verifying beyond any doubt that something is the limit. It is also what must be referred to in establishing theorems which are good for finding limits.

Example 2.11.4 Let $a_n = n^2$

Then in this case $\lim_{n \rightarrow \infty} a_n$ does not exist. Sometimes this situation is also referred to by saying $\lim_{n \rightarrow \infty} a_n = \infty$.

Example 2.11.5 Let $a_n = (-1)^n$.

In this case, $\lim_{n \rightarrow \infty} (-1)^n$ does not exist. This follows from the definition. Let $\varepsilon = 1/2$. If there exists a limit, l , then eventually, for all n large enough, $|a_n - l| < 1/2$. However, $|a_n - a_{n+1}| = 2$ and so,

$$2 = |a_n - a_{n+1}| \leq |a_n - l| + |l - a_{n+1}| < 1/2 + 1/2 = 1$$

which cannot hold. Therefore, there can be no limit for this sequence.

The following is another of those theorems which is intuitively fairly clear. A proof is on Page 74.

Theorem 2.11.6 Suppose $\{a_n\}$ and $\{b_n\}$ are sequences and that

$$\lim_{n \rightarrow \infty} a_n = a \text{ and } \lim_{n \rightarrow \infty} b_n = b.$$

Also suppose x and y are real numbers. Then

$$\lim_{n \rightarrow \infty} xa_n + yb_n = xa + yb \tag{2.8}$$

$$\lim_{n \rightarrow \infty} a_nb_n = ab \tag{2.9}$$

If $b \neq 0$,

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{a}{b}. \tag{2.10}$$

Another very useful theorem for finding limits is the squeezing theorem. As in the case of limits of a function, there is a squeezing theorem.

Theorem 2.11.7 Suppose $\lim_{n \rightarrow \infty} a_n = a = \lim_{n \rightarrow \infty} b_n$ and $a_n \leq c_n \leq b_n$ for all n large enough. Then $\lim_{n \rightarrow \infty} c_n = a$.

Proof: Let $\varepsilon > 0$ be given and let n_1 be large enough that if $n \geq n_1$,

$$|a_n - a| < \varepsilon/2 \text{ and } |b_n - a| < \varepsilon/2.$$

Then for such n ,

$$|c_n - a| \leq |a_n - a| + |b_n - a| < \varepsilon.$$

The reason for this is that if $c_n \geq a$, then

$$|c_n - a| = c_n - a \leq b_n - a \leq |a_n - a| + |b_n - a|$$

because $b_n \geq c_n$. On the other hand, if $c_n \leq a$, then

$$|c_n - a| = a - c_n \leq a - a_n \leq |a - a_n| + |b - b_n|.$$

This proves the theorem.

As an example, consider the following.

Example 2.11.8 Let

$$c_n \equiv (-1)^n \frac{1}{n}$$

and let $b_n = \frac{1}{n}$, and $a_n = -\frac{1}{n}$. Then you may easily show that

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = 0.$$

Since $a_n \leq c_n \leq b_n$, it follows $\lim_{n \rightarrow \infty} c_n = 0$ also.

Theorem 2.11.9 $\lim_{n \rightarrow \infty} r^n = 0$ whenever $|r| < 1$.

Proof: If $0 < r < 1$ it follows $r^{-1} > 1$. Why? Letting $\alpha = \frac{1}{r} - 1$, it follows

$$r = \frac{1}{1 + \alpha}.$$

Therefore, by the Example 1.3.3,

$$0 < r^n = \frac{1}{(1 + \alpha)^n} \leq \frac{1}{1 + \alpha n}.$$

Therefore, $\lim_{n \rightarrow \infty} r^n = 0$ if $0 < r < 1$. Now in general, if $|r| < 1$, $|r^n| = |r|^n \rightarrow 0$ by the first part. This proves the theorem.

An important theorem is the one which states that if a sequence converges, so does every subsequence. You should review Definition 2.1.18 on Page 47 at this point. Recall that if $\{x_n\}$ is a sequence and $n_1 < n_2 < \dots$ is an increasing sequence of integers with $n_1 \geq 1$, then $\{x_{n_k}\}_{k=1}^{\infty}$ is called a subsequence.

Theorem 2.11.10 Let $\{x_n\}$ be a sequence with $\lim_{n \rightarrow \infty} x_n = x$ and let $\{x_{n_k}\}$ be a subsequence. Then $\lim_{k \rightarrow \infty} x_{n_k} = x$.

Proof: Let $\varepsilon > 0$ be given. Then there exists n_ε such that if $n > n_\varepsilon$, then $|x_n - x| < \varepsilon$. Suppose $k > n_\varepsilon$. Then $n_k \geq k > n_\varepsilon$ and so

$$|x_{n_k} - x| < \varepsilon$$

showing $\lim_{k \rightarrow \infty} x_{n_k} = x$ as claimed.

2.11.1 Sequences And Completeness

You recall the definition of completeness which stated that every nonempty set of real numbers which is bounded above has a least upper bound and that every nonempty set of real numbers which is bounded below has a greatest lower bound and this is a property of the real line known as the completeness axiom. Geometrically, this involved filling in the holes. There is another way of describing completeness in terms of sequences which I believe is more useful than the least upper bound and greatest lower bound property.

Definition 2.11.11 $\{a_n\}$ is a Cauchy sequence if for all $\varepsilon > 0$, there exists n_ε such that whenever $n, m \geq n_\varepsilon$,

$$|a_n - a_m| < \varepsilon.$$

A sequence is Cauchy means the terms are “bunching up to each other” as m, n get large.

Theorem 2.11.12 The set of terms in a Cauchy sequence in $\mathbb{R}$ is bounded above and below.

Proof: Let $\varepsilon = 1$ in the definition of a Cauchy sequence and let $n > n_1$. Then from the definition,

$$|a_n - a_{n_1}| < 1.$$

It follows that for all $n > n_1$,

$$|a_n| < 1 + |a_{n_1}|.$$

Therefore, for all n ,

$$|a_n| \leq 1 + |a_{n_1}| + \sum_{k=1}^{n_1} |a_k|.$$

This proves the theorem.

Theorem 2.11.13 If a sequence $\{a_n\}$ in $\mathbb{R}$ converges, then the sequence is a Cauchy sequence.

Proof: Let $\varepsilon > 0$ be given and suppose $a_n \rightarrow a$. Then from the definition of convergence, there exists n_ε such that if $n > n_\varepsilon$, it follows that

$$|a_n - a| < \frac{\varepsilon}{2}$$

Therefore, if $m, n \geq n_\varepsilon + 1$, it follows that

$$|a_n - a_m| \leq |a_n - a| + |a - a_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

showing that, since $\varepsilon > 0$ is arbitrary, $\{a_n\}$ is a Cauchy sequence.

Definition 2.11.14 The sequence, $\{a_n\}$, is monotone increasing if for all n , $a_n \leq a_{n+1}$. The sequence is monotone decreasing if for all n , $a_n \geq a_{n+1}$.

If someone says a sequence is monotone, it usually means monotone increasing.

There exist different descriptions of the completeness axiom. If you like you can simply add the three new criteria in the following theorem to the list of things which you mean when you say $\mathbb{R}$ is complete and skip the proof. All versions of completeness involve the notion of filling in holes and they are really just different ways of expressing this idea.

In practice, it is often more convenient to use the first of the three equivalent versions of completeness in the following theorem which states that every Cauchy sequence converges. In fact, this version of completeness, although it is equivalent to the completeness axiom for the real line, also makes sense in many situations where Definition 1.2.1 on Page 18 does not make sense. For example, the concept of completeness is often needed in settings where there is no order. This happens as soon as one does multivariable calculus. From now on completeness will mean any of the three conditions in the following theorem.

It is the concept of completeness and the notion of limits which sets analysis apart from algebra. You will find that every existence theorem, a theorem which asserts the existence of something, in calculus depends on the assumption that something is complete. Here is the theorem. It is proved on Page 79.

Theorem 2.11.15 *The following conditions are equivalent to completeness.*

1. *Every Cauchy sequence converges*
2. *Every monotone increasing sequence which is bounded above converges.*
3. *Every monotone decreasing sequence which is bounded below converges.*

Theorem 2.11.16 *Let $\{a_n\}$ be a monotone increasing sequence which is bounded above. Then $\lim_{n \rightarrow \infty} a_n = \sup \{a_n : n \geq 1\}$*

Proof: Let $a = \sup \{a_n : n \geq 1\}$ and let $\varepsilon > 0$ be given. Then from Proposition 9.1.6 on Page 218 there exists m such that $a - \varepsilon < a_m \leq a$. Since the sequence is increasing, it follows that for all $n \geq m$, $a - \varepsilon < a_n \leq a$. Thus $a = \lim_{n \rightarrow \infty} a_n$.

2.11.2 Decimals

You are all familiar with decimals. In the United States these are written in the form $.a_1a_2a_3 \cdots$ where the a_i are integers between 0 and 9.⁵ Thus .23417432 is a number written as a decimal. You also recall the meaning of such notation in the case of a terminating decimal. For example, .234 is defined as $\frac{2}{10} + \frac{3}{10^2} + \frac{4}{10^3}$. Now what is meant by a nonterminating decimal?

Definition 2.11.17 *Let $.a_1a_2 \cdots$ be a decimal. Define*

$$.a_1a_2 \cdots \equiv \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{a_k}{10^k}.$$

Proposition 2.11.18 *The above definition makes sense.*

Proof: Note the sequence $\{\sum_{k=1}^n \frac{a_k}{10^k}\}_{n=1}^{\infty}$ is an increasing sequence. Therefore, if there exists an upper bound, it follows from Theorem 2.11.16 that this sequence converges and so the definition is well defined.

$$\sum_{k=1}^n \frac{a_k}{10^k} \leq \sum_{k=1}^n \frac{9}{10^k} = 9 \sum_{k=1}^n \frac{1}{10^k}.$$

⁵In France and Russia they use a comma instead of a period. This looks very strange but that is just the way they do it.

Now

$$\begin{aligned}\frac{9}{10} \left(\sum_{k=1}^n \frac{1}{10^k} \right) &= \sum_{k=1}^n \frac{1}{10^k} - \frac{1}{10} \sum_{k=1}^n \frac{1}{10^k} = \sum_{k=1}^n \frac{1}{10^k} - \sum_{k=2}^{n+1} \frac{1}{10^k} \\ &= \frac{1}{10} - \frac{1}{10^{n+1}}\end{aligned}$$

and so

$$\sum_{k=1}^n \frac{1}{10^k} \leq \frac{10}{9} \left(\frac{1}{10} - \frac{1}{10^{n+1}} \right) \leq \frac{10}{9} \left(\frac{1}{10} \right) = \frac{1}{9}.$$

Therefore, since this holds for all n , it follows the above sequence is bounded above. It follows the limit exists.

2.11.3 Continuity And The Limit Of A Sequence

There is a very useful way of thinking of continuity in terms of limits of sequences found in the following theorem. In words, it says a function is continuous if it takes convergent sequences to convergent sequences whenever possible.

Theorem 2.11.19 *A function $f : D(f) \rightarrow \mathbb{R}$ is continuous at $x \in D(f)$ if and only if, whenever $x_n \rightarrow x$ with $x_n \in D(f)$, it follows $f(x_n) \rightarrow f(x)$.*

Proof: Suppose first that f is continuous at x and let $x_n \rightarrow x$. Let $\varepsilon > 0$ be given. By continuity, there exists $\delta > 0$ such that if $|y - x| < \delta$, then $|f(y) - f(x)| < \varepsilon$. However, there exists n_δ such that if $n \geq n_\delta$, then $|x_n - x| < \delta$ and so for all n this large,

$$|f(x) - f(x_n)| < \varepsilon$$

which shows $f(x_n) \rightarrow f(x)$.

Now suppose the condition about taking convergent sequences to convergent sequences holds at x . Suppose f fails to be continuous at x . Then there exists $\varepsilon > 0$ and $x_n \in D(f)$ such that $|x - x_n| < \frac{1}{n}$, yet

$$|f(x) - f(x_n)| \geq \varepsilon.$$

But this is clearly a contradiction because, although $x_n \rightarrow x$, $f(x_n)$ fails to converge to $f(x)$. It follows f must be continuous after all. This proves the theorem.

2.12 Exercises

1. Find $\lim_{n \rightarrow \infty} \frac{n}{3n+4}$.
2. Find $\lim_{n \rightarrow \infty} \frac{3n^4+7n+1000}{n^4+1}$.
3. Find $\lim_{n \rightarrow \infty} \frac{2^n+7(5^n)}{4^n+2(5^n)}$.
4. Find $\lim_{n \rightarrow \infty} n \tan \frac{1}{n}$. **Hint:** See Problem 19 on Page 63.
5. Find $\lim_{n \rightarrow \infty} n \sin \frac{2}{n}$. **Hint:** See Problem 19 on Page 63.
6. Find $\lim_{n \rightarrow \infty} \sqrt{(n \sin \frac{9}{n})}$. **Hint:** See Problem 19 on Page 63.
7. Find $\lim_{n \rightarrow \infty} \sqrt{(n^2+6n)} - n$. **Hint:** Multiply and divide by $\sqrt{(n^2+6n)} + n$.

8. Find $\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{10^k}$.
9. Suppose $\{x_n + iy_n\}$ is a sequence of complex numbers which converges to the complex number $x + iy$. Show this happens if and only if $x_n \rightarrow x$ and $y_n \rightarrow y$.
10. For $|r| < 1$, find $\lim_{n \rightarrow \infty} \sum_{k=0}^n r^k$. **Hint:** First show $\sum_{k=0}^n r^k = \frac{r^{n+1}}{r-1} - \frac{1}{r-1}$. Then recall Theorem 2.11.9.
11. Suppose $x = .343434\overline{34}$ where the bar over the last 34 signifies that this repeats forever. In elementary school you were probably given the following procedure for finding the number x as a quotient of integers. First multiply by 100 to get $100x = 34.343434\overline{34}$ and then subtract to get $99x = 34$. From this you conclude that $x = 34/99$. Fully justify this procedure. **Hint:** $.343434\overline{34} = \lim_{n \rightarrow \infty} 34 \sum_{k=1}^n \left(\frac{1}{100}\right)^k$ now use Problem 10.
12. Suppose $D(f) = [0, 1] \cup \{9\}$ and $f(x) = x$ on $[0, 1]$ while $f(9) = 5$. Is f continuous at the point, 9? Use whichever definition of continuity you like.
13. Suppose $x_n \rightarrow x$ and $x_n \leq c$. Show that $x \leq c$. Also show that if $x_n \rightarrow x$ and $x_n \geq c$, then $x \geq c$. **Hint:** If this is not true, argue that for all n large enough $x_n > c$.
14. Let $a \in [0, 1]$. Show $a = .a_1 a_2 a_3 \cdots$ for a unique choice of integers, $a_1, a_2, \cdots$ if it is possible to do this. Otherwise, give an example.
15. Find $\lim_{n \rightarrow \infty} n \sin n$ if it exists. If it does not exist, explain why it does not.
16. Recall the axiom of completeness states that a set which is bounded above has a least upper bound and a set which is bounded below has a greatest lower bound. Show that a monotone decreasing sequence which is bounded below converges to its greatest lower bound. **Hint:** Let a denote the greatest lower bound and recall that because of this, it follows that for all $\varepsilon > 0$ there exist points of $\{a_n\}$ in $[a, a + \varepsilon)$.

2.13 Uniform Continuity*

There is a theorem about the integral of a continuous function which requires the notion of uniform continuity. This is discussed in this section. Consider the function $f(x) = \frac{1}{x}$ for $x \in (0, 1)$. This is a continuous function because, by Theorem 2.4.1, it is continuous at every point of $(0, 1)$. However, for a given $\varepsilon > 0$, the δ needed in the ε, δ definition of continuity becomes very small as x gets close to 0. The notion of uniform continuity involves being able to choose a single δ which works on the whole domain of f . Here is the definition.

Definition 2.13.1 *Let $f : D \subseteq \mathbb{R} \rightarrow \mathbb{R}$ be a function. Then f is uniformly continuous if for every $\varepsilon > 0$, there exists a δ **depending only on** ε such that if $|x - y| < \delta$ then $|f(x) - f(y)| < \varepsilon$.*

It is an amazing fact that under certain conditions continuity implies uniform continuity.

Definition 2.13.2 *A set, $K \subseteq \mathbb{R}$ is sequentially compact if whenever $\{a_n\} \subseteq K$ is a sequence, there exists a subsequence, $\{a_{n_k}\}$ such that this subsequence converges to a point of K .*

The following theorem is part of the Heine Borel theorem. Its proof is on Page 2.15.13.

Theorem 2.13.3 *Every closed interval, $[a, b]$ is sequentially compact.*

Theorem 2.13.4 *Let $f : K \rightarrow \mathbb{R}$ be continuous where K is a sequentially compact set in $\mathbb{R}$. Then f is uniformly continuous on K .*

Proof: If this is not true, there exists $\varepsilon > 0$ such that for every $\delta > 0$ there exists a pair of points, x_δ and y_δ such that even though $|x_\delta - y_\delta| < \delta$, $|f(x_\delta) - f(y_\delta)| \geq \varepsilon$. Taking a succession of values for δ equal to $1, 1/2, 1/3, \dots$, and letting the exceptional pair of points for $\delta = 1/n$ be denoted by x_n and y_n ,

$$|x_n - y_n| < \frac{1}{n}, |f(x_n) - f(y_n)| \geq \varepsilon.$$

Now since K is sequentially compact, there exists a subsequence, $\{x_{n_k}\}$ such that $x_{n_k} \rightarrow z \in K$. Now $n_k \geq k$ and so

$$|x_{n_k} - y_{n_k}| < \frac{1}{k}.$$

Consequently, $y_{n_k} \rightarrow z$ also. (x_{n_k} is like a person walking toward a certain point and y_{n_k} is like a dog on a leash which is constantly getting shorter. Obviously y_{n_k} must also move toward the point also. You should give a precise proof of what is needed here.) By continuity of f and Problem 13 on Page 69,

$$0 = |f(z) - f(z)| = \lim_{k \rightarrow \infty} |f(x_{n_k}) - f(y_{n_k})| \geq \varepsilon,$$

an obvious contradiction. Therefore, the theorem must be true.

The following corollary follows from this theorem and Theorem 2.13.3.

Corollary 2.13.5 *Suppose I is a closed interval, $I = [a, b]$ and $f : I \rightarrow \mathbb{R}$ is continuous. Then f is uniformly continuous.*

2.14 Exercises

1. A function, $f : D \subseteq \mathbb{R} \rightarrow \mathbb{R}$ is Lipschitz continuous or just Lipschitz for short if there exists a constant, K such that

$$|f(x) - f(y)| \leq K|x - y|$$

for all $x, y \in D$. Show every Lipschitz function is uniformly continuous.

2. If $|x_n - y_n| \rightarrow 0$ and $x_n \rightarrow z$, show that $y_n \rightarrow z$ also.
3. Consider $f : (1, \infty) \rightarrow \mathbb{R}$ given by $f(x) = \frac{1}{x}$. Show f is uniformly continuous even though the set on which f is defined is not sequentially compact.
4. If f is uniformly continuous, does it follow that $|f|$ is also uniformly continuous? If $|f|$ is uniformly continuous does it follow that f is uniformly continuous? Answer the same questions with “uniformly continuous” replaced with “continuous”. Explain why.
5. Suppose K is a sequentially compact set and $f : K \rightarrow \mathbb{R}$. Show that f achieves both its maximum and its minimum on K . **Hint:** Let $M \equiv \sup \{f(x) : x \in K\}$. Argue there exists a sequence, $\{x_n\} \subseteq K$ such that $f(x_n) \rightarrow M$. Now use sequential compactness to get a subsequence, $\{x_{n_k}\}$ such that $\lim_{k \rightarrow \infty} x_{n_k} = x \in K$ and use the continuity of f to verify that $f(x) = M$. Incidentally, this shows f is bounded on K as well. A similar argument works to give the part about achieving the minimum.

2.15 Fundamental Theory*

In this section, proofs of some theorems which have not been proved yet are given.

2.15.1 Combinations Of Functions And Sequences

Theorem 2.15.1 *The following assertions are valid*

1. *The function, $af + bg$ is continuous at x when f, g are continuous at $x \in D(f) \cap D(g)$ and $a, b \in \mathbb{R}$.*
2. *If f and g are each real valued functions continuous at x , then fg is continuous at x . If, in addition to this, $g(x) \neq 0$, then f/g is continuous at x .*
3. *If f is continuous at x , $f(x) \in D(g) \subseteq \mathbb{R}$, and g is continuous at $f(x)$, then $g \circ f$ is continuous at x .*
4. *The function $f: \mathbb{R} \rightarrow \mathbb{R}$, given by $f(x) = |x|$ is continuous.*

Proof: First consider 1.) Let $\varepsilon > 0$ be given. By assumption, there exist $\delta_1 > 0$ such that whenever $|x - y| < \delta_1$, it follows $|f(x) - f(y)| < \frac{\varepsilon}{2(|a|+|b|+1)}$ and there exists $\delta_2 > 0$ such that whenever $|x - y| < \delta_2$, it follows that $|g(x) - g(y)| < \frac{\varepsilon}{2(|a|+|b|+1)}$. Then let $0 < \delta \leq \min(\delta_1, \delta_2)$. If $|x - y| < \delta$, then everything happens at once. Therefore, using the triangle inequality

$$\begin{aligned} & |af(x) + bf(x) - (ag(y) + bg(y))| \\ & \leq |a||f(x) - f(y)| + |b||g(x) - g(y)| \\ & < |a| \left(\frac{\varepsilon}{2(|a|+|b|+1)} \right) + |b| \left(\frac{\varepsilon}{2(|a|+|b|+1)} \right) < \varepsilon. \end{aligned}$$

Now consider 2.) There exists $\delta_1 > 0$ such that if $|y - x| < \delta_1$, then $|f(x) - f(y)| < 1$. Therefore, for such y ,

$$|f(y)| < 1 + |f(x)|.$$

It follows that for such y ,

$$\begin{aligned} |fg(x) - fg(y)| & \leq |f(x)g(x) - g(x)f(y)| + |g(x)f(y) - f(y)g(y)| \\ & \leq |g(x)||f(x) - f(y)| + |f(y)||g(x) - g(y)| \\ & \leq (1 + |g(x)| + |f(y)|)[|g(x) - g(y)| + |f(x) - f(y)|] \\ & \leq (2 + |g(x)| + |f(x)|)[|g(x) - g(y)| + |f(x) - f(y)|] \end{aligned}$$

Now let $\varepsilon > 0$ be given. There exists δ_2 such that if $|x - y| < \delta_2$, then

$$|g(x) - g(y)| < \frac{\varepsilon}{2(2 + |g(x)| + |f(x)|)},$$

and there exists δ_3 such that if $|x - y| < \delta_3$, then

$$|f(x) - f(y)| < \frac{\varepsilon}{2(2 + |g(x)| + |f(x)|)}$$

Now let $0 < \delta \leq \min(\delta_1, \delta_2, \delta_3)$. Then if $|x - y| < \delta$, all the above hold at once and so

$$|fg(x) - fg(y)| \leq$$

$$(2 + |g(x)| + |f(x)|) [|g(x) - g(y)| + |f(x) - f(y)|] \\ < (2 + |g(x)| + |f(x)|) \left(\frac{\varepsilon}{2(2 + |g(x)| + |f(x)|)} + \frac{\varepsilon}{2(2 + |g(x)| + |f(x)|)} \right) = \varepsilon.$$

This proves the first part of 2.) To obtain the second part, let δ_1 be as described above and let $\delta_0 > 0$ be such that for $|x-y| < \delta_0$,

$$|g(x) - g(y)| < |g(x)|/2$$

and so by the triangle inequality,

$$-|g(x)|/2 \leq |g(y)| - |g(x)| \leq |g(x)|/2$$

which implies $|g(y)| \geq |g(x)|/2$, and $|g(y)| < 3|g(x)|/2$.

Then if $|x-y| < \min(\delta_0, \delta_1)$,

$$\begin{aligned} \left| \frac{f(x)}{g(x)} - \frac{f(y)}{g(y)} \right| &= \left| \frac{f(x)g(y) - f(y)g(x)}{g(x)g(y)} \right| \\ &\leq \frac{|f(x)g(y) - f(y)g(x)|}{\left(\frac{|g(x)|^2}{2}\right)} \\ &= \frac{2|f(x)g(y) - f(y)g(x)|}{|g(x)|^2} \\ &\leq \frac{2}{|g(x)|^2} [|f(x)g(y) - f(y)g(y) + f(y)g(y) - f(y)g(x)|] \\ &\leq \frac{2}{|g(x)|^2} [|g(y)||f(x) - f(y)| + |f(y)||g(y) - g(x)|] \\ &\leq \frac{2}{|g(x)|^2} \left[\frac{3}{2}|g(x)||f(x) - f(y)| + (1 + |f(x)|)|g(y) - g(x)| \right] \\ &\leq \frac{2}{|g(x)|^2} (1 + 2|f(x)| + 2|g(x)|) [|f(x) - f(y)| + |g(y) - g(x)|] \\ &\equiv M [|f(x) - f(y)| + |g(y) - g(x)|] \end{aligned}$$

where M is defined by

$$M \equiv \frac{2}{|g(x)|^2} (1 + 2|f(x)| + 2|g(x)|)$$

Now let δ_2 be such that if $|x-y| < \delta_2$, then

$$|f(x) - f(y)| < \frac{\varepsilon}{2} M^{-1}$$

and let δ_3 be such that if $|x-y| < \delta_3$, then

$$|g(y) - g(x)| < \frac{\varepsilon}{2} M^{-1}.$$

Then if $0 < \delta \leq \min(\delta_0, \delta_1, \delta_2, \delta_3)$, and $|x-y| < \delta$, everything holds and

$$\left| \frac{f(x)}{g(x)} - \frac{f(y)}{g(y)} \right| \leq M [|f(x) - f(y)| + |g(y) - g(x)|]$$

$$< M \left[\frac{\varepsilon}{2} M^{-1} + \frac{\varepsilon}{2} M^{-1} \right] = \varepsilon.$$

This completes the proof of the second part of 2.)

Note that in these proofs no effort is made to find some sort of “best” δ . The problem is one which has a yes or a no answer. Either it is or it is not continuous.

Now consider 3.). If f is continuous at x , $f(x) \in D(g) \subseteq \mathbb{R}^p$, and g is continuous at $f(x)$, then $g \circ f$ is continuous at x . Let $\varepsilon > 0$ be given. Then there exists $\eta > 0$ such that if $|y - f(x)| < \eta$ and $y \in D(g)$, it follows that $|g(y) - g(f(x))| < \varepsilon$. From continuity of f at x , there exists $\delta > 0$ such that if $|x - z| < \delta$ and $z \in D(f)$, then $|f(z) - f(x)| < \eta$. Then if $|x - z| < \delta$ and $z \in D(g \circ f) \subseteq D(f)$, all the above hold and so

$$|g(f(z)) - g(f(x))| < \varepsilon.$$

This proves part 3.)

To verify part 4.), let $\varepsilon > 0$ be given and let $\delta = \varepsilon$. Then if $|x - y| < \delta$, the triangle inequality implies

$$\begin{aligned} |f(x) - f(y)| &= ||x| - |y|| \\ &\leq |x - y| < \delta = \varepsilon. \end{aligned}$$

This proves part 4.) and completes the proof of the theorem.

Theorem 2.15.2 *In this theorem, the symbol, $\lim_{y \rightarrow x}$ denotes limits from the right or left as well as limits at ∞ or $-\infty$. Suppose $\lim_{y \rightarrow x} f(y) = L$ and $\lim_{y \rightarrow x} g(y) = K$ where K and L are real numbers in $\mathbb{R}$. Then if $a, b \in \mathbb{R}$,*

$$\lim_{y \rightarrow x} (af(y) + bg(y)) = aL + bK, \quad (2.11)$$

$$\lim_{y \rightarrow x} fg(y) = LK \quad (2.12)$$

and if $K \neq 0$,

$$\lim_{y \rightarrow x} \frac{f(y)}{g(y)} = \frac{L}{K}. \quad (2.13)$$

Also, if h is a continuous function defined near L , then

$$\lim_{y \rightarrow x} h \circ f(y) = h(L). \quad (2.14)$$

Suppose $\lim_{y \rightarrow x} f(y) = L$. If $f(y) \leq a$ all y of interest, then $L \leq a$ and if $f(y) \geq a$ then $L \geq a$.

Proof: The proof of 2.11 is left for you. It is like a corresponding theorem for continuous functions. Next consider 2.12. Let $\varepsilon > 0$ be given. Then by the triangle inequality,

$$\begin{aligned} |fg(y) - LK| &\leq |fg(y) - f(y)K| + |f(y)K - LK| \\ &\leq |f(y)| |g(y) - K| + |K| |f(y) - L|. \end{aligned} \quad (2.15)$$

There exists δ_1 such that if $0 < |y - x| < \delta_1$, then

$$|f(y) - L| < 1,$$

and so for such y , and the triangle inequality, $|f(y)| < 1 + |L|$. Therefore, for $0 < |y - x| < \delta_1$,

$$|fg(y) - LK| \leq (1 + |K| + |L|) [|g(y) - K| + |f(y) - L|]. \quad (2.16)$$

Now let $0 < \delta_2$ be such that for $0 < |x - y| < \delta_2$,

$$|f(y) - L| < \frac{\varepsilon}{2(1 + |K| + |L|)}, \quad |g(y) - K| < \frac{\varepsilon}{2(1 + |K| + |L|)}.$$

Then letting $0 < \delta \leq \min(\delta_1, \delta_2)$, it follows from 2.16 that

$$|fg(y) - LK| < \varepsilon$$

and this proves 2.12. Limits as $x \rightarrow \pm\infty$ and one sided limits are handled similarly.

The proof of 2.13 is left to you. It is just like the theorem about the quotient of continuous functions being continuous provided the function in the denominator is non zero at the point of interest.

Consider 2.14. Since h is continuous near L , it follows that for $\varepsilon > 0$ given, there exists $\eta > 0$ such that if $|y - L| < \eta$, then

$$|h(y) - h(L)| < \varepsilon$$

Now since $\lim_{y \rightarrow x} f(y) = L$, there exists $\delta > 0$ such that if $0 < |y - x| < \delta$, then

$$|f(y) - L| < \eta.$$

Therefore, if $0 < |y - x| < \delta$,

$$|h(f(y)) - h(L)| < \varepsilon.$$

The same theorem holds for one sided limits and limits as the variable moves toward $\pm\infty$. The proofs are left to you. They are minor modifications of the above.

It only remains to verify the last assertion. Assume $f(y) \leq a$. It is required to show that $L \leq a$. If this is not true, then $L > a$. Letting ε be small enough that $a < L - \varepsilon$, it follows that ultimately, for y close enough to x , $f(y) \in (L - \varepsilon, L + \varepsilon)$ which requires $f(y) > a$ contrary to assumption. This proves the theorem.

Theorem 2.15.3 Suppose $\{a_n\}$ and $\{b_n\}$ are sequences and that

$$\lim_{n \rightarrow \infty} a_n = a \text{ and } \lim_{n \rightarrow \infty} b_n = b.$$

Also suppose x and y are real numbers. Then

$$\lim_{n \rightarrow \infty} xa_n + yb_n = xa + yb \quad (2.17)$$

$$\lim_{n \rightarrow \infty} a_nb_n = ab \quad (2.18)$$

If $b \neq 0$,

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{a}{b}. \quad (2.19)$$

Proof: The first of these claims is left for you to do. To do the second, let $\varepsilon > 0$ be given and choose n_1 such that if $n \geq n_1$ then

$$|a_n - a| < 1.$$

Then for such n , the triangle inequality implies

$$\begin{aligned} |a_nb_n - ab| &\leq |a_nb_n - a_nb| + |a_nb - ab| \\ &\leq |a_n||b_n - b| + |b||a_n - a| \\ &\leq (|a| + 1)|b_n - b| + |b||a_n - a|. \end{aligned}$$

Now let n_2 be large enough that for $n \geq n_2$,

$$|b_n - b| < \frac{\varepsilon}{2(|a| + 1)}, \text{ and } |a_n - a| < \frac{\varepsilon}{2(|b| + 1)}.$$

Such a number exists because of the definition of limit. Therefore, let

$$n_\varepsilon > \max(n_1, n_2).$$

For $n \geq n_\varepsilon$,

$$\begin{aligned} |a_n b_n - ab| &\leq (|a| + 1) |b_n - b| + |b| |a_n - a| \\ &< (|a| + 1) \frac{\varepsilon}{2(|a| + 1)} + |b| \frac{\varepsilon}{2(|b| + 1)} \leq \varepsilon. \end{aligned}$$

This proves 2.18. Next consider 2.19.

Let $\varepsilon > 0$ be given and let n_1 be so large that whenever $n \geq n_1$,

$$|b_n - b| < \frac{|b|}{2}.$$

Thus for such n ,

$$\begin{aligned} \left| \frac{a_n}{b_n} - \frac{a}{b} \right| &= \left| \frac{a_n b - ab_n}{bb_n} \right| \leq \frac{2}{|b|^2} [|a_n b - ab| + |ab - ab_n|] \\ &\leq \frac{2}{|b|} |a_n - a| + \frac{2|a|}{|b|^2} |b_n - b|. \end{aligned}$$

Now choose n_2 so large that if $n \geq n_2$, then

$$|a_n - a| < \frac{\varepsilon |b|}{4}, \text{ and } |b_n - b| < \frac{\varepsilon |b|^2}{4(|a| + 1)}.$$

Letting $n_\varepsilon > \max(n_1, n_2)$, it follows that for $n \geq n_\varepsilon$,

$$\begin{aligned} \left| \frac{a_n}{b_n} - \frac{a}{b} \right| &\leq \frac{2}{|b|} |a_n - a| + \frac{2|a|}{|b|^2} |b_n - b| \\ &< \frac{2}{|b|} \frac{\varepsilon |b|}{4} + \frac{2|a|}{|b|^2} \frac{\varepsilon |b|^2}{4(|a| + 1)} < \varepsilon. \end{aligned}$$

This proves the theorem.

2.15.2 The Intermediate Value Theorem

Next here is a proof of the intermediate value theorem.

Theorem 2.15.4 *Suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous and suppose $f(a) < c < f(b)$. Then there exists $x \in (a, b)$ such that $f(x) = c$.*

Proof: Let $d = \frac{a+b}{2}$ and consider the intervals $[a, d]$ and $[d, b]$. If $f(d) \geq c$, then on $[a, d]$, the function is $\leq c$ at one end point and $\geq c$ at the other. On the other hand, if $f(d) \leq c$, then on $[d, b]$ $f \geq 0$ at one end point and ≤ 0 at the other. Pick the interval on which f has values which are at least as large as c and values no larger than c . Now consider that interval, divide it in half as was done for the original interval and argue that on one of these smaller intervals, the function has values at least as large as c and values no larger than c . Continue in this way. Next apply the nested interval lemma to

get x in all these intervals. In the n^{th} interval, let x_n, y_n be points of this interval such that $f(x_n) \leq c, f(y_n) \geq c$. Now $|x_n - x| \leq (b - a)2^{-n}$ and $|y_n - x| \leq (b - a)2^{-n}$ and so $x_n \rightarrow x$ and $y_n \rightarrow x$. Therefore,

$$f(x) - c = \lim_{n \rightarrow \infty} (f(x_n) - c) \leq 0$$

while

$$f(x) - c = \lim_{n \rightarrow \infty} (f(y_n) - c) \geq 0.$$

Consequently $f(x) = c$ and this proves the theorem. (For the last step, see Problem 13 on Page 69).

Here is another proof of the intermediate value theorem.

Theorem 2.15.5 *Suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous and suppose $f(a) < c < f(b)$. Then there exists $x \in (a, b)$ such that $f(x) = c$.*

Proof: Since $f(a) < c$, the set, S defined as

$$S \equiv \{y \in [a, b] : f(t) < c \text{ for all } t \leq y\}$$

is nonempty. In particular $a \in S$. Also, since $S \subseteq [a, b]$, S is bounded above and so there exists a least upper bound, $x \leq b$. I claim $f(x) \leq c$. Suppose $f(x) > c$. Then letting $\varepsilon \equiv f(x) - c$, there exists $\delta > 0$ such that if $|y - x| < \delta$ and $y \in [a, b]$ then

$$f(x) - f(y) \leq |f(x) - f(y)| < \varepsilon \equiv f(x) - c$$

which implies that for such y , $f(y) > c$. Therefore, $x > a$ and for all $y \in [a, b] \cap (x - \delta, x)$, it follows $f(y) > c$ and this violates the definition of x as the least upper bound of S . In particular those $y \in [a, b] \cap (x - \delta, x)$ are upper bounds to S and these are smaller than x . Thus $f(x) \leq c$. Next suppose $f(x) < c$ which implies $x < b$. Then letting $\varepsilon \equiv c - f(x)$ there exists $\delta > 0$ such that for $|y - x| < \delta$ and $y \in [a, b]$,

$$f(y) - f(x) \leq |f(x) - f(y)| < \varepsilon = c - f(x)$$

and so for $y \in (x, x + \delta) \cap [a, b]$, it follows $f(y) < c$. But for $y < x$, $f(y) \leq c$ and so x fails to be an upper bound for S . Therefore, it must be the case that $f(x) = c$ and x is not equal to either a or b . This proves the theorem.

Lemma 2.15.6 *Let $\phi : [a, b] \rightarrow \mathbb{R}$ be a continuous function and suppose ϕ is 1 - 1 on (a, b) . Then ϕ is either strictly increasing or strictly decreasing on $[a, b]$.*

Proof: First it is shown that ϕ is either strictly increasing or strictly decreasing on (a, b) .

If ϕ is not strictly decreasing on (a, b) , then there exists $x_1 < y_1$, $x_1, y_1 \in (a, b)$ such that

$$(\phi(y_1) - \phi(x_1))(y_1 - x_1) > 0.$$

If for some other pair of points, $x_2 < y_2$ with $x_2, y_2 \in (a, b)$, the above inequality does not hold, then since ϕ is 1 - 1,

$$(\phi(y_2) - \phi(x_2))(y_2 - x_2) < 0.$$

Let $x_t \equiv tx_1 + (1 - t)x_2$ and $y_t \equiv ty_1 + (1 - t)y_2$. Then $x_t < y_t$ for all $t \in [0, 1]$ because

$$tx_1 \leq ty_1 \text{ and } (1 - t)x_2 \leq (1 - t)y_2$$

with strict inequality holding for at least one of these inequalities since not both t and $(1-t)$ can equal zero. Now define

$$h(t) \equiv (\phi(y_t) - \phi(x_t))(y_t - x_t).$$

Since h is continuous and $h(0) < 0$, while $h(1) > 0$, there exists $t \in (0, 1)$ such that $h(t) = 0$. Therefore, both x_t and y_t are points of (a, b) and $\phi(y_t) - \phi(x_t) = 0$ contradicting the assumption that ϕ is one to one. It follows ϕ is either strictly increasing or strictly decreasing on (a, b) .

This property of being either strictly increasing or strictly decreasing on (a, b) carries over to $[a, b]$ by the continuity of ϕ . Suppose ϕ is strictly increasing on (a, b) , a similar argument holding for ϕ strictly decreasing on (a, b) . If $x > a$, then pick $y \in (a, x)$ and from the above, $\phi(y) < \phi(x)$. Now by continuity of ϕ at a ,

$$\phi(a) = \lim_{x \rightarrow a+} \phi(x) \leq \phi(y) < \phi(x).$$

Therefore, $\phi(a) < \phi(x)$ whenever $x \in (a, b)$. Similarly $\phi(b) > \phi(x)$ for all $x \in (a, b)$. This proves the lemma.

Corollary 2.15.7 *Let $f : (a, b) \rightarrow \mathbb{R}$ be one to one and continuous. Then $f(a, b)$ is an open interval, (c, d) and $f^{-1} : (c, d) \rightarrow (a, b)$ is continuous.*

Proof: Since f is either strictly increasing or strictly decreasing, it follows that $f(a, b)$ is an open interval, (c, d) . Assume f is decreasing. Now let $x \in (a, b)$. Why is f^{-1} continuous at $f(x)$? Since f is decreasing, if $f(x) < f(y)$, then $y \equiv f^{-1}(f(y)) < x \equiv f^{-1}(f(x))$ and so f^{-1} is also decreasing. Let $\varepsilon > 0$ be given. Let $\varepsilon > \eta > 0$ and $(x - \eta, x + \eta) \subseteq (a, b)$. Then $f(x) \in (f(x + \eta), f(x - \eta))$. Let

$$\delta = \min(f(x) - f(x + \eta), f(x - \eta) - f(x)).$$

Then if

$$|f(z) - f(x)| < \delta,$$

it follows

$$z \equiv f^{-1}(f(z)) \in (x - \eta, x + \eta) \subseteq (x - \varepsilon, x + \varepsilon)$$

so

$$|f^{-1}(f(z)) - x| = |f^{-1}(f(z)) - f^{-1}(f(x))| < \varepsilon.$$

This proves the theorem in the case where f is strictly decreasing. The case where f is increasing is similar.

2.15.3 Extreme Value Theorem

Now consider the extreme value theorem. This is based on the nested interval lemma.

In Russia there is a kind of doll called a matrushka doll. You pick it up and notice it comes apart in the center. Separating the two halves you find an identical doll inside. Then you notice this inside doll also comes apart in the center. Separating the two halves, you find yet another identical doll inside. This goes on quite a while until the final doll is in one piece. The nested interval lemma is like a matrushka doll except the process never stops. It involves a sequence of intervals, the first containing the second, the second containing the third, the third containing the fourth and so on. The fundamental question is whether there exists a point in all the intervals.

Lemma 2.15.8 *Let $I_k = [a^k, b^k]$ and suppose that for all $k = 1, 2, \dots$,*

$$I_k \supseteq I_{k+1}.$$

Then there exists a point, $c \in \mathbb{R}$ which is an element of every I_k .

Proof: Since $I_k \supseteq I_{k+1}$, this implies

$$a^k \leq a^{k+1}, \quad b^k \geq b^{k+1}. \quad (2.20)$$

Consequently, if $k \leq l$,

$$a^l \leq a^k \leq b^k \leq b^l. \quad (2.21)$$

Now define

$$c \equiv \sup \{a^l : l = 1, 2, \dots\}$$

By the first inequality in 2.20, and 2.21

$$a^k \leq c = \sup \{a^l : l = k, k+1, \dots\} \leq b^k \quad (2.22)$$

for each $k = 1, 2, \dots$. Thus $c \in I_k$ for every k and this proves the lemma. If this went too fast, the reason for the last inequality in 2.22 is that from 2.21, b^k is an upper bound to $\{a^l : l = k, k+1, \dots\}$. Therefore, it is at least as large as the least upper bound.

This is really quite a remarkable result and may not seem so obvious. Consider the intervals $I_k \equiv (0, 1/k)$. Then there is no point which lies in all these intervals because no negative number can be in all the intervals and $1/k$ is smaller than a given positive number whenever k is large enough. Thus the only candidate for being in all the intervals is 0 and 0 has been left out of them all. The problem here is that the endpoints of the intervals were not included contrary to the hypotheses of the above lemma in which all the intervals included the endpoints.

With the nested interval lemma, it becomes possible to prove the following lemma which shows a function continuous on a closed interval in $\mathbb{R}$ is bounded.

Lemma 2.15.9 *Let $I = [a, b]$ and let $f : I \rightarrow \mathbb{R}$ be continuous. Then f is bounded. That is there exist numbers, m and M such that for all $x \in [a, b]$,*

$$m \leq f(x) \leq M.$$

Proof: Let $I \equiv I_0$ and suppose f is not bounded on I_0 . Consider the two sets, $[a, \frac{a+b}{2}]$ and $[\frac{a+b}{2}, b]$. Since f is not bounded on I_0 , it follows that f must fail to be bounded on at least one of these sets. Let I_1 be one of these on which f is not bounded. Now do to I_1 what was done to I_0 to obtain $I_2 \subseteq I_1$ and for any two points, $x, y \in I_2$

$$|x - y| \leq 2^{-1} \frac{b-a}{2} \leq 2^{-2} (b-a).$$

Continue in this way obtaining sets, I_k such that $I_k \supseteq I_{k+1}$ and for any two points in I_k , x, y , $|x - y| \leq 2^{-k} (b-a)$. By the nested interval lemma, there exists a point, c which is contained in each I_k . Also, by continuity, there exists a $\delta > 0$ such that if $|c - y| < \delta$, then

$$|f(c) - f(y)| < 1. \quad (2.23)$$

Let k be so large that $2^{-k} (b-a) < \delta$. Then for every $y \in I_k$, $|c - y| < \delta$ and so 2.23 holds for all such y . But this implies that for all $y \in I_k$,

$$|f(y)| \leq |f(c)| + 1$$

which shows that f is bounded on I_k contrary to the way I_k was chosen. This contradiction proves the lemma.

Example 2.15.10 *Let $f(x) = 1/x$ for $x \in (0, 1)$.*

Clearly, f is not bounded. Does this violate the conclusion of the nested interval lemma? It does not because the end points of the interval involved are not in the interval. The same function defined on $[.000001, 1)$ would have been bounded although in this case the boundedness of the function would not follow from the above lemma because it fails to include the right endpoint.

Theorem 2.15.11 *Let $I = [a, b]$ and let $f : I \rightarrow \mathbb{R}$ be continuous. Then f achieves its maximum and its minimum on I . This means there exist, $x_1, x_2 \in I$ such that for all $x \in I$,*

$$f(x_1) \leq f(x) \leq f(x_2).$$

Proof: By completeness of $\mathbb{R}$ and Lemma 2.15.9 $f(I)$ has a least upper bound, M . If for all $x \in I$, $f(x) \neq M$, then by Theorem 2.4.1, the function, $g(x) \equiv (M - f(x))^{-1} = \frac{1}{M - f(x)}$ is continuous on I . Since M is the least upper bound of $f(I)$ there exist points, $x \in I$ such that $(M - f(x))$ is as small as desired. Consequently, g is not bounded above, contrary to Lemma 2.15.9. Therefore, there must exist some $x \in I$ such that $f(x) = M$. This proves f achieves its maximum. The argument for the minimum is similar. Alternatively, you could consider the function $h(x) = M - f(x)$. Then use what was just proved to conclude h achieves its maximum at some point, x_1 . Thus $h(x_1) \geq h(x)$ for all $x \in I$ and so $M - f(x_1) \geq M - f(x)$ for all $x \in I$ which implies $f(x_1) \leq f(x)$ for all $x \in I$. This proves the theorem.

2.15.4 Different Versions Of Completeness

Theorem 2.15.12 *The following conditions are equivalent to completeness.*

1. *Every Cauchy sequence converges*
2. *Every monotone increasing sequence which is bounded above converges.*
3. *Every monotone decreasing sequence which is bounded below converges.*

Proof: Suppose every Cauchy sequence converges and let S be a non empty set which is bounded above. In what follows, $s_n \in S$ and b_n will be an upper bound of S . If, in the process about to be described, $s_n = b_n$, this will have shown the existence of a least upper bound to S . Therefore, assume $s_n < b_n$ for all n . Let b_1 be an upper bound of S and let s_1 be an element of S . Suppose $s_1, \dots, s_n$ and $b_1, \dots, b_n$ have been chosen such that $s_k \leq s_{k+1}$ and $b_k \geq b_{k+1}$. Consider $\frac{s_n + b_n}{2}$, the point on $\mathbb{R}$ which is mid way between s_n and b_n . If this point is an upper bound, let

$$b_{n+1} \equiv \frac{s_n + b_n}{2}$$

and $s_{n+1} = s_n$. If the point is not an upper bound, let

$$s_{n+1} \in \left(\frac{s_n + b_n}{2}, b_n \right)$$

and let $b_{n+1} = b_n$. It follows this specifies an increasing sequence $\{s_n\}$ and a decreasing sequence $\{b_n\}$ such that

$$0 \leq b_n - s_n \leq 2^{-n} (b_1 - s_1).$$

Now if $n > m$,

$$\begin{aligned} 0 \leq b_m - b_n &= |b_m - b_n| \\ &= \sum_{k=m}^{n-1} b_k - b_{k+1} \leq \sum_{k=m}^{n-1} b_k - s_k \leq \sum_{k=m}^{n-1} 2^{-k} (b_1 - s_1) \end{aligned}$$

$$= \frac{2^{-m} - 2^{-n}}{2^{-1}} (b_1 - s_1) \leq 2^{-m+1} (b_1 - s_1)$$

and $\lim_{m \rightarrow \infty} 2^{-m} = 0$ by Theorem 2.11.9. Therefore, $\{b_n\}$ is a Cauchy sequence. Similarly, $\{s_n\}$ is a Cauchy sequence. Let $l \equiv \lim_{n \rightarrow \infty} s_n$ and let $l_1 \equiv \lim_{n \rightarrow \infty} b_n$. If n is large enough,

$$|l - s_n| < \varepsilon/3, |l_1 - b_n| < \varepsilon/3, \text{ and } |b_n - s_n| < \varepsilon/3.$$

Then

$$\begin{aligned} |l - l_1| &\leq |l - s_n| + |s_n - b_n| + |b_n - l_1| \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, $l = l_1$. Why? Then l must be the least upper bound of S . It is an upper bound because if there were $s > l$ where $s \in S$, then by the definition of limit, $b_n < s$ for some n , violating the assumption that each b_n is an upper bound for S . On the other hand, if $l_0 < l$, then for all n large enough, $s_n > l_0$, which implies l_0 is not an upper bound. This shows 1 implies completeness.

First note that 2 and 3 are equivalent. Why? Suppose 2 and consequently 3. Then the same construction yields the two monotone sequences, one increasing and the other decreasing. The sequence $\{b_n\}$ is bounded below by s_m for all m and the sequence $\{s_n\}$ is bounded above by b_m for all m . Why? Therefore, the two sequences converge. The rest of the argument is the same as the above. Thus 2 and 3 imply completeness.

Now suppose completeness and let $\{a_n\}$ be an increasing sequence which is bounded above. Let a be the least upper bound of the set of points in the sequence. If $\varepsilon > 0$ is given, there exists n_ε such that $a - \varepsilon < a_{n_\varepsilon}$. Since $\{a_n\}$ is a monotone sequence, it follows that whenever $n > n_\varepsilon$, $a - \varepsilon < a_n \leq a$. This proves $\lim_{n \rightarrow \infty} a_n = a$ and proves convergence. Since 3 is equivalent to 2, this is also established. It follows 3 and 2 are equivalent to completeness. It remains to show that completeness implies every Cauchy sequence converges.

Suppose completeness and let $\{a_n\}$ be a Cauchy sequence. Let

$$\inf \{a_k : k \geq n\} \equiv A_n, \sup \{a_k : k \geq n\} \equiv B_n$$

Then A_n is an increasing sequence while B_n is a decreasing sequence and $B_n \geq A_n$. Furthermore,

$$\lim_{n \rightarrow \infty} B_n - A_n = 0.$$

The details of these assertions are easy and are left to the reader. Also, $\{A_n\}$ is bounded below by any lower bound for the original Cauchy sequence while $\{B_n\}$ is bounded above by any upper bound for the original Cauchy sequence. By the equivalence of completeness with 3 and 2, it follows there exists a such that $a = \lim_{n \rightarrow \infty} A_n = \lim_{n \rightarrow \infty} B_n$. Since $B_n \geq a_n \geq A_n$, the squeezing theorem implies $\lim_{n \rightarrow \infty} a_n = a$ and this proves the equivalence of these characterizations of completeness.

2.15.5 Sequential Compactness Of Closed Intervals

Theorem 2.15.13 *Every closed interval, $[a, b]$ is sequentially compact.*

Proof: Let $\{x_n\} \subseteq [a, b] \equiv I_0$. Consider the two intervals $[a, \frac{a+b}{2}]$ and $[\frac{a+b}{2}, b]$ each of which has length $(b-a)/2$. At least one of these intervals contains x_n for infinitely many values of n . Call this interval I_1 . Now do for I_1 what was done for I_0 . Split it in half and let I_2 be an interval which contains x_n for infinitely many values of n . Continue this way obtaining a sequence of nested intervals $I_0 \supseteq I_1 \supseteq I_2 \supseteq I_3 \cdots$ where

the length of I_n is $(b-a)/2^n$. Now pick n_1 such that $x_{n_1} \in I_1$, n_2 such that $n_2 > n_1$ and $x_{n_2} \in I_2$, n_3 such that $n_3 > n_2$ and $x_{n_3} \in I_3$, etc. (This can be done because in each case the intervals contained x_n for infinitely many values of n .) By the nested interval lemma there exists a point, c contained in all these intervals. Furthermore,

$$|x_{n_k} - c| < (b-a)2^{-k}$$

and so $\lim_{k \rightarrow \infty} x_{n_k} = c \in [a, b]$. This proves the theorem.

2.16 Exercises

1. If X_i are sets and for some j , $X_j = \emptyset$, the empty set. Verify carefully that $\prod_{i=1}^n X_i = \emptyset$.
2. Suppose $f(x) + f(\frac{1}{x}) = 7x$ and f is a function defined on $\mathbb{R} \setminus \{0\}$, the nonzero real numbers. Does there exist any such function?
3. Does there exist a function f , satisfying $f(x) - f(\frac{1}{x}) = 3x$ which has both x and $\frac{1}{x}$ in the domain of f ?
4. In the situation of the Fibonacci sequence show that the formula for the n^{th} term can be found and is given by

$$a_n = \frac{\sqrt{5}}{5} \left(\frac{1 + \sqrt{5}}{2} \right)^n - \frac{\sqrt{5}}{5} \left(\frac{1 - \sqrt{5}}{2} \right)^n.$$

Hint: You might be able to do this by induction but a better way would be to look for a solution to the recurrence relation, $a_{n+2} \equiv a_n + a_{n+1}$ of the form r^n . You will be able to show that there are two values of r which work, one of which is $r = \frac{1+\sqrt{5}}{2}$. Next you can observe that if r_1^n and r_2^n both satisfy the recurrence relation then so does $cr_1^n + dr_2^n$ for any choice of constants c, d . Then you try to pick c and d such that the conditions, $a_1 = 1$ and $a_2 = 1$ both hold.

5. In an ordinary annuity, you make constant payments, P at the beginning of each payment period. These accrue interest at the rate of r per payment period. This means at the start of the first payment period, there is the payment $P \equiv A_1$. Then this produces an amount rP in interest so at the beginning of the second payment period, you would have $rP + P + P \equiv A_2$. Thus $A_2 = A_1(1+r) + P$. Then at the beginning of the third payment period you would have $A_2(1+r) + P \equiv A_3$. Continuing in this way, you see that the amount in the bank at the beginning of the n^{th} payment period would be A_n given by $A_n = A_{n-1}(1+r) + P$ and $A_1 = P$. Thus A is a function defined on the positive integers given recursively as just described and A_n is the amount at the beginning of the n^{th} payment period. Now if you wanted to find out A_n for large n , how would you do it? One way would be to use the recurrence relation n times. A better way would be to find a formula for A_n . Look for one in the form $A_n = Cz^n + s$ where C, z and s are to be determined. Show that $C = \frac{P}{r}$, $z = (1+r)$, and $s = -\frac{P}{r}$.
6. A well known puzzle consists of three pegs and several disks each of a different diameter, each having a hole in the center which allows it to be slid down each of the pegs. These disks are piled one on top of the other on one of the pegs, in order of decreasing diameter, the larger disks always being below the smaller disks. The problem is to move the whole pile of disks to another peg such that you never place a disk on a smaller disk. If you have n disks, how many moves

will it take? Of course this depends on n . If $n = 1$, you can do it in one move. If $n = 2$, you would need 3. Let A_n be the number required for n disks. Then in solving the puzzle, you must first obtain the top $n - 1$ disks arranged in order on another peg before you can move the bottom disk of the original pile. This takes A_{n-1} moves. Explain why $A_n = 2A_{n-1} + 1$, $A_1 = 1$ and give a formula for A_n . Look for one in the form $A_n = Cr^n + s$. This puzzle is called the Tower of Hanoi. When you have found a formula for A_n , explain why it is not possible to do this puzzle if n is very large.

7. Suppose f is a function defined on $\mathbb{R}$ and f is continuous at 0. Suppose also that $f(x + y) = f(x) + f(y)$. Show that if this is so, then f must be continuous at every value of $x \in \mathbb{R}$. Next show that for every rational number, r , $f(r) = rf(1)$. Finally explain why $f(r) = rf(1)$ for every r a real number. **Hint:** To do this last part, you need to use the density of the rational numbers and continuity of f .
8. Suppose f is a continuous function defined on $[0, 1]$ which maps $[0, 1]$ into $[0, 1]$. Show there exists $x \in [0, 1]$ such that $x = f(x)$. **Hint:** Consider $h(x) \equiv x - f(x)$ and the intermediate value theorem.
9. Using the binomial theorem, Problem 2 on Page 27, prove that for all $n \in \mathbb{N}$, $(1 + \frac{1}{n})^n \leq (1 + \frac{1}{n+1})^{n+1}$. **Hint:** Show first that $\binom{n}{k} = \frac{n \cdot (n-1) \cdots (n-k+1)}{k!}$. By the binomial theorem,

$$\left(1 + \frac{1}{n}\right)^n = \sum_{k=0}^n \binom{n}{k} \left(\frac{1}{n}\right)^k = \sum_{k=0}^n \frac{\overbrace{n \cdot (n-1) \cdots (n-k+1)}^{k \text{ factors}}}{k! n^k}.$$

Now consider the term $\frac{n \cdot (n-1) \cdots (n-k+1)}{k! n^k}$ and note that a similar term occurs in the binomial expansion for $\left(1 + \frac{1}{n+1}\right)^{n+1}$ except you replace n with $n + 1$ wherever this occurs. Argue the term got bigger and then note that in the binomial expansion for $\left(1 + \frac{1}{n+1}\right)^{n+1}$, there are more terms.

10. Prove by induction that for all $k \geq 4$, $2^k \leq k!$
11. First show that $2^k < k!$ for all $k > 4$. Now verify for all $n \in \mathbb{N}$, $(1 + \frac{1}{n})^n \leq 3$.
12. Prove $\lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n$ exists and equals a number less than 3.
13. Using Problem 11, prove $n^{n+1} \geq (n+1)^n$ for all integers, $n \geq 3$.
14. Let $A_n = \sum_{k=2}^n \frac{1}{k(k-1)}$ for $n \geq 2$. Show $\lim_{n \rightarrow \infty} A_n$ exists. **Hint:** Show there exists an upper bound to the A_n as follows.

$$\begin{aligned} \sum_{k=2}^n \frac{1}{k(k-1)} &= \sum_{k=2}^n \left(\frac{1}{k-1} - \frac{1}{k} \right) \\ &= \frac{1}{2} - \frac{1}{n-1} \leq \frac{1}{2}. \end{aligned}$$

15. Let $H_n = \sum_{k=1}^n \frac{1}{k^2}$ for $n \geq 2$. Show $\lim_{n \rightarrow \infty} H_n$ exists. **Hint:** Use the above problem to obtain the existence of an upper bound.

16. Let a be a positive number and let $x_1 = b > 0$ where $b^2 > a$. Explain why there exists such a number, b . Now having defined x_n , define $x_{n+1} \equiv \frac{1}{2} \left(x_n + \frac{a}{x_n} \right)$. Verify that $\{x_n\}$ is a decreasing sequence and that it satisfies $x_n^2 \geq a$ for all n and is therefore, bounded below. Explain why $\lim_{n \rightarrow \infty} x_n$ exists. If x is this limit, show that $x^2 = a$. Explain how this shows that every positive real number has a square root. This is an example of a recursively defined sequence. Note this does not give a formula for x_n , just a rule which tells how to define x_{n+1} if x_n is known.
17. Let $a_1 = 0$ and suppose that $a_{n+1} = \frac{9}{9-a_n}$. Write a_2, a_3, a_4 . Now prove that for all n , it follows that $a_n \leq \frac{9}{2} - \frac{3}{2}\sqrt{5}$ (By Problem 7 on Page 58 there is no problem with the existence of various roots of positive numbers.) and so the sequence is bounded above. Next show that the sequence is increasing and so it converges. Find the limit of the sequence. **Hint:** You should prove these things by induction. Finally, to find the limit, let $n \rightarrow \infty$ in both sides and argue that the limit, a , must satisfy $a = \frac{9}{9-a}$.

18. If $x \in \mathbb{R}$, show there exists a sequence of rational numbers, $\{x_n\}$ such that $x_n \rightarrow x$ and a sequence of irrational numbers, $\{x'_n\}$ such that $x'_n \rightarrow x$. Now consider the following function.

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}.$$

Show using the sequential version of continuity in Theorem 2.11.19 that f is discontinuous at every point.

19. If $x \in \mathbb{R}$, show there exists a sequence of rational numbers, $\{x_n\}$ such that $x_n \rightarrow x$ and a sequence of irrational numbers, $\{x'_n\}$ such that $x'_n \rightarrow x$. Now consider the following function.

$$f(x) = \begin{cases} x & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}.$$

Show using the sequential version of continuity in Theorem 2.11.19 that f is continuous at 0 and nowhere else.

20. The nested interval lemma, Lemma 2.7.8 and Theorem 2.11.19 can be used to give an easy proof of the intermediate value theorem. Suppose $f(a) > 0$ and $f(b) < 0$ for f a continuous function defined on $[a, b]$. The intermediate value theorem states that under these conditions, there exists $x \in (a, b)$ such that $f(x) = 0$. Prove this theorem as follows: Let $c = \frac{a+b}{2}$ and consider the intervals $[a, c]$ and $[c, b]$. Show that on one of these intervals, f is nonnegative at one end and nonpositive at the other. Now consider that interval, divide it in half as was done for the original interval and argue that on one of these smaller intervals, the function has different signs at the two endpoints. Continue in this way. Next apply the nested interval lemma to get x in all these intervals and argue there exist sequences, $x_n \rightarrow x$ and $y_n \rightarrow x$ such that $f(x_n) < 0$ and $f(y_n) > 0$. By continuity, you can assume $f(x_n) \rightarrow f(x)$ and $f(y_n) \rightarrow f(x)$. Show this requires that $f(x) = 0$.
21. If $\lim_{n \rightarrow \infty} a_n = a$, does it follow that $\lim_{n \rightarrow \infty} |a_n| = |a|$? Prove or else give a counter example.
22. Show the following converge to 0.

(a) $\frac{n^5}{1.01^n}$

(b) $\frac{10^n}{n!}$

23. * Suppose $\lim_{n \rightarrow \infty} x_n = x$. Show that then $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n x_k = x$. Give an example where $\lim_{n \rightarrow \infty} x_n$ does not exist but $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n x_k$ does.
24. Prove $\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1$. **Hint:** Let $e_n \equiv \sqrt[n]{n} - 1$ so that $(1 + e_n)^n = n$. Now observe that $e_n > 0$ and use the binomial theorem or Example 1.3.3 to conclude $1 + ne_n + \frac{n(n-1)}{2}e_n^2 \leq n$. This nice approach to establishing this limit using only elementary algebra is in Rudin [23].
25. Find $\lim_{n \rightarrow \infty} (x^n + 5)^{1/n}$ for $x \geq 0$. There are two cases here, $x < 1$ and $x \geq 1$. Show that if $x \geq 1$, the limit is x while if $x < 1$ the limit equals 1. **Hint:** Use the argument of Problem 24. This interesting example is in [8].

Derivatives

3.0.1 Outcomes

1. Understand the relation between velocity and the derivative.
2. Define and understand the definition of the derivative, both graphically and in terms of a limit. Be able to find derivatives using this definition.
3. Define and use the derivative to find local extrema.
4. Understand the proof of the mean value theorem, both the Lagrange and the Cauchy version.
5. Use the derivative as an aid to curve sketching.

3.1 Velocity

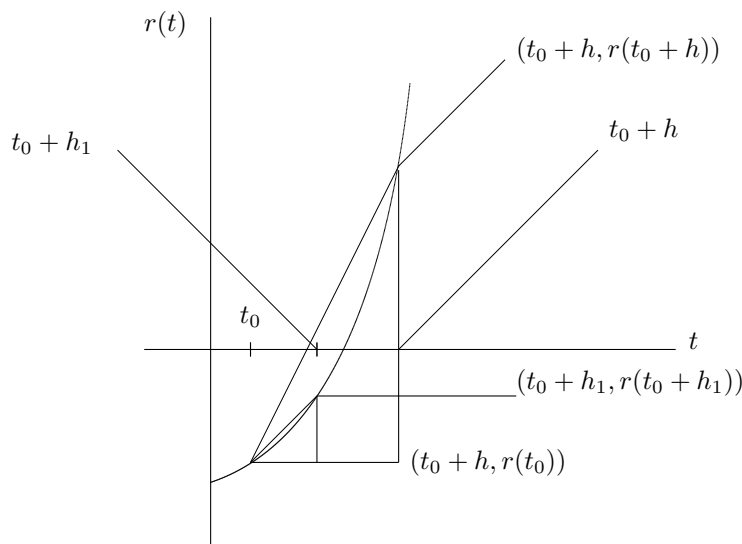
Imagine an object which is moving along the real line in the positive direction and that at time $t > 0$, the position of the object is $r(t) = -10 + 30t + t^2$ where distance is measured in kilometers and t in hours. Thus at $t = 0$, the object is at the point -10 kilometers and when $t = 1$, the object is at 21 kilometers. The average velocity during this time is the distance traveled divided by the elapsed time. Thus the average velocity would be $\frac{21 - (-10)}{1} = 31$ kilometers per hour. It came out positive because the object moved in the positive direction along the real line, from -10 to 21. Suppose it was desired to find something which deserves to be referred to as the instantaneous velocity when $t = 1/2$? If the object were a car, it is reasonable to suppose that the magnitude of the average velocity of the object over a very small interval of time would be very close to the number that would appear on the speedometer. For example, if considering the average velocity of the object on the interval $[\frac{1}{2}, \frac{1}{2} + .0001]$, this average velocity would be pretty close to the thing which deserves to be called the instantaneous velocity at $t = \frac{1}{2}$ hours. Thus the velocity at $t = \frac{1}{2}$ would be close to

$$\begin{aligned} & (r(.5 + .0001) - r(.5)) / .0001 \\ &= \left(30(.5 + .01) + (.5 + .0001)^2 - 30(.5) - (.5)^2 \right) / .0001 = 31.0001 \end{aligned}$$

Of course, you would expect to be even closer using a time interval of length .000001 instead of just .0001. In general, consider a time interval of length h and then define the instantaneous velocity to be the number which all these average velocities get close to as h gets smaller and smaller. Thus in this case form the average velocity on the interval, $[\frac{1}{2}, \frac{1}{2} + h]$ to get

$$\left(30(.5 + h) + (.5 + h)^2 - \left(30(.5) + (.5)^2 \right) \right) / h = 30 + 2(.5) + h.$$

What number does this average get close to as h gets smaller and smaller? Clearly it gets close to 31 and for this reason, the velocity at time .5 is defined as 31. It is positive because the object is moving in the positive direction. If the object were moving in the negative direction, the number would be negative. The notion just described of finding an instantaneous velocity has a geometrical application to finding the slope of a line tangent to a curve.



In the above picture, you see the slope of the line joining the two points $(t_0, r(t_0))$ and $(t_0 + h, r(t_0 + h))$ is given by

$$\frac{r(t_0 + h) - r(t_0)}{h}$$

which equals the average velocity on the time interval, $[t_0, t_0 + h]$. You can also see the effect of making h closer and closer to zero as illustrated by changing h to the smaller h_1 in the picture. The slope of the resulting line segment appears to get closer and closer to what ought to be considered the slope of the line tangent to the curve at the point $(t_0, r(t_0))$.

It is time to make this heuristic material much more precise.

3.2 The Derivative

The derivative of a function of one variable is a function given by the following definition.

Definition 3.2.1 *The derivative of a function, $f'(x)$, is defined as the following limit whenever the limit exists. If the limit does not exist, then neither does $f'(x)$.*

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \equiv f'(x) \quad (3.1)$$

The function of h on the left is called the difference quotient.

Note that the difference quotient on the left of the equation is a function of h which is not defined at $h = 0$. This is why, in the definition of limit, $|h| > 0$. **It is not**

necessary to have the function defined at the point in order to consider its limit. The distinction between the limit of a function and its value is very important and must be kept in mind. Also it is clear from setting $y = x + h$ that

$$f'(x) = \lim_{y \rightarrow x} \frac{f(y) - f(x)}{y - x}. \quad (3.2)$$

Theorem 3.2.2 *If $f'(x)$ exists, then f is continuous at x .*

Proof: Suppose $\varepsilon > 0$ is given and choose $\delta_1 > 0$ such that if $|h| < \delta_1$,

$$\left| \frac{f(x+h) - f(x)}{h} - f'(x) \right| < 1.$$

then for such h , the triangle inequality implies

$$|f(x+h) - f(x)| < |h| + |f'(x)| |h|.$$

Now letting $\delta < \min\left(\delta_1, \frac{\varepsilon}{1+|f'(x)|}\right)$ it follows if $|h| < \delta$, then

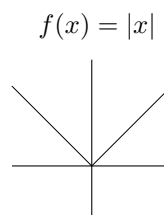
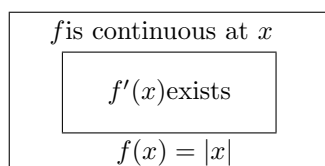
$$|f(x+h) - f(x)| < \varepsilon.$$

Letting $y = h + x$, this shows that if $|y - x| < \delta$,

$$|f(y) - f(x)| < \varepsilon$$

which proves f is continuous at x .

It is very important to remember that just because f is continuous, does not mean f has a derivative. The following picture describes the situation.



As indicated in the above picture the function $f(x) = |x|$ does not have a derivative at $x = 0$. To see this,

$$\lim_{h \rightarrow 0+} \frac{f(h) - f(0)}{h} = \lim_{h \rightarrow 0+} \frac{h}{h} = 1$$

while

$$\lim_{h \rightarrow 0-} \frac{f(h) - f(0)}{h} = \lim_{h \rightarrow 0-} \frac{-h}{h} = -1.$$

Thus the two limits, one from the right and one from the left do not agree as they would have to do if the function had a derivative at $x = 0$. See Problem 18 on Page 63. Geometrically, this lack of differentiability is manifested by there being a pointy place in the graph of $y = |x|$ at $x = 0$. In short, the pointy places don't have derivatives.

Example 3.2.3 *Let $f(x) = c$ where c is a constant. Find $f'(x)$.*

Set up the difference quotient,

$$\frac{f(x+h) - f(x)}{h} = \frac{c - c}{h} = 0$$

Therefore,

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} 0 = 0$$

Example 3.2.4 Let $f(x) = cx$ where c is a constant. Find $f'(x)$.

Set up the difference quotient,

$$\frac{f(x+h) - f(x)}{h} = \frac{c(x+h) - cx}{h} = \frac{ch}{h} = c.$$

Therefore,

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} c = c.$$

Example 3.2.5 Let $f(x) = \sqrt{x}$ for $x > 0$. Find $f'(x)$.

Set up the difference quotient,

$$\begin{aligned} \frac{f(x+h) - f(x)}{h} &= \frac{\sqrt{x+h} - \sqrt{x}}{h} = \frac{x+h-x}{h(\sqrt{x+h} + \sqrt{x})} \\ &= \frac{1}{\sqrt{x+h} + \sqrt{x}} \end{aligned}$$

and so

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{1}{\sqrt{x+h} + \sqrt{x}} = \frac{1}{2\sqrt{x}}.$$

There are rules of derivatives which make finding the derivative very easy.

Theorem 3.2.6 Let $a, b \in \mathbb{R}$ and suppose $f'(t)$ and $g'(t)$ exist. Then the following formulas are obtained.

$$(af + bg)'(t) = af'(t) + bg'(t). \quad (3.3)$$

$$(fg)'(t) = f'(t)g(t) + f(t)g'(t). \quad (3.4)$$

The formula, 3.4 is referred to as the product rule.

If $g(t) \neq 0$,

$$\left(\frac{f}{g}\right)'(t) = \frac{f'(t)g(t) - g'(t)f(t)}{g^2(t)}. \quad (3.5)$$

Formula 3.5 is referred to as the quotient rule.

If f is differentiable at ct where $c \neq 0$, Then letting $g(t) \equiv f(ct)$,

$$g'(t) = cf'(ct). \quad (3.6)$$

Written with a slight abuse of notation,

$$(f(ct))' = cf'(ct). \quad (3.7)$$

If f is differentiable on (a, b) and if $g(t) \equiv f(t+c)$, then g is differentiable on $(a-c, b-c)$ and

$$g'(t) = f'(t+c). \quad (3.8)$$

Written with a slight abuse of notation,

$$(f(t+c))' = f'(t+c) \quad (3.9)$$

For p an integer and $f'(t)$ exists, let $g_p(t) \equiv f(t)^p$. Then

$$(g_p)'(t) = pf(t)^{p-1}f'(t). \quad (3.10)$$

(In the case where $p < 0$, assume $f(t) \neq 0$.)

Written with a slight abuse of notation, an easy to remember version of 3.10 says

$$(f(t)^p)' = pf(t)^{p-1}f'(t).$$

Proof: The first formula is left for you to prove. Consider the second, 3.4.

$$\begin{aligned}\frac{fg(t+h) - fg(t)}{h} &= \frac{f(t+h)g(t+h) - f(t+h)g(t)}{h} + \frac{f(t+h)g(t) - f(t)g(t)}{h} \\ &= f(t+h) \frac{(g(t+h) - g(t))}{h} + \frac{(f(t+h) - f(t))}{h} g(t)\end{aligned}$$

Taking the limit as $h \rightarrow 0$ and using Theorem 3.2.2 to conclude $\lim_{h \rightarrow 0} f(t+h) = f(t)$, it follows from Theorem 2.9.4 that 3.4 follows. Next consider the quotient rule.

$$\begin{aligned}h^{-1} \left(\frac{f}{g}(t+h) - \frac{f}{g}(t) \right) &= \frac{f(t+h)g(t) - g(t+h)f(t)}{hg(t)g(t+h)} \\ &= \frac{f(t+h)(g(t) - g(t+h)) + g(t+h)(f(t+h) - f(t))}{hg(t)g(t+h)} \\ &= \frac{-f(t+h)}{g(t)g(t+h)} \frac{(g(t+h) - g(t))}{h} + \frac{g(t+h)}{g(t)g(t+h)} \frac{(f(t+h) - f(t))}{h}\end{aligned}$$

and from Theorem 2.9.4 on Page 60,

$$\left(\frac{f}{g} \right)'(t) = \frac{g(t)f'(t) - g'(t)f(t)}{g^2(t)}.$$

Now consider Formula 3.6.

$$\begin{aligned}(g(t+h) - g(t))h^{-1} &= h^{-1}(f(ct+ch) - f(ct)) \\ &= c \frac{f(ct+ch) - f(ct)}{ch} \\ &= c \frac{f(ct+h_1) - f(ct)}{h_1}\end{aligned}$$

where $h_1 = ch$. Then $h_1 \rightarrow 0$ if and only if $h \rightarrow 0$ and so taking the limit as $h \rightarrow 0$ yields

$$g'(t) = cf'(ct)$$

as claimed. Formulas 3.7 and 3.8 are left as an exercise.

First consider 3.10 in the case where p equals a nonnegative integer. If $p = 0$, 3.10 holds because $g_0(t) = 1$ and so by Example 3.2.3,

$$g'_0(t) = 0 = 0(f(t))^{-1}f'(t).$$

Next suppose 3.10 holds for p an integer. Then

$$(g_{p+1}(t)) = f(t)g_p(t)$$

and so by the product rule,

$$\begin{aligned}g'_{p+1}(t) &= f'(t)g_p(t) + f(t)g'_p(t) \\ &= f'(t)(f(t))^p + f(t)(pf(t)^{p-1}f'(t)) \\ &= (p+1)f(t)^p f'(t).\end{aligned}$$

If the formula holds for some integer, p then it holds for $-p$. Here is why.

$$g_{-p}(t) = g_p(t)^{-1}$$

and so

$$\frac{g_{-p}(t+h) - g_{-p}(t)}{h} = \left(\frac{g_p(t) - g_p(t+h)}{h} \right) \left(\frac{1}{g_p(t)g_p(t+h)} \right).$$

Taking the limit as $h \rightarrow 0$ and using the formula for p ,

$$\begin{aligned} g'_{-p}(t) &= -p f(t)^{p-1} f'(t) (f(t))^{-2p} \\ &= -p (f(t))^{-p-1} f'(t). \end{aligned}$$

This proves the theorem.

Example 3.2.7 Let $p(x) = 3 + 5x + 6x^2 - 7x^3$. Find $p'(x)$.

From the above theorem, and abusing the notation,

$$\begin{aligned} p'(x) &= (3 + 5x + 6x^2 - 7x^3)' \\ &= 3' + (5x)' + (6x^2)' + (-7x^3)' \\ &= 0 + 5 + (6)(2)(x)(x)' + (-7)(3)(x^2)(x)' \\ &= 5 + 12x - 21x^2. \end{aligned}$$

Note the process is to take the exponent and multiply by the coefficient and then make the new exponent one less in each term of the polynomial in order to arrive at the answer. This is the general procedure for differentiating a polynomial as shown in the next example.

Example 3.2.8 Let a_k be a number for $k = 0, 1, \dots, n$ and let $p(x) = \sum_{k=0}^n a_k x^k$. Find $p'(x)$.

Use Theorem 3.2.6

$$\begin{aligned} p'(x) &= \left(\sum_{k=0}^n a_k x^k \right)' = \sum_{k=0}^n a_k (x^k)' \\ &= \sum_{k=0}^n a_k k x^{k-1} (x)' = \sum_{k=0}^n a_k k x^{k-1} \end{aligned}$$

Example 3.2.9 Find the derivative of the function $f(x) = \frac{x^2+1}{x^3}$.

Use the quotient rule

$$f'(x) = \frac{2x(x^3) - 3x^2(x^2+1)}{x^6} = -\frac{1}{x^4}(x^2+3)$$

Example 3.2.10 Let $f(x) = (x^2+1)^4(x^3)$. Find $f'(x)$.

Use the product rule and 3.10. Abusing the notation for the sake of convenience,

$$\begin{aligned} \left((x^2+1)^4(x^3) \right)' &= \left((x^2+1)^4 \right)' (x^3) + (x^3)' \left((x^2+1)^4 \right) \\ &= 4(x^2+1)^3(2x)(x^3) + 3x^2(x^2+1)^4 \\ &= 4x^4(x^2+1)^3 + 3x^2(x^2+1)^4 \end{aligned}$$

Example 3.2.11 Let $f(x) = x^3/(x^2+1)^2$. Find $f'(x)$.

Use the quotient rule to obtain

$$f'(x) = \frac{3x^2(x^2+1)^2 - 2(x^2+1)(2x)x^3}{(x^2+1)^4} = \frac{3x^2(x^2+1)^2 - 4x^4(x^2+1)}{(x^2+1)^4}$$

Obviously, one could consider taking the derivative of the derivative and then the derivative of that and so forth. The main thing to consider about this is the notation. The second derivative is denoted with two primes.

Example 3.2.12 Let $f(x) = x^3 + 2x^2 + 1$. Find $f''(x)$ and $f'''(x)$.

To find $f''(x)$ take the derivative of the derivative. Thus $f'(x) = 3x^2 + 4x$ and so $f''(x) = 6x + 4$. Then $f'''(x) = 6$.

When high derivatives are taken, say the 5^{th} derivative, it is customary to write $f^{(5)}(t)$ putting the number of derivatives in parentheses.

3.3 Exercises

- Find derivatives of the following functions.

(a) $3x^2 + 4x^3 - 7x + 11$

(b) $\frac{1}{x^3+1}$

(c) $(x^4 + x + 5)(x^5 + 7x)$

(d) $\frac{x^3+x}{x^2+1}$

(e) $(x^2 + 2)^4$

(f) $(x^3 + 2x + 1)^{-1}$

- For $f(x) = -3x^7 + 4x^5 + 2x^3 + x^2 - 5x$, find $f^{(3)}(x)$.

- For $f(x) = -x^7 + x^5 + x^3 - 2x$, find $f^{(3)}(x)$.

- Find $f'(x)$ for the given functions

(a) $f(x) = \frac{3x^3 - x - 1}{3x^2 + 1}$

(b) $f(x) = \frac{3x^3 + 3x - 1}{x^2 + 1}$

(c) $f(x) = \frac{(x^2 + 5)^4}{(x^2 + 1)^2}$

(d) $f(x) = (4x + 7x^3)^6$

(e) $f(x) =$

$$(-2x^3 + x + 3)^3(-2x^2 + 3x - 1)$$

- For $f(x) = \sqrt{4x^2 + 1}$, find $f'(x)$ from the definition of the derivative.

- For $f(x) = \sqrt[3]{3x^2 + 1}$, find $f'(x)$ from the definition of the derivative.

Hint: You might use

$$b^3 - a^3 = (b^2 + ab + a^2)(b - a) \text{ for}$$

$$b = \sqrt[3]{3(x+h)^2 + 1} \text{ and } a = \sqrt[3]{3x^2 + 1}.$$

- For $f(x) = (-3x + 5)^5$, find $f'(x)$ from the definition of the derivative.

Hint: You might use the formula

$$b^n - a^n = (b - a) \cdot$$

$$(b^{n-1} + b^{n-2}a + \cdots + a^{n-2}b + b^{n-1})$$

- Let

$$f(x) = (x+2)^2 \sin(1/(x+2)) + 6(x-5)(x+2)$$

for $x \neq -2$ and define $f(-2) \equiv 0$. Find $f'(-2)$ from the definition of the derivative if this is possible.

Hint: Note that $|\sin(z)| \leq 1$ for any real value of z .

- Let $f(x) = (x+5) \sin(1/(x+5))$ for $x \neq -5$ and define $f(-5) \equiv 0$. Show $f'(-5)$ does not exist. **Hint:** Verify that $\lim_{h \rightarrow 0} \sin(1/h)$ does not exist and then explain why this shows $f'(-5)$ does not exist.

- Suppose f is a continuous function and $y = mx + b$ is the equation of a straight line with slope m which intersects the graph of f only at the point $(x, f(x))$. Does it follow that $f'(x) = m$?

11. Give an example of a function, f which is not continuous but for which

$$\lim_{n \rightarrow \infty} \frac{f(x + (1/n)) - f(x - (1/n))}{2(1/n)}$$

exists for every x .

12. Suppose f is defined on (a, b) and $f'(x)$ exists for some $x \in (a, b)$. Show $f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x-h)}{2h}$.

13. Let $f(x) = x^2$ if x is rational and $f(x) = 0$ if x is irrational. Show that $f'(0) = 0$ but that f fails to be continuous at every other point.

14. Suppose $|f(x)| \leq x^2$ for all $x \in (-1, 1)$. Let $h(x) = x + f(x)$. Find $h'(0)$.

15. Find a formula for the derivative of a product of three functions. Generalize to a product of n functions where n is a positive integer.

16. Suppose f is a function defined on $\mathbb{R}$ and it satisfies the functional equation, $f(a+b) = f(a) + f(b)$. Suppose also that $f'(0) = k$. Find $f'(x)$, $f(x)$.

17. Suppose f is a function defined on $\mathbb{R}$ and it satisfies the functional equation, $f(a+b) = f(a) + f(b) + ab$. Suppose also that $\lim_{h \rightarrow 0} \frac{f(h)}{h} = 7$. Find $f'(x)$, $f(x)$.

18. From Corollary 1.4.4 on Page 24 the following inequality holds.

$$\sin x + (1 - \cos x) \geq x \geq \sin x. \quad (3.11)$$

Using this inequality, show

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1.$$

Hint: For $x > 0$, divide both sides of the inequality by $\sin x$. This yields

$$1 + \frac{1 - \cos x}{x} \geq \frac{x}{\sin x} \geq 1.$$

Now

$$\begin{aligned} 0 &\leq \frac{1 - \cos x}{x} = \frac{1 - \cos^2 x}{x(1 + \cos x)} \\ &= \frac{\sin^2 x}{x(1 + \cos x)} \leq \frac{\sin x}{(1 + \cos x)}. \end{aligned} \quad (3.12)$$

If $x < 0$, $\frac{\sin x}{x} = \frac{\sin(-x)}{(-x)}$ and $-x > 0$.

19. Show that $\lim_{h \rightarrow 0} \left(\frac{1 - \cos(h)}{h} \right) = 0$.

Hint: For $h > 0$, consider the inequality 3.12 with x replaced with h .

20. Show using Problems 18 and 19 and the definition of the derivative that $\sin'(x) = \cos x$. Also show $\cos'(x) = -\sin(x)$. **Hint:** Just write down the difference quotient and use the limits in these problems.

$$\begin{aligned} &\frac{\sin(x+h) - \sin x}{h} = \\ &\frac{\sin(x)(\cos(h) - 1)}{h} \\ &+ \cos(x) \frac{\sin(h)}{h}, \text{ etc.} \end{aligned}$$

21. Now that you know $\sin'(x) = \cos(x)$, use the definition of derivative to find $f'(x)$ where $f(x) = \sin(3x)$. What about $g'(x)$ where $g(x) = \cos(3x)$? Now show that both $\cos(\omega t)$ and $\sin(\omega t)$ are solutions to the differential equation, $y'' + \omega^2 y = 0$. This is the equation for undamped oscillations and it will be discussed more later.

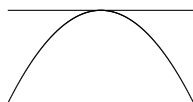
3.4 Local Extrema

When you are on top of a hill, you are at a local maximum although there may be other hills higher than the one on which you are standing. Similarly, when you are at the bottom of a valley, you are at a local minimum even though there may be other valleys deeper than the one you are in. The word, "local" is applied to the situation because if you confine your attention only to points close to your location, you are indeed at

either the top or the bottom.

Definition 3.4.1 Let $f : D(f) \rightarrow \mathbb{R}$ where here $D(f)$ is only assumed to be some subset of $\mathbb{R}$. Then $x \in D(f)$ is a local minimum (maximum) if there exists $\delta > 0$ such that whenever $y \in (x - \delta, x + \delta) \cap D(f)$, it follows $f(y) \geq (\leq) f(x)$. The plural of minimum is minima and the plural of maximum is maxima.

Derivatives can be used to locate local maxima and local minima. The following picture suggests how to do this. This picture is of the graph of a function having a local maximum and the tangent line to it.



Note how the tangent line is horizontal. If you were not at a local maximum or local minimum, the function would be falling or climbing and the tangent line would not be horizontal. This heuristic observation will be made more precise later.

Theorem 3.4.2 Suppose $f : (a, b) \rightarrow \mathbb{R}$ and suppose $x \in (a, b)$ is a local maximum or minimum. Then $f'(x) = 0$.

Proof: Suppose x is a local maximum. If $h > 0$ and is sufficiently small, then $f(x + h) \leq f(x)$ and so from Theorem 2.9.4 on Page 60,

$$f'(x) = \lim_{h \rightarrow 0+} \frac{f(x+h) - f(x)}{h} \leq 0.$$

Similarly,

$$f'(x) = \lim_{h \rightarrow 0-} \frac{f(x+h) - f(x)}{h} \geq 0.$$

The case when x is local minimum is similar. This proves the theorem.

Definition 3.4.3 Points where the derivative of a function equals zero are called critical points. It is also customary to refer to points where the derivative of a function does not exist as critical points.

Example 3.4.4 It is desired to find two positive numbers whose sum equals 16 and whose product is to be as large as possible.

The numbers are x and $16 - x$ and $f(x) = x(16 - x)$ is to be made as large as possible. The value of x which will do this would be a local maximum so by Theorem 3.4.2 the procedure is to take the derivative of f and find values of x where it equals zero. Thus $16 - 2x = 0$ and the only place this occurs is when $x = 8$. Therefore, the two numbers are 8 and 8.

Example 3.4.5 A farmer wants to fence a rectangular piece of land next to a straight river. What are the dimensions of the largest rectangle if there are exactly 600 meters of fencing available.

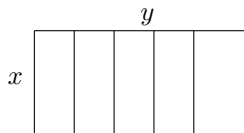
The two sides perpendicular to the river have length x and the third side has length $y = (600 - 2x)$. Thus the function to be maximized is $f(x) = 2x(600 - x) = 1200x - 2x^2$. Taking the derivative and setting it equal to zero gives

$$f'(x) = 1200 - 4x = 0$$

and so $x = 300$. Therefore, the desired dimensions are 300×600 .

Example 3.4.6 A rectangular playground is to be enclosed by a fence and divided in 5 pieces by 4 fences parallel to one side of the playground. 1704 feet of fencing is used. Find dimensions of the playground which will have the largest total area.

Let x denote the length of one of these dividing fences and let y denote the length of the playground as shown in the following picture.



Thus $6x + 2y = 1704$ so $y = \frac{1704-6x}{2}$ and the function to maximize is

$$f(x) = x \left(\frac{1704 - 6x}{2} \right) = x(852 - 3x).$$

Therefore, to locate the value of x which will make $f(x)$ as large as possible, take $f'(x)$ and set it equal to zero.

$$852 - 6x = 0$$

and so $x = 142$ feet and $y = \frac{1704-6 \times 142}{2} = 426$ feet.

Revenue is defined to be the amount of money obtained in some transaction. Profit is defined as the revenue minus the costs.

Example 3.4.7 Sam, the owner of Spider Sam's Tarantulas and Creepy Critters finds he can sell 6 tarantulas every day at the regular price of \$30 each. At his last spider celebration sale he reduced the price to \$24 and was able to sell 12 tarantulas every day. He has to pay \$.05 per day to exhibit a tarantula and his fixed costs are \$30 per day, mainly to maintain the thousands of tarantulas he keeps on his tarantula breeding farm. What price should he charge to maximize his profit.

He assumes the demand for tarantulas is a linear function of price. Thus if y is the number of tarantulas demanded at price x , it follows $y = 36 - x$. Therefore, the revenue for price x equals $R(x) = (36 - x)x$. Now you have to subtract off the costs to get the profit. Thus

$$P(x) = (36 - x)x - (36 - x)(.05) - 30.$$

It follows the profit is maximized when $P'(x) = 0$ so $-2.0x + 36.05 = 0$ which occurs when $x = \$18.025$. Thus Sam should charge about \$18 per tarantula.

Exercise 3.4.8 Lisa, the owner of Lisa's gags and gadgets sells 500 whoopee cushions per year. It costs \$.25 per year to store a whoopee cushion. To order whoopee cushions it costs \$8 plus \$.90 per cushion. How many times a year and in what lot size should whoopee cushions be ordered to minimize inventory costs?

Let x be the times per year an order is sent for a lot size of $\frac{500}{x}$. If the demand is constant, it is reasonable to suppose there are about $\frac{500}{2x}$ whoopee cushions which have to be stored. Thus the cost to store whoopee cushions is $\frac{125.0}{2x} = .25 \left(\frac{500}{2x} \right)$. Each time an order is made for a lot size of $\frac{500}{x}$ it costs $8 + \frac{450.0}{x} = 8 + .9 \left(\frac{500}{x} \right)$ and this is done x times a year. Therefore, the total inventory cost is $x \left(8 + \frac{450.0}{x} \right) + \frac{125.0}{2x} = C(x)$. The problem is to minimize $C(x) = x \left(8 + \frac{450.0}{x} \right) + \frac{125.0}{2x}$. Taking the derivative yields

$$C'(x) = \frac{16x^2 - 125}{2x^2}$$

and so the value of x which will minimize $C(x)$ is $\frac{5}{4}\sqrt{5} = 2.79\cdots$ and the lot size is $\frac{500}{\frac{5}{4}\sqrt{5}} = 178.88$. Of course you would round these numbers off. Order 179 whoopee cushions 3 times a year.

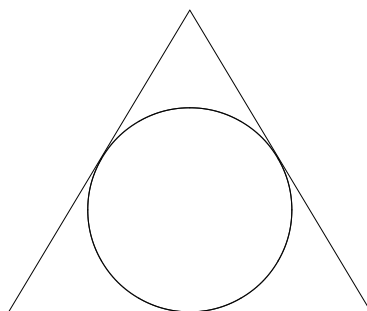
3.5 Exercises

1. If $f'(x) = 0$, is it necessary that x is either a local minimum or local maximum?
Hint: Consider $f(x) = x^3$.
2. Two positive numbers add to 32. Find the numbers if their product is to be as large as possible.
3. The product of two positive numbers equals 16. Find the numbers if their sum is to be as small as possible.
4. The product of two positive numbers equals 16. Find the numbers if twice the first plus three times the second is to be as small as possible.
5. Theodore, the owner of Theodore's tarantulas finds he can sell 6 tarantulas at the regular price of \$20 each. At his last spider celebration day sale he reduced the price to \$14 and was able to sell 14 tarantulas. He has to pay \$.05 per day to maintain a tarantula and his fixed costs are \$30 per day. What price should he charge to maximize his profit.
6. Lisa, the owner of Lisa's gags and gadgets sells 500 whoopee cushions per year. It costs \$.25 per year to store a whoopee cushion. To order whoopee cushions it costs \$2 plus \$.25 per cushion. How many times a year and in what lot size should whoopee cushions be ordered to minimize inventory costs?
7. A continuous function, f defined on $[a, b]$ is to be maximized. It was shown above in Theorem 3.4.2 that if the maximum value of f occurs at $x \in (a, b)$, and if f is differentiable there, then $f'(x) = 0$. However, this theorem does not say anything about the case where the maximum of f occurs at either a or b . Describe how to find the point of $[a, b]$ where f achieves its maximum. Does f have a maximum? Explain.
8. Find the maximum and minimum values and the values of x where these are achieved for the function, $f(x) = x + \sqrt{25 - x^2}$.
9. A piece of wire of length L is to be cut in two pieces. One piece is bent into the shape of an equilateral triangle and the other piece is bent to form a square. How should the wire be cut to maximize the sum of the areas of the two shapes? How should the wire be bent to minimize the sum of the areas of the two shapes?
Hint: Be sure to consider the case where all the wire is devoted to one of the shapes separately. This is a possible solution even though the derivative is not zero there.
10. A cylindrical can is to be constructed of material which costs 3 cents per square inch for the top and bottom and only 2 cents per square inch for the sides. The can needs to hold 90π cubic inches. Find the dimensions of the cheapest can.
Hint: The volume of a cylinder is $\pi r^2 h$ where r is the radius of the base and h is the height. The area of the cylinder is $2\pi r^2 + 2\pi r h$.
11. A rectangular sheet of tin has dimensions 10 cm. by 20 cm. It is desired to make a topless box by cutting out squares from each corner of the rectangular sheet and

then folding the rectangular tabs which remain. Find the volume of the largest box which can be made in this way.

12. Let $f(x) = \frac{1}{3}x^3 - x^2 - 8x$ on the interval $[-1, 10]$. Find the point of $[-1, 10]$ at which f achieves its minimum.
13. Lets find the point on the graph of $y = \frac{x^2}{4}$ which is closest to $(0, 1)$. One way to do it is to observe that a typical point on the graph is of the form $(x, \frac{x^2}{4})$ and then to minimize the function, $f(x) = x^2 + \left(\frac{x^2}{4} - 1\right)^2$. Taking the derivative of f yields $x + \frac{1}{4}x^3$ and setting this equal to 0 leads to the solution, $x = 0$. Therefore, the point closest to $(0, 1)$ is $(0, 0)$. Now lets do it another way. Lets use $y = \frac{x^2}{4}$ to write $x^2 = 4y$. Now for (x, y) on the graph, it follows it is of the form $(\sqrt{4y}, y)$. Therefore, minimize $f(y) = 4y + (y - 1)^2$. Take the derivative to obtain $2 + 2y$ which requires $y = -1$. However, on this graph, y is never negative. What on earth is the problem?
14. A rectangular garden 200 square feet in area is to be fenced off against rabbits. Find the least possible length of fencing if one side of the garden is already protected by a barn.
15. A feed lot is to be enclosed by a fence and divided in 5 pieces by 4 fences parallel to one side. 1272 feet of fencing is used. Find dimensions of the feed lot which will have the largest total area.
16. Find the dimensions of the largest rectangle that can be inscribed in a semicircle of radius 8.
17. Find the dimensions of the largest rectangle that can be inscribed in the ellipse, $\frac{x^2}{9} + \frac{y^2}{4} = 1$.
18. Find the largest volume for a cylinder inscribed in a ball of radius eight cm. The volume of a cylinder equals $\pi r^2 h$ where h is the height and r is the radius.
19. A function, f , is said to be odd if $f(-x) = -f(x)$ and a function is said to be even if $f(-x) = f(x)$. Show that if f is even, then f' is odd and if f is odd, then f' is even. Sketch the graph of a typical odd function and a typical even function.
20. Recall sin is an odd function and cos is an even function. Determine whether each of the trig functions is odd, even or neither.
21. Find the x values of the critical points of the function $f(x) = 3x^2 - 5x^3$.
22. Find the x values of the critical points of the function $f(x) = \sqrt{3x^2 - 6x + 8}$.
23. Find the extreme points of the function, $f(x) = x + \frac{25}{x}$ and tell whether the extreme point is a local maximum or a local minimum or neither.
24. A piece of property is to be fenced on the front and two sides. Fencing for the sides costs \$3.50 per foot and fencing for the front costs \$5.60 per foot. What are the dimensions of the largest such rectangular lot if the available money is \$1400?
25. In a particular apartment complex of 200 units, it is found that all units remain occupied when the rent is \$400 per month. For each \$40 increase in the rent, 5 units become vacant, on the average. Occupied units require \$80 per month for maintenance, while vacant units require none. Fixed costs for the buildings are \$20000 per month. What rent should be charged for maximum profit and what is the maximum profit?

26. Find the point on the curve, $y = \sqrt{25 - 2x}$ which is closest to $(0, 0)$.
27. A street is 200 feet long and there are two lights located at the ends of the street. One of the lights is $\frac{1}{8}$ times as bright as the other. Assuming the brightness of light from one of these street lights is proportional to the brightness of the light and the reciprocal of the square of the distance from the light, locate the darkest point on the street.
28. * Find the volume of the smallest right circular cone which can be circumscribed about a sphere of radius 4 inches.



29. Two cities are located on the same side of a straight river. One city is at a distance of 3 miles from the river and the other city is at a distance of 8 miles from the river. The distance between the two points on the river which are closest to the respective cities is 40 miles. Find the location of a pumping station which is to pump water to the two cities which will minimize the length of pipe used.
30. A hungry spider is located on the wall four feet off the floor directly above a point four feet from the corner of the room. This is a Daring Jumping Spider¹, not a lazy web spinner who just sits in the web and waits for its prey. This kind of spider stalks its dinner like a lion hunting an impala. At a point on the floor which is 6 feet from the wall and 8 feet from the corner is a possible dinner, a plump juicy fly temporarily distracted as it slurps on a succulent morsel of rotting meat. What path should the hungry spider follow? Describe a way to do this problem geometrically without using any calculus.

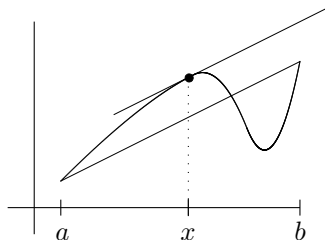
3.6 Mean Value Theorem

The mean value theorem is one of the most important theorems about the derivative. The best versions of many other theorems depend on this fundamental result. The mean value theorem says that under suitable conditions, there exists a point in (a, b) , x , such that $f'(x)$ equals the slope of the secant line joining $(a, f(a))$ and $(b, f(b))$,

$$\frac{f(b) - f(a)}{b - a}.$$

The following picture is descriptive of this situation.

¹These are very beautiful spiders if you don't look too close. They are small furry and black with white markings. You sometimes see them running along walls and ceilings. Like many spiders they can climb even very slippery surfaces like glass with no difficulties.



This theorem is an existence theorem and like the other existence theorems in analysis, it depends on the completeness axiom. The following is known as Rolle's² theorem.

Theorem 3.6.1 Suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous,

$$f(a) = f(b),$$

and

$$f : (a, b) \rightarrow \mathbb{R}$$

has a derivative at every point of (a, b) . Then there exists $x \in (a, b)$ such that $f'(x) = 0$.

Proof: Suppose first that $f(x) = f(a)$ for all $x \in [a, b]$. Then any $x \in (a, b)$ is a point such that $f'(x) = 0$. If f is not constant, either there exists $y \in (a, b)$ such that $f(y) > f(a)$ or there exists $y \in (a, b)$ such that $f(y) < f(b)$. In the first case, the maximum of f is achieved at some $x \in (a, b)$ and in the second case, the minimum of f is achieved at some $x \in (a, b)$. Either way, Theorem 3.4.2 on Page 93 implies $f'(x) = 0$. This proves Rolle's theorem.

The next theorem is known as the Cauchy mean value theorem.

Theorem 3.6.2 Suppose f, g are continuous on $[a, b]$ and differentiable on (a, b) . Then there exists $x \in (a, b)$ such that

$$f'(x)(g(b) - g(a)) = g'(x)(f(b) - f(a)).$$

Proof: Let

$$h(x) \equiv f(x)(g(b) - g(a)) - g(x)(f(b) - f(a)).$$

Then letting $x = a$ and then letting $x = b$, a short computation shows $h(a) = h(b)$. Also, h is continuous on $[a, b]$ and differentiable on (a, b) . Therefore Rolle's theorem applies and there exists $x \in (a, b)$ such that

$$h'(x) = f'(x)(g(b) - g(a)) - g'(x)(f(b) - f(a)) = 0.$$

This proves the theorem.

The usual mean value theorem, sometimes called the Lagrange mean value theorem, illustrated by the above picture is obtained by letting $g(x) = x$.

Corollary 3.6.3 Let f be continuous on $[a, b]$ and differentiable on (a, b) . Then there exists $x \in (a, b)$ such that $f(b) - f(a) = f'(x)(b - a)$.

²Rolle is remembered for Rolle's theorem and not for anything else he did. Ironically, he did not like calculus.

Corollary 3.6.4 Suppose $f'(x) = 0$ for all $x \in (a, b)$ where $a \geq -\infty$ and $b \leq \infty$. Then $f(x) = f(y)$ for all $x, y \in (a, b)$. Thus f is a constant.

Proof: If this is not true, there exists x_1 and x_2 such that $f(x_1) \neq f(x_2)$. Then by the mean value theorem,

$$0 \neq \frac{f(x_1) - f(x_2)}{x_1 - x_2} = f'(z)$$

for some z between x_1 and x_2 . This contradicts the hypothesis that $f'(x) = 0$ for all x . This proves the theorem.

Corollary 3.6.5 Suppose $f'(x) > 0$ for all $x \in (a, b)$ where $a \geq -\infty$ and $b \leq \infty$. Then f is strictly increasing on (a, b) . That is, if $x < y$, then $f(x) < f(y)$. If $f'(x) \geq 0$, then f is increasing in the sense that whenever $x < y$ it follows that $f(x) \leq f(y)$.

Proof: Let $x < y$. Then by the mean value theorem, there exists $z \in (x, y)$ such that

$$0 < f'(z) = \frac{f(y) - f(x)}{y - x}.$$

Since $y > x$, it follows $f(y) > f(x)$ as claimed. Replacing $<$ by $\leq$ in the above equation and repeating the argument gives the second claim.

Corollary 3.6.6 Suppose $f'(x) < 0$ for all $x \in (a, b)$ where $a \geq -\infty$ and $b \leq \infty$. Then f is strictly decreasing on (a, b) . That is, if $x < y$, then $f(x) > f(y)$. If $f'(x) \leq 0$, then f is decreasing in the sense that for $x < y$, it follows that $f(x) \geq f(y)$.

Proof: Let $x < y$. Then by the mean value theorem, there exists $z \in (x, y)$ such that

$$0 > f'(z) = \frac{f(y) - f(x)}{y - x}.$$

Since $y > x$, it follows $f(y) < f(x)$ as claimed. The second claim is similar except instead of a strict inequality in the above formula, you put $\geq$.

3.7 Exercises

1. Sally drives her Saturn over the 110 mile toll road in exactly 1.3 hours. The speed limit on this toll road is 70 miles per hour and the fine for speeding is 10 dollars per mile per hour over the speed limit. How much should Sally pay?
2. Two cars are careening down a freeway weaving in and out of traffic. Car A passes car B and then car B passes car A as the driver makes obscene gestures. This infuriates the driver of car A who passes car B while firing his handgun at the driver of car B. Show there are at least two times when both cars have the same speed. Then show there exists at least one time when they have the same acceleration. The acceleration is the derivative of the velocity.
3. Show the cubic function, $f(x) = 5x^3 + 7x - 18$ has only one real zero.
4. Suppose $f(x) = x^7 + |x| + x - 12$. How many solutions are there to the equation, $f(x) = 0$?
5. Let $f(x) = |x - 7| + (x - 7)^2 - 2$ on the interval $[6, 8]$. Then $f(6) = 0 = f(8)$. Does it follow from Rolle's theorem that there exists $c \in (6, 8)$ such that $f'(c) = 0$? Explain your answer.

6. Suppose f and g are differentiable functions defined on $\mathbb{R}$. Suppose also that it is known that $|f'(x)| > |g'(x)|$ for all x and that $|f'(t)| > 0$ for all t . Show that whenever $x \neq y$, it follows $|f(x) - f(y)| > |g(x) - g(y)|$. **Hint:** Use the Cauchy mean value theorem, Theorem 3.6.2.
7. Show that, like continuous functions, functions which are derivatives have the intermediate value property. This means that if $f'(a) < 0 < f'(b)$ then there exists $x \in (a, b)$ such that $f'(x) = 0$. **Hint:** Argue the minimum value of f occurs at an interior point of $[a, b]$.
8. Consider the function

$$f(x) \equiv \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}.$$

Is it possible that this function could be the derivative of some function? Why?

9. Suppose $a \in I$, an open interval and that a function f , defined on I has $n + 1$ derivatives. Then for each $m \leq n$ the following formula holds for $x \in I$.

$$f(x) = \sum_{k=0}^m f^{(k)}(a) \frac{(x-a)^k}{k!} + f^{(m+1)}(y) \frac{(x-a)^{m+1}}{(m+1)!} \quad (3.13)$$

where y is some point between x and a . Note that if $n = 0$, this reduces to the Lagrange form of the mean value theorem so the formula holds if $m = 0$. Suppose it holds for some $0 \leq m-1 < n$. The task is to show then that it also holds for m . It will then follow that the above formula will hold for all $m \leq n$ as claimed. This formula is very important. The last term is called the Lagrange form of the remainder in Taylor series. It will be discussed carefully later. For now, try to prove the formula using the following steps. If the formula holds for $m-1$, then you can apply it to f' .

$$f'(x) = \sum_{k=0}^{m-1} \frac{f^{(k+1)}(a)}{k!} (x-a)^k + \frac{f^{(m+1)}(y)}{m!} (x-a)^m. \quad (3.14)$$

Now let $g(x) = f(x) - \sum_{k=0}^m \frac{f^{(k)}(a)}{k!} (x-a)^k$ and let $h(x) = (x-a)^{m+1}$. Now from the Cauchy mean value theorem there exists z between x and a such that

$$\frac{g(x) - g(a)}{h(x) - h(a)} = \frac{g'(z)}{h'(z)} = \frac{f'(z) - \sum_{k=1}^m \frac{f^{(k)}(a)}{(k-1)!} (z-a)^{k-1}}{(m+1)(z-a)^m}.$$

Now explain why $\sum_{k=1}^m \frac{f^{(k)}(a)}{(k-1)!} (z-a)^{k-1} = \sum_{k=0}^{m-1} \frac{f^{(k+1)}(a)}{k!} (z-a)^k$ and then use 3.14. Explain why this implies

$$\frac{f(x) - \sum_{k=0}^m \frac{f^{(k)}(a)}{k!} (x-a)^k}{(x-a)^{m+1}} = \frac{\frac{f^{(m+1)}(y)}{m!} (z-a)^m}{(m+1)(z-a)^m}$$

which yields the desired formula for m . A way to think of 3.13 is as a generalized mean value theorem. Note that the key result which made it work was the Cauchy mean value theorem.

3.8 Curve Sketching

The theorems and corollaries given above can be used to aid in sketching the graphs of functions. The second derivative will also help in determining the shape of the function.

Definition 3.8.1 A differentiable function, f , defined on an interval, (a, b) , is concave up if f' is an increasing function. A differentiable function defined on an interval, (a, b) , is concave down if f' is a decreasing function. A point where the graph of the function changes from being concave up to concave down or from concave down to concave up is called an inflection point.

From the geometric description of the derivative as the slope of a tangent line to the graph of the function, to say the derivative is an increasing function means that as you move from left to right, the slopes of the lines tangent to the graph of f become larger. Thus the graph of the function is bent up in the shape of a smile. It may also help to think of it as a cave when you view it from above, hence the term concave up. If the derivative is decreasing, it follows that as you move from left to right the slopes of the lines tangent to the graph of f become smaller. Thus the graph of the function is bent down in the form of a frown. It is concave down because it is like a cave when viewed from beneath. The following theorem will give a convenient criterion in terms of the second derivative for finding whether a function is concave up or concave down. The term, concavity, is used to refer to this property. Thus you determine the concavity of a function when you find whether it is concave up or concave down.

Theorem 3.8.2 Suppose $f''(x) > 0$ for $x \in (a, b)$. Then f is concave up on (a, b) . Suppose $f''(x) < 0$ on (a, b) . Then f is concave down.

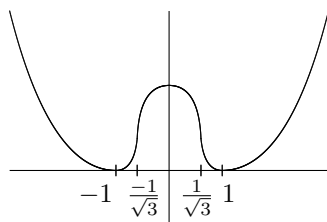
Proof: This follows immediately from Corollaries 3.6.6 and 3.6.5 applied to the first derivative. The following picture may help in remembering this.



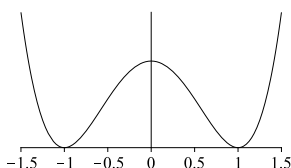
In this picture, the plus signs and the smile on the left correspond to the second derivative being positive. The smile gives the way in which the graph of the function is bent. In the second face, the minus signs correspond to the second derivative being negative. The frown gives the way in which the graph of the function is bent.

Example 3.8.3 Sketch the graph of the function, $f(x) = (x^2 - 1)^2 = x^4 - 2x^2 + 1$

Take the derivative of this function, $f'(x) = 4x^3 - 4x = 4x(x - 1)(x + 1)$ which equals zero at $-1, 0$, and 1 . It is positive on $(-1, 0)$, and $(1, \infty)$ and negative on $(0, 1)$ and $(-\infty, -1)$. Therefore, $x = 0$ corresponds to a local maximum and $x = -1$ and $x = 1$ correspond to local minima. The second derivative is $f''(x) = 12x^2 - 4$ and this equals zero only at the points $-1/\sqrt{3}$ and $1/\sqrt{3}$. The second derivative is positive on the intervals $(1/\sqrt{3}, \infty)$ and $(-\infty, -1/\sqrt{3})$ so the function, f is smiling on these intervals. The second derivative is negative on the interval $(-1/\sqrt{3}, 1/\sqrt{3})$ and so the original function is frowning on this interval. This describes in words the qualitative shape of the function. It only remains to draw a picture which incorporates this description. The following is such a sketch. It is not intended to be an accurate drawing made to scale, only to be a qualitative picture of what was just determined.



A better graph of this function is the following, done by a computer algebra system. However the computer worked a lot harder.



In general, if you are interested in getting a nice graph of a function, you should use a computer algebra system. An effective way to accomplish your graphing is to go to the help menu and copy and paste an example from this menu changing it as needed. Both mathematica and Maple have good help menus. Keep in mind there are certain conventions which must be followed. For example to write x raised to the second power you enter x^2 . In Maple, you also need to place an asterisk between quantities which are multiplied since otherwise it will not know you are multiplying and won't work. There are also easy to use versions of Maple available which involve essentially pointing and clicking. You won't learn any calculus from playing with a computer algebra system but you might have a lot of fun.

3.9 Exercises

1. Sketch the graph of the function, $f(x) = x^3 - 3x + 1$ showing the intervals on which the function is concave up and down and identifying the intervals on which the function is increasing.
2. Find intervals on which the function, $f(x) = \sqrt{1 - x^2}$ is increasing and intervals on which it is concave up and concave down. Sketch a graph of the function.
3. Sketch the graphs of $y = x^4$, $y = x^3$, and $y = -x^4$. What do these graphs tell you about the case when the second derivative equals zero?
4. Sketch the graph of $f(x) = 1/(1 + x^2)$ showing the intervals on which the function is increasing or decreasing and the intervals on which the graph is concave up and concave down.
5. Sketch the graph of $f(x) = x/(1 + x^2)$ showing the intervals on which the function is increasing or decreasing and the intervals on which the graph is concave up and concave down.
6. Show that inflection points can be identified by looking at those points where the second derivative equals zero but that not every point where the second derivative equals zero is an inflection point. **Hint:** For the last part consider $y = x^3$ and $y = x^4$.

7. Suppose $f''(x) = 0$ and $f'''(x) \neq 0$. Does it follow that x must be an inflection point?
8. Find all inflection points for the function, $f(x) = x^2 / (1 + x^2)$.

Some Important Special Functions

4.0.1 Outcomes

1. Understand the definition and derivatives of the circular functions.
2. Use the rules of exponents and find derivatives of the exponential functions and the logarithm functions.
3. Understand the number, e .

4.1 The Circular Functions

The Trigonometric functions are also called the circular functions. Thus this section will be on the functions, $\cos$, $\sin$, $\tan$, $\sec$, $\csc$, and $\cot$. The first thing to do is to give an important lemma. There are several approaches to this lemma. To see it done in terms of areas of a circular sector, see Apostol, [1], or almost any other calculus book. However, the book by Apostol has no loose ends in the presentation unlike most other books which use this approach. The proof given here is a modification of that found in Tierney, [27] and Rose, [22] and is based on arc length.

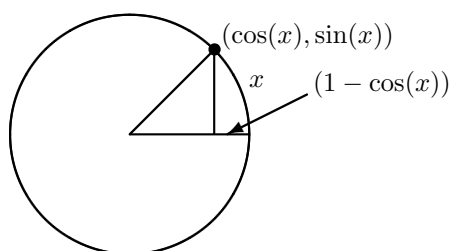
Lemma 4.1.1 *The following limits hold.*

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = 1 \quad (4.1)$$

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x} = 0 \quad (4.2)$$

Proof: First consider 4.1. In the following picture, it follows from Corollary 1.4.4 on Page 24 that for small positive x ,

$$\sin x + (1 - \cos x) \geq x \geq \sin x. \quad (4.3)$$



Now divide by $\sin x$ to get

$$1 + \frac{1 - \cos x}{|\sin x|} = 1 + \frac{1 - \cos x}{\sin x} \geq \frac{x}{\sin x} \geq 1.$$

For small negative values of x , it is also true that

$$1 + \frac{1 - \cos x}{|\sin x|} \geq \frac{x}{\sin x} \geq 1.$$

(Why?) From the trig. identities, it follows that for all small values of x ,

$$1 + \frac{\sin^2 x}{|\sin x|(1 + \cos x)} = 1 + \frac{|\sin x|}{(1 + \cos x)} \geq \frac{x}{\sin x} \geq 1$$

and so from the squeezing theorem, Theorem 2.9.5 on Page 60,

$$\lim_{x \rightarrow 0} \frac{x}{\sin x} = 1$$

and consequently, from the limit theorems,

$$\lim_{x \rightarrow 0} \frac{\sin x}{x} = \lim_{x \rightarrow 0} \frac{1}{\left(\frac{x}{\sin x}\right)} = 1.$$

Finally,

$$\frac{1 - \cos x}{x} = \frac{1 - \cos^2 x}{x(1 + \cos x)} = \sin x \frac{\sin x}{x} \frac{1}{1 + \cos x}.$$

Therefore, from Theorem 2.5.1 on Page 54 which says $\lim_{x \rightarrow 0} \sin(x) = 0$, and the limit theorems,

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x} = 0.$$

This proves the Lemma.

With this, it is easy to find the derivative of $\sin$. Using Lemma 4.1.1,

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\sin(x+h) - \sin x}{h} &= \lim_{h \rightarrow 0} \frac{\sin(x) \cos(h) + \cos(x) \sin(h) - \sin x}{h} \\ &= \lim_{h \rightarrow 0} \frac{(\sin x)(\cos(h) - 1)}{h} + \cos x \frac{\sin(h)}{h} \\ &= \cos x. \end{aligned}$$

The derivative of $\cos$ can be found the same way. Alternatively,

$$\cos(x) = \sin(x + \pi/2)$$

and so

$$\begin{aligned} \cos'(x) &= \sin'(x + \pi/2) \\ &= \cos(x + \pi/2) \\ &= \cos x \cos(\pi/2) - \sin x \sin(\pi/2) \\ &= -\sin x. \end{aligned}$$

The following theorem is now obvious and the proofs of the remaining parts are left for you.

Theorem 4.1.2 *The derivatives of the trig. functions are as follows.*

$$\begin{aligned}\sin'(x) &= \cos x \\ \cos'(x) &= -\sin x \\ \tan'(x) &= \sec^2(x) \\ \cot'(x) &= -\csc^2(x) \\ \sec'(x) &= \sec x \tan x \\ \csc'(x) &= -\csc x \cot x\end{aligned}$$

Here are some examples of extremum problems which involve the use of the trig. functions.

Example 4.1.3 *Two hallways intersect at a right angle. One is 5 feet wide and the other is 2 feet wide. What is the length of the longest thin rod which can be carried horizontally from one hallway to the other?*

You must minimize the length of the rod which touches the inside corner of the two halls and extends to the outside walls. Letting θ be the angle between this rod and the outside wall for the hall having width 2, minimize

$$f(\theta) = \overbrace{2 \csc \theta + 5 \sec \theta}^{\text{length of rod}}.$$

Therefore, using the rules of differentiation,

$$f'(\theta) = \frac{2 \cos^3 \theta - 5 \sin \theta + 5 \sin \theta \cos^2 \theta}{(\cos^2 \theta)(-1 + \cos^2 \theta)} = 0$$

should be solved to get the angle where this length is as small as possible. Thus

$$2 \cos^3 \theta - 5 \sin \theta + 5 \sin \theta \cos^2 \theta = 0.$$

and $2 \cos^3 \theta - 5 \sin^3 \theta = 0$ and so $\tan \theta = \frac{1}{5} \sqrt[3]{2} (\sqrt[3]{5})^2$. Drawing a triangle, you see that at this value of θ , you have $\sec \theta = \frac{\sqrt{(\sqrt[3]{5})^2 + (\sqrt[3]{2})^2}}{\sqrt[3]{5}}$ and $\csc \theta = \frac{\sqrt{(\sqrt[3]{5})^2 + (\sqrt[3]{2})^2}}{\sqrt[3]{2}}$. Therefore, the minimum is obtained by substituting these values in to the equation for $f(\theta)$ yielding $\left(\sqrt{((\sqrt[3]{5})^2 + (\sqrt[3]{2})^2)} \right)^3$.

Example 4.1.4 *A fence 9 feet high is 2 feet from a building. What is the length of the shortest ladder which will lean against the top of the fence and touch the building?*

Let θ be the angle of the ladder with the ground. Then the length of this ladder making this angle with the ground and leaning on the top of the fence while touching the building is

$$f(\theta) = \frac{9}{\sin \theta} + \frac{2}{\cos \theta}.$$

Then the final answer is $\left(\sqrt{3\sqrt[3]{3} + (\sqrt[3]{2})^2} \right)^3$. The details are similar to the problem of the two hallways.

4.2 Exercises

1. Prove all parts of Theorem 4.1.2.
2. Prove $\tan'(x) = 1 + \tan^2(x)$.
3. Find and prove a formula for the derivative of $\sin^m(x)$ for m an integer.
4. Find the derivative of the function, $\sin^6(5x)$.
5. Find the derivative of the function, $\tan^7(4x)$.
6. Find the derivative of the function, $\frac{\sec^3(2x)}{\tan^3(3x)}$.
7. Find all intervals where $\sin(2x)$ is concave down.
8. Find the intervals where $\cos(3x)$ is increasing.
9. Two hallways intersect at a right angle. One is 3 feet wide and the other is 4 feet wide. What is the length of the longest thin rod which can be carried horizontally from one hallway to the other?
10. A fence 5 feet high is 2 feet from a building. What is the length of the shortest ladder which will lean against the top of the fence and touch the building?
11. Suppose $f(x) = A \cos \omega x + B \sin \omega x$. Show there exists an angle, ϕ such that $f(x) = \sqrt{A^2 + B^2} \sin(\omega x + \phi)$. The number, $\sqrt{A^2 + B^2}$ gives the “amplitude” and ϕ is called the “phase shift” while ω is called the “frequency”. This is very important because it is descriptive of what is going on. The amplitude gives the height of the periodic function, f . **Hint:** Remember a point on the unit circle determines an angle. Write $f(x)$ in the form

$$\sqrt{A^2 + B^2} \left(\frac{A}{\sqrt{A^2 + B^2}} \cos \omega x + \frac{B}{\sqrt{A^2 + B^2}} \sin \omega x \right)$$

and note that $\left(\frac{B}{\sqrt{A^2 + B^2}}, \frac{A}{\sqrt{A^2 + B^2}} \right)$ is a point on the unit circle.

12. Repeat Problem 11 but this time show $f(x) = \sqrt{A^2 + B^2} \cos(\omega x + \phi)$. How could you find ϕ ?
13. A square picture is 6 feet high and is fastened to the wall with its lowest edge one foot above the eye level of an observer. Where should he stand to maximize the angle subtended by the picture at his eye?
14. A 100 foot tall lamp post with a light on top creates a shadow from a falling ball dropped from a height of 100 feet at a distance of 50 feet from the lamp post. Thus the distance the ball has fallen at time t is $16t^2$. Find the velocity of the shadow when the ball has dropped a distance of 64 feet.

4.3 The Exponential And Log Functions

4.3.1 The Rules Of Exponents

As mentioned earlier, b^m means to multiply b by itself m times assuming m is a positive integer. $b^0 \equiv 1$ provided $b \neq 0$. In the case where $b = 0$ the symbol is undefined. If

$m < 0$, b^m is defined as $\frac{1}{b^{-m}}$. Then the following algebraic properties are obtained. Be sure you understand these properties for x and y integers.

$$b^{x+y} = b^x b^y, (ab)^x = a^x b^x \quad (4.4)$$

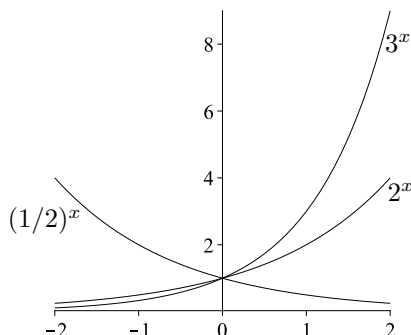
$$b^{xy} = (b^x)^y, b^{-1} = \frac{1}{b} \quad (4.5)$$

These properties are called the rules of exponents.

When x and y are not integers, the meaning of b^x is no longer clear. For example, suppose $b = -1$ and $x = 1/2$. What exactly is meant by $(-1)^{1/2}$? Even in the case where $b > 0$ there are difficulties. If x is a rational number, m/n and $b > 0$ the symbol $b^{m/n}$ means $\sqrt[n]{b^m}$. That is its definition and it is a useful exercise for you to verify 4.4 and 4.5 hold with this definition. There are no mathematical questions about the existence of this number. To see this, consider Problem 7 on Page 58. The problem is not one of theory but of practicality. Could you use this definition to find $2^{\frac{1234567812345}{1234567812344}}$? Consider what you would do. First find the number $2^{\frac{1234567812345}{1234567812344}}$ and then $\dots$? Can you find this number? It is just too big. However, a calculator can find $2^{\frac{1234567812345}{1234567812344}}$. It yields $2^{\frac{1234567812345}{1234567812344}} = 2.000\,000\,000\,001\,123$ as an approximate answer. Clearly something else must be going on. To make matters even worse, what would you do with $2^{\sqrt{2}}$? As mentioned earlier, $\sqrt{2}$ is irrational and so cannot be written as the quotient of two integers. These are serious difficulties and must be dealt with.

4.3.2 The Exponential Functions, A Wild Assumption

Using your calculator or a computer you can obtain graphs of the functions, $y = b^x$ for various choices of b . The following picture gives a few of these graphs.



These graphs suggest that if $b < 1$ the function, $y = b^x$ is decreasing while if $b > 1$, the function is increasing but just how was the calculator or computer able to draw those graphs? Also, do the laws of exponents continue to hold for all real values of x ? The short answer is that they do and this is shown later but for now here is a wild assumption which glosses over these issues.



Wild Assumption 4.3.1 *For every $b > 0$ there exists a unique differentiable function $\exp_b(x) \equiv b^x$ valid for all real values of x such that 4.4 and 4.5 both hold for all $x, y \in \mathbb{R}$, $\exp_b(m/n) = \sqrt[n]{b^m}$ whenever m, n are integers, (In particular, $b^0 = 1$.) and $b^x > 0$ for all $x \in \mathbb{R}$. Furthermore, if $b \neq 1$ and $h \neq 0$, then $\exp_b(h) = b^h \neq 1$.*

Instead of writing $\exp_b(x)$ I will often write b^x and I will also be somewhat sloppy and regard b^x as the name of a function and not just as $\exp_b(x)$, a given function defined at x . This is done to conform with usual usage. Also, the last claim in Wild Assumption 4.3.1 follows from the first part of this assumption. See Problem 1. I want it to be completely clear that the Wild Assumption is just that. No reason for believing in such an assumption has been given notwithstanding the pretty pictures drawn by the calculator. Later in the book, the wild assumption will be completely justified. Based on Wild Assumption 4.3.1 one can easily find out all about b^x . In the following theorem, the number $\ln b$ defined there is the slope of the tangent line to the graph of $y = b^x$ at the point $(0, 1)$.

Theorem 4.3.2 *Let $\exp_b$ be defined in Wild Assumption 4.3.1 for $b > 0$. Then there exists a unique number, denoted by $\ln b$ for $b > 0$ satisfying*

$$\exp'_b(x) = (\ln b) \exp_b(x). \quad (4.6)$$

Furthermore,

$$\ln(ab) = \ln(a) + \ln(b), \quad \ln 1 = 0, \quad (4.7)$$

and for all $y \in \mathbb{R}$,

$$\ln(b^y) = y \ln b. \quad (4.8)$$

$$\ln\left(\frac{a}{b}\right) = \ln(a) - \ln(b) \quad (4.9)$$

The function, $x \rightarrow \ln x$ is differentiable and defined for all $x > 0$ and

$$\ln'(x) = \frac{1}{x}. \quad (4.10)$$

The function, $x \rightarrow \ln x$ is one to one on $(0, \infty)$. Also, $\ln$ maps $(0, \infty)$ onto $(-\infty, \infty)$.

Proof: First consider 4.6.

$$\lim_{h \rightarrow 0} \frac{\exp_b(x+h) - \exp_b(x)}{h} = \lim_{h \rightarrow 0} \frac{b^{x+h} - b^x}{h} = \lim_{h \rightarrow 0} \left(\frac{b^h - 1}{h} \right) b^x.$$

The expression, $\lim_{h \rightarrow 0} \left(\frac{b^h - 1}{h} \right)$ is assumed to exist thanks to Wild Assumption 4.3.1 and this is denoted by $\ln b$. This proves 4.6.

To verify 4.7, if $b = 1$ then $b^x = 1^x$ for all $x \in \mathbb{R}$. Now by 4.4 and 4.5,

$$1^x 1^x = 1^{x+x} = (1^2)^x = 1^x$$

and so, dividing both sides by 1^x , an operation justified by Wild Assumption 4.3.1, $1^x = 1$ for all $x \in \mathbb{R}$. Therefore, $\exp_1(x) = 1$ for all x and so $\exp'_1(x) = \ln 1 \exp_1(x) = 0$. Thus $\ln 1 = 0$ as claimed. Next, by the product rule and Wild Assumption 4.3.1,

$$\begin{aligned} \ln(ab) (ab)^x &= ((ab)^x)' = (a^x b^x)' = (a^x)' b^x + a^x (b^x)' \\ &= (\ln a) a^x b^x + (\ln b) b^x a^x = [\ln a + \ln b] (ab)^x. \end{aligned}$$

Therefore, $\ln(ab) = \ln a + \ln b$ as claimed.

Next consider 4.8. Keeping y fixed, consider the function $x \rightarrow b^{xy} = (b^y)^x$. Then,

$$\ln(b^y) \exp_b(xy) = \ln(b^y) (b^y)^x = ((b^y)^x)' = g'(x)$$

where $g(x) \equiv b^{xy}$. Using 3.7 on Page 88, it is also the case that

$$g'(x) = y \exp_b'(xy)$$

and so

$$y \ln(b) \exp_b(xy) = y \exp_b'(xy) = \ln(b^y) \exp_b(xy)$$

Now dividing both sides by $\exp_b(xy)$ verifies 4.8.

To obtain 4.9 from this, note

$$\ln\left(\frac{a}{b}\right) = \ln(ab^{-1}) = \ln(a) + \ln(b^{-1}) = \ln(a) - \ln(b).$$

It remains to verify 4.10. From 3.2 and the continuity of 2^x ,

$$\ln'(1) = \lim_{h \rightarrow 0} \frac{\ln(2^h) - \ln 1}{2^h - 1} = \lim_{h \rightarrow 0} \frac{h \ln 2}{2^h - 1} = \frac{\ln 2}{\ln 2} = 1.$$

$\ln 2 \neq 0$ because if it were, then $(2^x)' = (\ln 2) 2^x = 0$ and by Corollary 3.6.4, this would imply 2^x is a constant function which it is not. Now the first part of this lemma implies

$$\ln'(x) = \lim_{y \rightarrow x} \frac{\ln y - \ln x}{y - x} = \lim_{y \rightarrow x} \frac{1}{x} \frac{\ln\left(\frac{y}{x}\right) - \ln 1}{\left(\frac{y}{x}\right) - 1} = \frac{1}{x} \ln'(1) = \frac{1}{x}.$$

It remains to verify $\ln$ is one to one. Suppose $\ln x = \ln y$. Then by the mean value theorem, there exists t between x and y such that $(1/t)(x - y) = \ln x - \ln y = 0$. Therefore, $x = y$ and this shows $\ln$ is one to one as claimed.

It only remains to verify that $\ln$ maps $(0, \infty)$ onto $(-\infty, \infty)$. By Wild Assumption 4.3.1 and the mean value theorem, Corollary 3.6.3, there exists $y \in (0, 1)$ such that

$$0 < \frac{2^1 - 1}{1} = \ln(2) 2^y$$

Since $2^y > 0$ it follows $\ln(2) > 0$ and so $x \rightarrow 2^x$ is strictly increasing. Therefore, by Corollary 3.6.3

$$\frac{2^1 - 1}{1} = \ln(2) 2^y \leq \ln(2) 2^1$$

and it follows that

$$\frac{1}{2} \leq \ln 2. \quad (4.11)$$

Also, from 4.7

$$0 = \ln(2) + \ln\left(\frac{1}{2}\right) \geq \frac{1}{2} + \ln\left(\frac{1}{2}\right)$$

which shows that

$$\ln\left(\frac{1}{2}\right) \leq -\frac{1}{2}. \quad (4.12)$$

It follows from 4.11 and 4.12 that $\ln$ achieves values which are arbitrarily large and arbitrarily large in the negative direction. Therefore, by the intermediate value theorem, $\ln$ achieves all values.

More precisely, let $y \in \mathbb{R}$. Then choose n large enough that $\frac{n}{2} > y$ and $-\frac{n}{2} < y$. Then from 4.11 and 4.12

$$\ln\left(\left(\frac{1}{2}\right)^n\right) \leq \frac{-n}{2} < y < \frac{n}{2} < \ln(2^n).$$

By the intermediate value theorem, there exists $x \in \left(\left(\frac{1}{2}\right)^n, 2^n\right)$ such that $\ln x = y$. This proves the theorem.

Example 4.3.3 Find $f'(x)$ if $f(x) = 7^{2x}$.

$7^{2x} = (7^x)^2$ and so the derivative is $2(7^x)(7^x)' = 2(7^x) \ln 7 (7^x) = 2(\ln 7) 7^{2x}$.

Example 4.3.4 Find $f'(x)$ if $f(x) = 7^{(x^2)}$.

Set up the difference quotient,

$$\frac{7^{(x+h)^2} - 7^{x^2}}{h} = 7^{x^2} \left(\frac{7^{2xh+h^2} - 1}{h} \right) = 7^{x^2} \left(\frac{7^{2xh+h^2} - 1}{2xh + h^2} \right) \frac{2xh + h^2}{h}.$$

Taking the limit as $h \rightarrow 0$ and using the definition of $\ln 7$, this limit equals $7^{x^2} \ln(7) 2x$.

4.3.3 The Special Number, e

Since $\ln$ is one to one onto $\mathbb{R}$, it follows there exists a unique number, e such that $\ln(e) = 1$. Therefore,

$$\exp'_e(x) \equiv (e^x)' = \ln(e) e^x \equiv \ln(e) \exp_e(x) = \exp_e(x)$$

showing that $\exp_e$ has the remarkable property that it equals its own derivative. This wonderful number is called Euler's number and it can be shown to equal approximately 2.7183. It is customary to write $\exp(x)$ for $\exp_e(x)$. Thus

$$\exp'(x) = \exp(x). \quad (4.13)$$

4.3.4 The Function $\ln|x|$

The function, $\ln$ is only defined on positive numbers. However, it is possible to write $\ln|x|$ whenever $x \neq 0$. What is the derivative of this function?

Corollary 4.3.5 Let $f(x) = \ln|x|$ for $x \neq 0$. Then

$$f'(x) = \frac{1}{x}.$$

Proof: If $x > 0$ the formula is just 4.10. Suppose then that $x < 0$. Then $\ln|x| = \ln(-x)$ so by 3.7 on Page 88,

$$(\ln|x|)' = (\ln(-x))' = (\ln((-1)x))' = \frac{1}{-x}(-1) = \frac{1}{x}.$$

This proves the corollary.

4.3.5 Logarithm Functions

Next a new function called $\log_b$ will be defined.

Definition 4.3.6 For all $b > 0$ and $b \neq 1$

$$\log_b(x) \equiv \frac{\ln x}{\ln b}. \quad (4.14)$$

Notice this definition implies 4.7 - 4.9 all hold with $\ln$ replaced with $\log_b$. Also, $\log_b : (0, \infty) \rightarrow \mathbb{R}$ is one to one and onto.

The fundamental relationship between the exponential function, b^x and $\log_b x$ is in the following proposition. This proposition shows this new function is $\log_b$ you may have studied in high school.

Proposition 4.3.7 *Let $b > 0$ and $b \neq 1$. Then for all $x > 0$,*

$$b^{\log_b x} = x, \quad (4.15)$$

and for all $y \in \mathbb{R}$,

$$\log_b b^y = y, \quad (4.16)$$

Also,

$$\log'_b(x) = \frac{1}{\ln b} \frac{1}{x}. \quad (4.17)$$

Proof: Formula 4.15 follows from 4.8.

$$\ln(b^{\log_b x}) = \log_b x \ln b = \ln x$$

and so, since $\ln$ is one to one, it follows 4.15 holds.

$$\log_b b^y \equiv \frac{\ln(b^y)}{\ln b} = \frac{y \ln b}{\ln b} = y$$

and this verifies 4.16. Formula 4.17 is obvious from 4.14.

The functions, $\log_b$ are only defined on positive numbers. However, it is possible to write $\log_b |x|$ whenever $x \neq 0$. What is the derivative of these functions?

Corollary 4.3.8 *Let $f(x) = \log_b |x|$ for $x \neq 0$. Then*

$$f'(x) = \frac{1}{(\ln b)x}.$$

Proof: If $x > 0$ the formula is just 4.17. Suppose then that $x < 0$. Then $\log_b |x| = \log_b (-x)$ so by 3.7 on Page 88,

$$\begin{aligned} (\log_b |x|)' &= (\log_b (-x))' = (\log_b ((-1)x))' \\ &= \frac{1}{-x \ln b} (-1) = \frac{1}{x \ln b}. \end{aligned}$$

This proves the corollary.

Example 4.3.9 *Using properties of logarithms, simplify the expression, $\log_3 (\frac{1}{9})$.*

From 4.7 - 4.9,

$$\log_3 \left(\frac{1}{9} \right) = \log_3 \left(\frac{1}{9} \right) + \log_3 (x) = \log_3 (3^{-2}) + \log_3 (x) = -2 + \log_3 (x).$$

Example 4.3.10 *Using properties of logarithms, solve $5^{x-1} = 3^{2x+2}$.*

Take $\ln$ of both sides. Thus $(x-1) \ln 5 = (2x+2) \ln 3$. Then solving this for x yields $x = \frac{\ln 5 + 2 \ln 3}{\ln 5 - 2 \ln 3}$.

Example 4.3.11 *Solve $\log_3 (x) + 2 = \log_9 (x+3)$.*

From the given equation,

$$3^{\log_3(x)+2} = 3^{\log_9(x+3)} = 9^{\frac{1}{2}(\log_9(x+3))} = 9^{\log_9 \sqrt{x+3}}$$

and so $9x = \sqrt{x+3}$. Therefore, $x = \frac{1+\sqrt{1+12 \times 81}}{2(81)} = \frac{1}{162} + \frac{1}{162} \sqrt{973}$. In the use of the quadratic formula, only one solution was possible. (Why?)

Example 4.3.12 Compare $\ln(x)$ and $\log_e(x)$

Recall that

$$\log_e(x) \equiv \frac{\ln x}{\ln e}.$$

Since $\ln e = 1$ from the definition of e , it follows $\log_e(x) = \ln x$. These logarithms are called natural logarithms.

Example 4.3.13 Find the derivative of the function, $f(x) = \log_5(x)$.

$$f'(x) = \left(\frac{\ln(x)}{\ln 5} \right)' = \frac{1}{\ln 5} \frac{1}{x}.$$

Example 4.3.14 Find the derivative of $f(x) = \ln(x^3)$.

$$f'(x) = (\ln(x^3))' = (3 \ln x)' = \frac{3}{x}.$$

4.4 Exercises

1. Prove the last part of the Wild Assumption follows from the first part of this assumption. That is, show that if $b \neq 1$, then $\exp_b(h) \neq 1$ if $h \neq 0$ follows from the first part. **Hint:** If $b^h = 1$ for $h \neq 0$, show $b^x = 1$ for all $x \in \mathbb{R}$.
2. Simplify
 - (a) $\log_4(16x)$.
 - (b) $\log_3(27x^3)$
 - (c) $(\log_b a)(\log_a b)$ for a, b positive real numbers not equal to 1.
3. Simplify
 - (a) $\log_3(\log_3(27) - \log_3(9))$
 - (b) $\log_2(3) \log_3(2)$
 - (c) $\log_b(x) \log_a(b)$
 - (d) $\log_{10}(100000^{1/3})$
4. Find the derivatives of the following functions. You may want to do this by looking at the definition of the derivative in some cases.
 - (a) $\log_5(x^2)$
 - (b) $\log_3(5x + 1)$
 - (c) $\log_6(\sqrt{x})$
 - (d) $\log_3(\sqrt{x^2 + 1})$.
5. Find the derivatives of the following functions. You may want to do this by looking at the definition of the derivative in some cases.
 - (a) 3^{x^2}
 - (b) 2^{3x+1}
 - (c) 5^{2x+7}

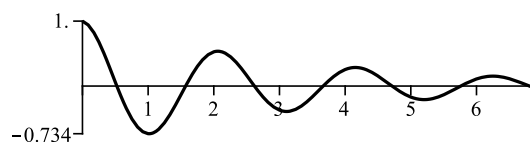
- (d) $7x^3$
 (e) $3\sqrt{x^2+1}$
6. Explain why the function $6^x(1 - x \ln 6)$ is never larger than 1. **Hint:** Consider $f(x) = 6^x(1 - x \ln 6)$ and find its maximum value.
 7. Solve $\log_2(x) + 3 = \log_2(3x + 8)$.
 8. Solve $\log_4(x) + 3 = \log_2(x + 8)$.
 9. Solve the equation $5^{2x+9} = 7^x$ in terms of logarithms.
 10. Using properties of logarithms, simplify the expression, $\log_4\left(\frac{1}{64}x\right)$.
 11. Using properties of logarithms, solve $4^{x-1} = 3^{2x+2}$.
 12. The Wild Assumption gave the existence of a function, b^x satisfying certain properties. Show there can be no more than one such function. **Hint:** Recall the rational numbers were dense in $\mathbb{R}$ and so one can obtain a rational number arbitrarily close to a given real number. Exploit this and the assumed continuity of $\exp_b$ to obtain uniqueness.
 13. Prove the function, b^x is concave up and $\log_b(x)$ is concave down whenever $b > 1$.
 14. Using properties of logarithms and exponentials, solve $3 + \ln(-3x) = 4 + \ln 3x^2$.
 15. Let f be a differentiable function and suppose $f(a) \geq 0$ and that $f'(x) \geq 0$ for $x \geq a$. Show that $f(x) \geq f(a)$ for all $x \geq a$. **Hint:** Use the mean value theorem.
 16. Let e be defined by $\ln(e) = 1$ and suppose $e < x < y$. Find a relationship between x^y and y^x . **Hint:** Use Problem 15 and at some point consider the function $h(x) = \frac{\ln x}{x}$.
 17. Suppose f is any function defined on the positive real numbers and $f'(x) = g(x)$ where g is an odd function. ($g(-x) = -g(x)$.) Show $(f(|x|))' = g(x)$.
 18. You know $(e^x)' = e^x$. Use the definition of the derivative to verify that $(e^{ax})' = ae^{ax}$. Show that Ae^{ax} solves the differential equation, $y' = ay$ along with the initial condition, $y(0) = A$.
 19. Suppose $y' = ay$ and $y(0) = A$. Does it follow that $y = Ae^{ax}$? You know from Problem 18 that Ae^{ax} is one possibility. Is it the only possibility? **Hint:** Consider $y(x)e^{-ax}$ and use the product rule and Problem 18 to show $(y(x)e^{-ax})' = 0$. Then review the mean value theorem. What does the mean value theorem say about functions whose derivatives are always equal to zero?
 20. Show from the definition of the derivative that

$$(\sin(ax))' = a \cos(ax), \quad (\cos(ax))' = -a \sin(ax).$$

With this and Problem 18 find the derivatives of the following functions using the rules of derivatives. You will need to use the product and maybe the quotient rule along with properties of exponents.

- (a) $\sin(3x)5^{6x}$
- (b) $\tan(2x)3^{3x+1}$
- (c) $\sin(2x)\cos(3x)\tan(4x)2^{3x}$

21. Using Problems 18 and 20 try to determine real numbers, b and a such that $y(t) = e^{bt} \cos at$ and $y(t) = e^{bt} \sin at$ both are solutions of the differential equation $y'' + 2\alpha y' + \beta^2 y = 0$ given that $\alpha^2 - \beta^2 < 0$. This is the equation of damped oscillation and will be discussed more carefully later. Typically y measures some sort of displacement from an equilibrium position and t represents time. Here is a picture of the graph of $y(t) = e^{-.3t} \cos(3t)$. This is called damped oscillation.



Properties And Applications Of Derivatives

5.0.1 Outcomes

1. Understand the proof of the chain rule and use the chain rule to differentiate composite functions.
2. Use the chain rule to find the derivative of an inverse function and to do implicit differentiation.
3. Understand the proof that inverse functions of differentiable functions can also be differentiated.
4. Understand the function x^r for r a real number and be able to differentiate function which involve raising to a real exponent.
5. Understand and use logarithmic differentiation.
6. Understand and use the inverse trigonometric functions and the inverse hyperbolic functions.
7. Understand and use L'Hôpital's rule.
8. Understand and use compound interest.
9. Understand and use the chain rule to solve related rates problems.
10. Use the derivative to set up and solve optimization problems.
11. Use the Newton Raphson method to find solutions to nonlinear equations.

5.1 The Chain Rule And Derivatives Of Inverse Functions

5.1.1 The Chain Rule

The chain rule is one of the most important of differentiation rules. Special cases of it are in Theorem 3.2.6. Now it is time to consider the theorem in full generality.

Theorem 5.1.1 *Suppose $f : (a, b) \rightarrow (c, d)$ and $g : (c, d) \rightarrow \mathbb{R}$. Also suppose that $f'(x)$ exists and that $g'(f(x))$ exists. Then $(g \circ f)'(x)$ exists and*

$$(g \circ f)'(x) = g'(f(x)) f'(x).$$

Proof: Define

$$H(h) \equiv \begin{cases} \frac{g(f(x+h)) - g(f(x))}{f(x+h) - f(x)} & \text{if } f(x+h) - f(x) \neq 0 \\ g'(f(x)) & \text{if } f(x+h) - f(x) = 0 \end{cases}.$$

Then for $h \neq 0$,

$$\frac{g(f(x+h)) - g(f(x))}{h} = H(h) \frac{f(x+h) - f(x)}{h}.$$

Note that $\lim_{h \rightarrow 0} H(h) = g'(f(x))$ due to Theorems 2.9.6 on Page 61 and 3.2.2 on Page 87. Therefore, taking the limit and using Theorem 2.9.4,

$$\lim_{h \rightarrow 0} \frac{g(f(x+h)) - g(f(x))}{h} = g'(f(x)) f'(x).$$

This proves the chain rule.

Example 5.1.2 Let $f(x) = \ln |\ln(x^4 + 1)|$. Find $f'(x)$.

From the chain rule,

$$\begin{aligned} f'(x) &= \ln'(\ln(x^4 + 1)) (\ln(x^4 + 1))' = \frac{1}{\ln(x^4 + 1)} \frac{1}{x^4 + 1} (x^4)' \\ &= \left(\frac{1}{\ln(x^4 + 1)} \right) \left(\frac{1}{x^4 + 1} \right) (4x^3). \end{aligned}$$

Example 5.1.3 Let $f(x) = (2 + \ln|x|)^3$. Find $f'(x)$.

Use the chain rule again. Thus

$$f'(x) = 3(2 + \ln|x|)^2 (2 + \ln|x|)' = \frac{3}{x} (2 + \ln|x|)^2.$$

5.1.2 Implicit Differentiation

Sometimes a function is not given explicitly in terms of a formula. For example, you might have $x^2 + y^2 = 4$. This relation defines y as a function of x near a given point such as $(0, 1)$. Near this point, $y = \sqrt{4 - x^2}$. Near the point, $(0, -1)$, you have $y = -\sqrt{4 - x^2}$. Near the point, $(1, 0)$, you can't solve for y in terms of x but you can solve for x in terms of y . Thus near $(1, 0)$, $x = \sqrt{4 - y^2}$. This was a simple example but in general, you can't use algebra to solve for one of the variables in terms of the others even if the relation defines that variable as a function of the others. Here is an example in which, even though it is impossible to find $y(x)$ you can still find the derivative of y . The procedure by which this is accomplished is nothing more than the chain rule and other rules of differentiation.

Example 5.1.4 Suppose y is a differentiable function of x and $y^3 + 2yx = x^3 + 7 + \ln|y|$. Find $y'(x)$.

This illustrates the technique of implicit differentiation. If you believe y is some differentiable function of x , then you can differentiate both sides with respect to x and write, using the chain rule and product rule.

$$3y^2 y' + 2xy' + 2y = 3x^2 + \frac{y'}{y}.$$

Now you can solve for y' and obtain $y' = -\frac{2y-3x^2}{3y^3+2xy-1}y$.

Of course there are significant mathematical considerations which are being ignored when it is assumed y is a differentiable function of x . It turns out that for problems like this, the equation relating x and y actually does define y as a differentiable function of x near points where it makes sense to formally solve for y' as just done. The theorems which give this justification are called the implicit and inverse function theorems. They are some of the most profound theorems in mathematics and are topics for advanced calculus. The interested reader should consult the book by Rudin, [23] for this and generalizations of all the hard theorems given in this book. I have also given a simpler proof of the Implicit function theorem in an appendix of this book. One case is of special interest in which $y = f(x)$ and it is desired to find $\frac{dx}{dy}$ or in other words, the derivative of the inverse function.

It happens that if f is a differentiable one to one function defined on an interval, $[a, b]$, and $f'(x)$ exists and is non zero then the inverse function, f^{-1} has a derivative at the point $f(x)$. Recall that f^{-1} is defined according to the formula

$$f^{-1}(f(x)) = x.$$

Definition 5.1.5 Let $f : [a, b] \rightarrow \mathbb{R}$ be a continuous function. Define

$$f'(a) \equiv \lim_{x \rightarrow a+} \frac{f(x) - f(a)}{x - a}, \quad f'(b) \equiv \lim_{x \rightarrow b-} \frac{f(x) - f(b)}{x - b}.$$

Recall the notation $x \rightarrow a+$ means that only $x > a$ are considered in the definition of limit. The notation $x \rightarrow b-$ is defined similarly. Thus, this definition includes the derivative of f at the endpoints of the interval and to save notation,

$$f'(x_1) \equiv \lim_{x \rightarrow x_1} \frac{f(x) - f(x_1)}{x - x_1}$$

where it is understood that x is always in $[a, b]$.

Theorem 5.1.6 Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous and one to one. Suppose $f'(x_1)$ exists for some $x_1 \in [a, b]$ and $f'(x_1) \neq 0$. Then $(f^{-1})'(f(x_1))$ exists and is given by the formula, $(f^{-1})'(f(x_1)) = \frac{1}{f'(x_1)}$.

Proof: By Lemma 2.7.4, and Corollary 2.7.6 on Page 57 f is either strictly increasing or strictly decreasing and f^{-1} is continuous. Therefore there exists $\eta > 0$ such that if $0 < |f(x_1) - f(x)| < \eta$, then

$$0 < |x_1 - x| = |f^{-1}(f(x_1)) - f^{-1}(f(x))| < \delta$$

where δ is small enough that for $0 < |x_1 - x| < \delta$,

$$\left| \frac{x - x_1}{f(x) - f(x_1)} - \frac{1}{f'(x_1)} \right| < \varepsilon.$$

It follows that if $0 < |f(x_1) - f(x)| < \eta$,

$$\left| \frac{f^{-1}(f(x)) - f^{-1}(f(x_1))}{f(x) - f(x_1)} - \frac{1}{f'(x_1)} \right| = \left| \frac{x - x_1}{f(x) - f(x_1)} - \frac{1}{f'(x_1)} \right| < \varepsilon$$

Therefore, since $\varepsilon > 0$ is arbitrary,

$$\lim_{y \rightarrow f(x_1)} \frac{f^{-1}(y) - f^{-1}(f(x_1))}{y - f(x_1)} = \frac{1}{f'(x_1)}$$

and this proves the theorem.

The following obvious corollary comes from the above by not bothering with end points.

Corollary 5.1.7 *Let $f : (a, b) \rightarrow \mathbb{R}$ be continuous and one to one. Suppose $f'(x_1)$ exists for some $x_1 \in (a, b)$ and $f'(x_1) \neq 0$. Then $(f^{-1})'(f(x_1))$ exists and is given by the formula, $(f^{-1})'(f(x_1)) = \frac{1}{f'(x_1)}$.*

This is one of those theorems which is very easy to remember if you neglect the difficult questions and simply focus on formal manipulations. Consider the following.

$$f^{-1}(f(x)) = x.$$

Now use the chain rule on both sides to write

$$(f^{-1})'(f(x)) f'(x) = 1,$$

and then divide both sides by $f'(x)$ to obtain

$$(f^{-1})'(f(x)) = \frac{1}{f'(x)}.$$

Of course this gives the conclusion of the above theorem rather effortlessly and it is formal manipulations like this which aid in remembering formulas such as the one given in the theorem.

Example 5.1.8 *Let $f(x) = \ln(1+x^2) + x^3 + 7$. Show that f has an inverse and find $(f^{-1})'(7)$.*

I am not able to find a formula for the inverse function. This is typical in useful applications so you need to get used to this idea. The methods of algebra are insufficient to solve hard problems in analysis. You need something more. The question is to determine whether f has an inverse. To do this,

$$f'(x) = \frac{2x}{1+x^2} + 3x^2 + 7 > -1 + 3x^2 + 7 > 0$$

By Corollary 3.6.5 on Page 99, this function is strictly increasing on $\mathbb{R}$ and so it has an inverse function although I have no idea how to find an explicit formula for this inverse function. However, I can see that $f(0) = 7$ and so by the formula for the derivative of an inverse function,

$$(f^{-1})'(7) = (f^{-1})'(f(0)) = \frac{1}{f'(0)} = \frac{1}{7}.$$

Example 5.1.9 *Suppose $f(a) = 0$ and $f'(x) = \sqrt{1+x^4 + \ln(1+x^2)}$. Find $(f^{-1})'(0)$.*

The function, f is one to one because it is strictly increasing due to the fact that its derivative is positive for all x . As in the last example, I have no idea how to find a formula for f^{-1} but I do see that $f(a) = 0$ and so

$$(f^{-1})'(0) = (f^{-1})'(f(a)) = \frac{1}{f'(a)} = \frac{1}{\sqrt{1+a^4 + \ln(1+a^2)}}.$$

The chain rule has a particularly attractive form in Leibniz's notation. Suppose $y = g(u)$ and $u = f(x)$. Thus $y = g \circ f(x)$. Then from the above theorem

$$(g \circ f)'(x) = g'(f(x)) f'(x) = g'(u) f'(x)$$

or in other words,

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}.$$

Notice how the du 's cancel. This particular form is a very useful crutch and is used extensively in applications.

5.2 Exercises

1. In each of the following, find $\frac{dy}{dx}$.

- (a) $y = e^{\sin x}$
- (b) $y = \sqrt{7 + x^2} + \sin x$
- (c) $y = \ln(x^2 + 1)$
- (d) $y = \sin(\ln(x^2 + 1))$
- (e) $y = \ln(\sin(x) + 3)$
- (f) $y = \ln(\tan(x))$
- (g) $y = \sin^2(\ln(\tan(x)))$
- (h) $y = \sin((x + \tan x)^6)$
- (i) $y = \ln(\sin(x^2 + 7))$
- (j) $y = \tan(\cos(x^2))$
- (k) $y = \log_2(\sin(x) + 6)$
- (l) $y = \sin(\log_3(x^2 + 1))$
- (m) $y = \frac{\sqrt{x^3+7}}{\sqrt{\sin(x)+4}}$
- (n) $y = 3^{\tan(\sin(x))}$
- (o) $y = \left(\frac{x^2+2x}{\tan(x^2+1)}\right)^6$

2. In each of the following, assume the relation defines y as a function of x for values of x and y of interest and use the process of implicit differentiation to find $y'(x)$.

- (a) $xy^2 + \sin(y) = x^3 + 1$
- (b) $y^3 + x \cos(y^2) = x^4$
- (c) $y \cos(x) = \tan(y) \cos(x^2) + 2$
- (d) $(x^2 + y^2)^6 = x^3y + 3$
- (e) $\frac{xy^2+y}{y^5+x} + \cos(y) = 7$
- (f) $\sqrt{x^2 + y^4} \sin(y) = 3x$
- (g) $y^3 \sin(x) + y^2x^2 = 2x^2y + \ln|y|$

$$\begin{aligned} \text{(h)} \quad y^2 \sin(y)x + \log_3(xy) \\ = y^2 + 11 \end{aligned}$$

$$\text{(i)} \quad \sin(\tan(xy^2)) + y^3 = 16$$

$$\text{(j)} \quad \cos(\sec(\tan(y))) = 3x$$

3. In each of the following, assume the relation defines y as a function of x for values of x and y of interest and use the process of implicit differentiation to show y satisfies the given differential equation.

$$\text{(a)} \quad x^2y + \sin y = 7, \quad (x^2 + \cos y)y' + 2xy = 0.$$

$$\text{(b)} \quad x^2y^3 + \sin(y^2) = 5, \quad 2xy^3 + (3x^2y^2 + 2(\cos(y^2))y)y' = 0.$$

$$\begin{aligned} \text{(c)} \quad y^2 \sin(y) + xy = 6, \\ (2y(\sin(y)) + y^2(\cos(y)) + x) \cdot y' + y = 0. \end{aligned}$$

$$\begin{aligned} \text{(d)} \quad \tan(x^2 + y) + x^x = 7, \\ (1 + \tan^2(x^2 + y))y' + (1 + \tan^2(x^2 + y))2x + x^x(\ln x + 1) = 0. \end{aligned}$$

4. Show that if $D(g) \subseteq U \subseteq D(f)$, and if f and g are both one to one, then $f \circ g$ is also one to one.

5. Using Problem 4 show that the following functions are one to one and find the derivative of the inverse function at the indicated point.

$$\text{(a)} \quad y = e^{x^3+1}, e^2$$

$$\text{(b)} \quad y = (x^3 + x + 2), 0$$

$$\text{(c)} \quad y = \tan\left(x + \frac{\pi}{4}\right), 1$$

$$\text{(d)} \quad y = \tan\left(-x + \frac{\pi}{4}\right), 1$$

5.3 The Function x^r For r A Real Number

Theorem 4.3.2 on Page 110 says that for $x > 0$, and for r a real number,

$$\ln(x^r) = r \ln(x) \tag{5.1}$$

By this theorem, it also follows that $\ln^{-1} : \mathbb{R} \rightarrow (0, \infty)$ exists. Then by Corollary 5.1.7 $\ln^{-1}$ is differentiable. Also, for all $x \in \mathbb{R}$,

$$\ln(\ln^{-1}(x)) = x, \quad \ln(\exp_e(x)) = \ln(e^x) = x \ln e = x.$$

Since $\ln$ is one to one, $\ln^{-1}(x) = \exp_e(x) = \exp(x)$. Thus

$$\exp(\ln x) = \ln^{-1}(\ln(x)) = x, \quad \ln(\exp(y)) = y. \quad (5.2)$$

From 4.13 on Page 112, $\exp'(x) = \exp(x)$. Recall why this was. From the definition of $\ln$ and the fact $\ln e = 1$,

$$\exp'(x) = \exp'_e(x) \equiv \ln(e) \exp_e(x) = \exp(x).$$

With this understanding, it becomes possible to find derivatives of functions raised to arbitrary real powers. First, note that upon taking $\exp$ of both sides of 5.1 and using 5.2,

$$x^r = \exp(r \ln x). \quad (5.3)$$

Theorem 5.3.1 For $x > 0$, $(x^r)' = rx^{r-1}$.

Proof: Differentiate both sides of 5.3 using the chain rule. From the Wild Assumption on Page 110, in particular, the part about the validity of the laws of exponents,

$$(x^r)' = \exp'(r \ln x) \frac{r}{x} = \exp(r \ln x) \frac{r}{x} = \frac{r}{x} x^r = rx^{r-1}. \quad (5.4)$$

and this shows from 5.4 that $(x^r)' = rx^{r-1}$ as claimed.

Example 5.3.2 Suppose $f(x)$ is a non zero differentiable function. Find the derivative of $|f(x)|^r$.

From 5.3,

$$|f(x)|^r = \exp(r \ln |f(x)|).$$

Therefore,

$$\begin{aligned} (|f(x)|^r)' &= \exp(r \ln |f(x)|) (r \ln |f(x)|)' \\ &= |f(x)|^r r \frac{f'(x)}{f(x)} = r |f(x)|^{r-2} f(x) f'(x). \end{aligned}$$

5.3.1 Logarithmic Differentiation

Example 5.3.3 Let $f(x) = (1 + x^2)^x$. Find $f'(x)$.

One way to do this is to take $\ln$ of both sides and use the chain rule to differentiate both sides with respect to x . Thus

$$\ln(f(x)) = x \ln(1 + x^2)$$

and so, taking the derivative of both sides, using the chain and product rules,

$$\frac{f'(x)}{f(x)} = \frac{2x^2}{1 + x^2} + \ln(1 + x^2).$$

Then solve for $f'(x)$ to obtain

$$f'(x) = (1 + x^2)^x \left(\frac{2x^2}{1 + x^2} + \ln(1 + x^2) \right).$$

This process is called logarithmic differentiation.

Example 5.3.4 Let $f(x) = \frac{\sqrt[3]{x^3 + \sin(x)}}{\sqrt[6]{x^4 + 2x}}$. Find $f'(x)$.

You could use the quotient and chain rules but it is easier to use logarithmic differentiation.

$$\ln(f(x)) = \frac{1}{3} \ln(x^3 + \sin(x)) - \frac{1}{6} \ln(x^4 + 2x).$$

Differentiating both sides,

$$\frac{f'(x)}{f(x)} = \frac{1}{3} \left(\frac{3x^2 + \cos x}{x^3 + \sin x} \right) - \frac{1}{6} \frac{4x^3 + 2}{x^4 + 2x}.$$

Therefore, the answer is

$$f'(x) = \frac{\sqrt[3]{x^3 + \sin(x)}}{\sqrt[6]{x^4 + 2x}} \left(\frac{1}{3} \left(\frac{3x^2 + \cos x}{x^3 + \sin x} \right) - \frac{1}{6} \frac{4x^3 + 2}{x^4 + 2x} \right).$$

I think you can see the advantage of doing it this way over using the quotient rule.

5.4 Exercises

- Let $f(x) \equiv x^3 + 1$. Find $f^{-1}(y)$. Now find $(f^{-1})'(2)$.
- Let $f(x) \equiv x^3 + 7x + 3$. Explain why f has an inverse. Find $(f^{-1})'(3)$.
- Derive the quotient rule from the product rule and the chain rule. This shows you don't need to remember the wretched quotient rule if you don't want to. It follows from two rules which you cannot survive without.
- What is wrong with the following "proof" of the chain rule? Here $g'(f(x))$ exists and $f'(x)$ exists.

$$\begin{aligned} & \lim_{h \rightarrow 0} \frac{g(f(x+h)) - g(f(x))}{h} \\ &= \lim_{h \rightarrow 0} \frac{g(f(x+h)) - g(f(x))}{f(x+h) - f(x)} \frac{f(x+h) - f(x)}{h} \\ &= g'(f(x)) f'(x). \end{aligned}$$

- Is the derivative of a function always continuous? **Hint:** Consider a differentiable function f which is periodic of period 1 and non constant. (Periodic of period 1 means $f(x+1) = f(x)$ for all $x \in \mathbb{R}$.) Now consider

$$h(x) = \begin{cases} x^2 f\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

Show $h'(0) = 0$. What is $h'(x)$ for $x \neq 0$? Is h' also periodic of period 1?

- Let $f(x) = x^3 + 1$. Find an explicit formula for f^{-1} and use it to compute $(f^{-1})'(9)$. Then use the formula given in the theorem of this section to see you get the same answer.
- Find the derivatives of the following functions.

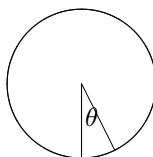
- $\sin(x^2) \ln(x^2 + 1)$
- $\ln(1 + x^2)$

- (c) $(x^3 + 1)^6 \sin(x^2 + 7)$
- (d) $\ln \left((x^3 + 1)^6 \sin(x^2 + 7) \right)$
- (e) $\tan(\sec(\sin(x^2 + 1)))$
- (f) $(\sin^2(x^2 + 5))^{\sqrt{7}}$

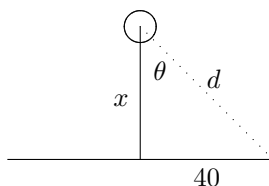
8. Use 5.3 or logarithmic differentiation to differentiate the following functions.

- (a) $(2 + \sin(x^2 + 6))^{\tan x}$
- (b) x^x
- (c) $(x^x)^x$
- (d) $(\tan^2(x^4 + 4) + 1)^{\cos x}$
- (e) $(\sin^2(x))^{\tan x}$

9. A search light at a prison revolves five times every minute and is located 200 yards from a long straight wall. How fast in yards per minute is the light moving along the wall at a distance of 100 yards from the point of the wall closest to the light? If you are an escaping convict, where is the most dangerous location on this wall?
10. A circular disk of paper has a circular sector removed and then the edges are joined to form a cone. What is the angle of the removed circular sector which will create the cone of largest volume?



11. A light pole with a light on top of it is intended to illuminate the edge of a circle of radius 40 feet centered at the base of the light pole. The brightness of the illumination is of the form $k \frac{\cos \theta}{d^2}$ where θ and d are given in the following picture. Find the height, x which will result in the brightest illumination at the edge of this circle.



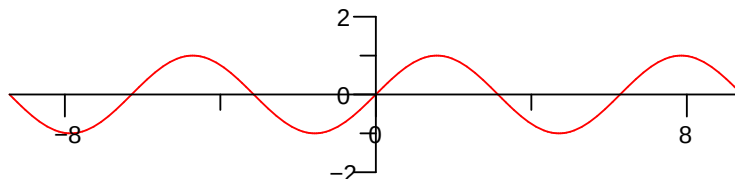
12. * Let θ be the angle between the two equal sides of an isosceles triangle. Suppose the sum of the lengths of the sides of this triangle is L . Find the value of θ which will maximize the area of the triangle.
13. A window has total perimeter equal to L . It consists of a square surmounted by an equilateral triangle. (All sides equal) Find the dimensions which will make the largest window.
14. * A car is proceeding down a road near Crystal Falls Michigan at 60 miles per hour. Fifty yards in front of the car there is a suicidal deer standing 50 feet from the road¹. Consider the angle formed by the road and the line of sight to the

¹Crystal Falls Michigan is in the upper peninsula. It is a very beautiful place, especially in the fall when the leaves change colors but it is hazardous to drive there, especially at dusk, because of suicidal deer which jump out in front of cars unexpectedly. These are very large deer.

deer. How fast is this angle changing? What happens to the rate of change of this angle as the driver gets closer to the deer? This is actually a related rates problem which will be discussed more later.

5.5 The Inverse Trigonometric Functions

It is desired to consider the inverse trigonometric functions. Graphing the function $y = \sin x$, is clear $\sin$ is not one to one on $\mathbb{R}$ and so it is not possible to define an inverse function.



However, a little thing like this will not prevent the definition of useful inverse trig. functions. Observe the function, $\sin$, is one to one on the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ shown in the above picture as the interval containing zero on which the function climbs from -1 to 1 . Also note that for y on this interval, $\cos(y) \geq 0$. Now the $\arcsin$ function is defined as the inverse of the $\sin$ when its domain is restricted to $[-\frac{\pi}{2}, \frac{\pi}{2}]$. In words, $\arcsin(x)$ is defined to be the angle whose sine is x which lies in $[-\frac{\pi}{2}, \frac{\pi}{2}]$. From Theorem 5.1.6 on Page 119 about the derivative of the inverse function, the derivative of $x \rightarrow \arcsin(x)$ exists for all $x \in [-1, 1]$. The formula in this theorem could be used to find the derivative of $\arcsin$ but it is more useful to simply use that theorem to resolve the existence question and apply the chain rule to find the formula. It is a mistake to memorize too many formulas. Let $y = \arcsin(x)$ so $\sin(y) = x$. Now taking the derivative of both sides,

$$\cos(y) y' = 1$$

and so

$$y' = \frac{1}{\cos y} = \frac{1}{\sqrt{1 - \sin^2(y)}} = \frac{1}{\sqrt{1 - x^2}}.$$

The positive value of the square root is used because for $y \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\cos(y) \geq 0$. Thus

$$\frac{1}{\sqrt{1 - x^2}} = \arcsin'(x). \quad (5.5)$$

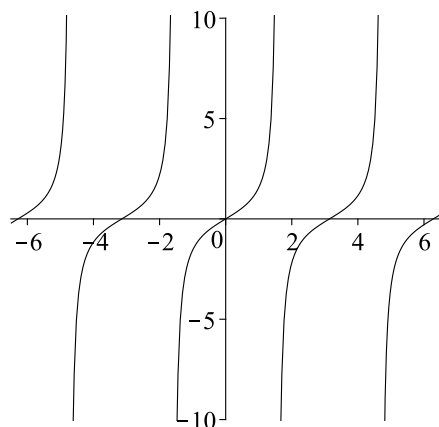
Next consider the inverse tangent function. You are aware that $\tan$ is periodic of period π because

$$\begin{aligned} \tan(x + \pi) &= \frac{\sin(x + \pi)}{\cos(x + \pi)} = \frac{\sin(x)\cos(\pi) + \cos(x)\sin(\pi)}{\cos(x)\cos(\pi) - \sin(x)\sin(\pi)} \\ &= \frac{-\sin(x)}{-\cos(x)} = \tan(x). \end{aligned}$$

Therefore, it is impossible to take the inverse of $\tan$. However, $\tan$ is one to one on $(-\frac{\pi}{2}, \frac{\pi}{2})$ and

$$\lim_{x \rightarrow \frac{\pi}{2}^-} \tan(x) = +\infty, \quad \lim_{x \rightarrow -\frac{\pi}{2}^+} \tan(x) = -\infty$$

as shown in the following graph of $y = \tan(x)$.



Therefore, $\arctan(x)$ for $x \in (-\infty, \infty)$ is defined according to the rule: $\arctan(x)$ is the angle whose tangent is x which is in $(-\frac{\pi}{2}, \frac{\pi}{2})$. By Theorem 5.1.6 $\arctan$ has a derivative. Therefore, letting $y = \arctan(x)$, $\tan(y) = x$ and by the chain rule,

$$\sec^2(y) y' = 1.$$

Therefore,

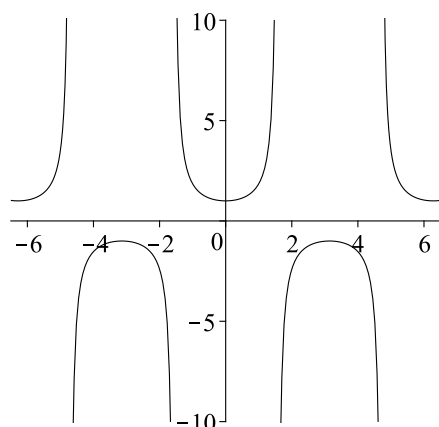
$$y' = \frac{1}{\sec^2(y)} = \frac{1}{1 + \tan^2(y)} = \frac{1}{1 + x^2}.$$

and so

$$\frac{1}{1 + x^2} = \arctan'(x). \quad (5.6)$$

The inverse secant function can be defined similarly. There is no agreement on the best way to restrict the domain of $\sec$. I will follow the way of doing it which is used in the book by Salas and Hille [25] recognizing that there are good reasons for doing it other ways also. The graph of $\sec$ is represented below. There is a vertical asymptote at $x = \frac{\pi}{2}$. Thus

$$\lim_{x \rightarrow \frac{\pi}{2}^-} \sec(x) = +\infty, \quad \lim_{x \rightarrow \frac{\pi}{2}^+} \sec(x) = -\infty$$



As in the case of $\arcsin$ and $\arctan$, $\operatorname{arcsec}(x)$ is the angle whose secant is x which lies in $[0, \frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi]$. Let $y = \operatorname{arcsec}(x)$ so $x = \sec(y)$ and using the chain rule,

$$1 = \sec(y) \tan(y) y'.$$

Now from the trig. identity, $1 + \tan^2(y) = \sec^2(y)$,

$$\begin{aligned} y' &= \frac{1}{\sec(y) \tan(y)} \\ &= \frac{1}{x(\pm\sqrt{x^2-1})} \end{aligned}$$

and it is necessary to consider what to do with $\pm$. If $y \in [0, \frac{\pi}{2})$, both $x = \sec(y)$ and $\tan(y)$ are nonnegative and so in this case,

$$y' = \frac{1}{x\sqrt{x^2-1}} = \frac{1}{|x|\sqrt{x^2-1}}.$$

If $y \in (\frac{\pi}{2}, \pi]$, then $x = \sec(y) < 0$ and $\tan(y) \leq 0$ so

$$y' = \frac{1}{x(-1)\sqrt{x^2-1}} = \frac{1}{(-x)\sqrt{x^2-1}} = \frac{1}{|x|\sqrt{x^2-1}}.$$

Thus either way,

$$y' = \frac{1}{|x|\sqrt{x^2-1}}.$$

This yields the formula

$$\frac{1}{|x|\sqrt{x^2-1}} = \operatorname{arcsec}'(x). \quad (5.7)$$

As in the case of $\ln$, there is an interesting and useful formula involving $\operatorname{arcsec}(|x|)$. For $x < 0$, this function equals $\operatorname{arcsec}(-x)$ and so by the chain rule, its derivative equals

$$\frac{1}{|-x|\sqrt{x^2-1}}(-1) = \frac{1}{x\sqrt{x^2-1}}$$

If $x > 0$, this function equals $\operatorname{arcsec}(x)$ and its derivative equals

$$\frac{1}{|x|\sqrt{x^2-1}} = \frac{1}{x\sqrt{x^2-1}}$$

and so either way,

$$(\operatorname{arcsec}(|x|))' = \frac{1}{x\sqrt{x^2-1}}. \quad (5.8)$$

5.6 The Hyperbolic And Inverse Hyperbolic Functions

The hyperbolic functions are given by

$$\sinh(x) \equiv \frac{e^x - e^{-x}}{2}, \quad \cosh(x) \equiv \frac{e^x + e^{-x}}{2}, \quad \tanh(x) \equiv \frac{\sinh(x)}{\cosh(x)}.$$

The first of these is called the hyperbolic sine and the second the hyperbolic cosine. I imagine you can guess what the third is called. If you guessed “hyperbolic tangent” you got it right. The other hyperbolic functions are defined by analogy to the circular functions.

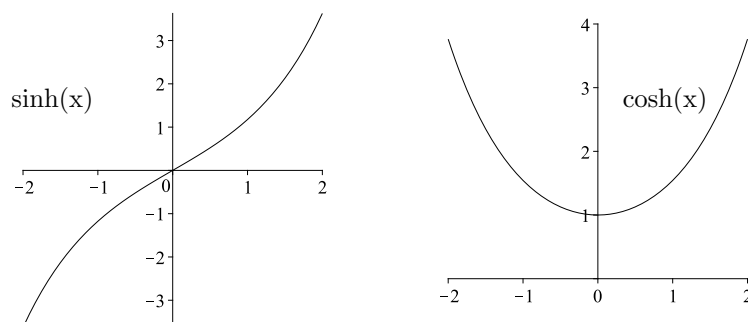
The reason these are called hyperbolic functions is that

$$\cosh^2 t - \sinh^2 t = 1 \quad (5.9)$$

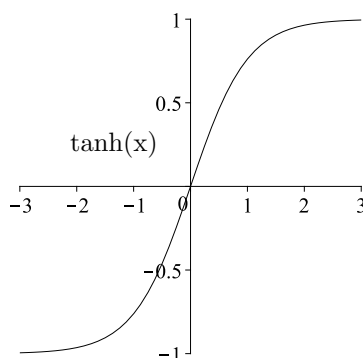
and so the point, $(\cosh t, \sinh t)$ is a point on the hyperbola whose equation is $x^2 - y^2 = 1$. This is not important but is the source for the term hyperbolic. Using the chain rule,

$$\cosh'(x) = \sinh(x), \quad \sinh'(x) = \cosh x.$$

Also, you see that $\sinh(0) = 0$, $\cosh(0) = 1$ and that $\sinh(x) < 0$ if $x < 0$ while $\sinh(x) > 0$ for $x > 0$, but $\cosh(x) > 0$ for all x . Therefore, $\sinh$ is an increasing function, concave down for $x < 0$ and concave up for $x > 0$ because $\sinh''(x) = \sinh(x)$ while $\cosh$ is decreasing for $x < 0$ and increasing for $x > 0$. Since $\cosh''(x) = \cosh(x)$, it is concave up for all x . Thus the graphs of these functions are as follows.



Also, a short computation shows $\tanh(x) = (1 - e^{-2x}) / (1 + e^{-2x})$. Its graph is below.



Since $x \rightarrow \sinh(x)$ is strictly increasing, it has an inverse function, $\sinh^{-1}(x)$. If $y = \sinh^{-1}(x)$, then $\sinh(y) = x$ and so using the chain rule and the theorem about the existence of the derivative of the inverse function, $y' \cosh(y) = 1$. From the identity 5.9 $\cosh(y) = \sqrt{1 + \sinh^2(y)} = \sqrt{1 + x^2}$. Therefore,

$$y' = \frac{1}{\sqrt{1 + x^2}}$$

which gives the formula

$$\frac{1}{\sqrt{1 + x^2}} = (\sinh^{-1})'(x). \quad (5.10)$$

The derivative of the hyperbolic tangent is also easy to find. This yields after a short computation

$$(\tanh x)' = 1 - \tanh^2 x. \quad (5.11)$$

Another notation for the inverse hyperbolic functions which is sometimes used is arsinh or arcosh or artanh by analogy with the inverse trig. functions.

5.7 Exercises

- Verify 5.11.
- Simplify the following.
 - $\sin(\arctan(1))$
 - $\cos(\arctan \sqrt{3})$
 - $\tan(\arcsin(\sqrt{3}/2))$
 - $\sec(\arcsin(\sqrt{2}/2))$
 - $\tan(\arcsin(1/2))$
 - $\cos(\arctan(\sqrt{3}) - \arcsin(\sqrt{2}/2))$
 - $\sin(\arctan(x))$
 - $\cos(\arcsin(x))$
 - $\cos(2 \arcsin(x))$
 - $\sec(2 \arctan(x))$
 - $\tan(2 \arcsin(x) + \arctan(x^2))$
 - $\sec(\arcsin(x) + 2 \arccos(x))$
- Find the derivatives and give the domains of the following functions.
 - $\arcsin(x^2 + 3x)$
 - $\arctan(3x + 5)$
 - $\operatorname{arcsec}(2x + 1)$
 - $\arctan(\sinh(x))$
 - $\sinh(\tan(3x))$
 - $\cosh(\csc(3x))$
 - $\tanh(\sec(3x))$
- For a and b positive constants, find

$$\frac{d}{dx} \left(\frac{\arctan\left(\frac{bx}{a}\right)}{ab} \right).$$
- Find the derivative of

$$\left(\arctan \frac{2}{\sqrt{(9x^2 - 4)}} + \operatorname{arcsec} \left| \frac{3x}{2} \right| \right).$$
- Use the process of implicit differentiation to find y' in the following examples.
 - $\arcsin(x^2y) + \arctan(xy^2) = \sinh(y) + 3$
 - $x^2 \tanh(yx) + \arctan(x^2) = 5$
 - $y^4 \sinh(x) + \arccos(x^2 + y^2) = 1$
 - $\sinh(x \tanh(x^2 + y^2)) = 2$
- Find the derivatives of the following functions.
 - $\sinh^{-1}(x^2 + 7)$
 - $\tanh^{-1}(x)$
 - $\sin(\sinh^{-1}(x^2 + 2))$
 - $\sin(\tanh(x))$
 - $x^2 \sinh(\sin(\cos(x)))$
 - $(\cosh(x^3))^{\sqrt{6}}$
 - $(1 + x^4)^{\sin x}$
- Simplify for $x > 0$, $\arcsin x + \arccos x$.
- A wonderful identity which was used to compute π for over 200 years² is the following.

$$\frac{\pi}{4} = 4 \arctan\left(\frac{1}{5}\right) - \arctan\left(\frac{1}{239}\right).$$

Establish this identity by taking the tangent of both sides and using an appropriate formula for the tangent of the difference of two angles. Use De Moivre's theorem to get some help in finding a formula for $\tan(4\theta)$.
- Find a formula for $\tanh^{-1}$ in terms of $\ln$.
- Find a formula for $\sinh^{-1}$ in terms of $\ln$.
- Prove $1 - \tanh^2 x = \operatorname{sech}^2 x$.
- Prove $\coth^2 x - 1 = \operatorname{csch}^2 x$.
- What about $\cosh^{-1}$? Define it by restricting the domain of $\cosh$ to be nonnegative numbers? What is $\cosh^{-1}$ in terms of $\ln$?

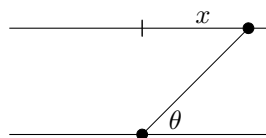
²John Machin computed π to 100 decimal places in 1706 through the use of this identity. Later in 1873 William Shanks did it to over 700 places using this identity. The next advance was in 1948, 808 decimal places. After this, computers began to be used and currently π is "known" to millions of decimal places. Many other schemes have been used besides this identity for computing π .

15. * Show $\arcsin x = \arctan \left(\frac{x}{\sqrt{1-x^2}} \right)$.

It is possible to start with the arctan function and obtain all the other trig functions in terms of this one. If you knew the function, arctan explain how to define sin and cos. This is interesting because there is a simple way to define arctan directly as a function of a real variable[15]. Approaches like these avoid all reference to plane geometry.

16. A divided highway is separated by a median which is $1/10$ of a mile wide. Two cars pass each other, one going west and the other east. The west

bound car is traveling at 70 miles per hour while the east bound car is traveling at 60 miles per hour. The situation is described in the following picture.



How fast is the distance between the two cars changing when x , shown in the picture, equals $1/10$ mile? How fast is the angle, θ , shown in the picture, changing at this instant?

5.8 L'Hôpital's Rule

There is an interesting rule which is often useful for evaluating difficult limits called L'Hôpital's³ rule. The best versions of this rule are based on the Cauchy Mean value theorem, Theorem 3.6.2 on Page 98.

Theorem 5.8.1 Let $[a, b] \subseteq [-\infty, \infty]$ and suppose f, g are functions which satisfy,

$$\lim_{x \rightarrow b-} f(x) = \lim_{x \rightarrow b-} g(x) = 0, \quad (5.12)$$

and f' and g' exist on (a, b) with $g'(x) \neq 0$ on (a, b) . Suppose also that

$$\lim_{x \rightarrow b-} \frac{f'(x)}{g'(x)} = L. \quad (5.13)$$

Then

$$\lim_{x \rightarrow b-} \frac{f(x)}{g(x)} = L. \quad (5.14)$$

Proof: By the definition of limit and 5.13 there exists $c < b$ such that if $t > c$, then

$$\left| \frac{f'(t)}{g'(t)} - L \right| < \frac{\varepsilon}{2}.$$

Now pick x, y such that $c < x < y < b$. By the Cauchy mean value theorem, there exists $t \in (x, y)$ such that

$$g'(t)(f(x) - f(y)) = f'(t)(g(x) - g(y)).$$

Since $g'(s) \neq 0$ for all $s \in (a, b)$ it follows $g(x) - g(y) \neq 0$. Therefore,

$$\frac{f'(t)}{g'(t)} = \frac{f(x) - f(y)}{g(x) - g(y)}$$

³L'Hôpital published the first calculus book in 1696. This rule, named after him, appeared in this book. The rule was actually due to Bernoulli who had been L'Hôpital's teacher. L'Hôpital did not claim the rule as his own but Bernoulli accused him of plagiarism. Nevertheless, this rule has become known as L'Hôpital's rule ever since. The version of the rule presented here is superior to what was discovered by Bernoulli and depends on the Cauchy mean value theorem which was found over 100 years after the time of L'Hôpital.

and so, since $t > c$,

$$\left| \frac{f(x) - f(y)}{g(x) - g(y)} - L \right| < \frac{\varepsilon}{2}.$$

Now letting $y \rightarrow b-$,

$$\left| \frac{f(x)}{g(x)} - L \right| \leq \frac{\varepsilon}{2} < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this shows 5.14.

The following corollary is proved in the same way.

Corollary 5.8.2 *Let $[a, b] \subseteq [-\infty, \infty]$ and suppose f, g are functions which satisfy,*

$$\lim_{x \rightarrow a+} f(x) = \lim_{x \rightarrow a+} g(x) = 0, \quad (5.15)$$

and f' and g' exist on (a, b) with $g'(x) \neq 0$ on (a, b) . Suppose also that

$$\lim_{x \rightarrow a+} \frac{f'(x)}{g'(x)} = L. \quad (5.16)$$

Then

$$\lim_{x \rightarrow a+} \frac{f(x)}{g(x)} = L. \quad (5.17)$$

Here is a simple example which illustrates the use of this rule.

Example 5.8.3 *Find $\lim_{x \rightarrow 0} \frac{5x + \sin 3x}{\tan 7x}$.*

The conditions of L'Hôpital's rule are satisfied because the numerator and denominator both converge to 0 and the derivative of the denominator is nonzero for x close to 0. Therefore, if the limit of the quotient of the derivatives exists, it will equal the limit of the original function. Thus,

$$\lim_{x \rightarrow 0} \frac{5x + \sin 3x}{\tan 7x} = \lim_{x \rightarrow 0} \frac{5 + 3 \cos 3x}{7 \sec^2(7x)} = \frac{8}{7}.$$

Sometimes you have to use L'Hôpital's rule more than once.

Example 5.8.4 *Find $\lim_{x \rightarrow 0} \frac{\sin x - x}{x^3}$.*

Note that $\lim_{x \rightarrow 0} (\sin x - x) = 0$ and $\lim_{x \rightarrow 0} x^3 = 0$. Also, the derivative of the denominator is nonzero for x close to 0. Therefore, if $\lim_{x \rightarrow 0} \frac{\cos x - 1}{3x^2}$ exists and equals L , it will follow from L'Hôpital's rule that the original limit exists and equals L . However, $\lim_{x \rightarrow 0} (\cos x - 1) = 0$ and $\lim_{x \rightarrow 0} 3x^2 = 0$ so L'Hôpital's rule can be applied again to consider $\lim_{x \rightarrow 0} \frac{-\sin x}{6x}$. From L'Hôpital's rule, if this limit exists and equals L , it will follow that $\lim_{x \rightarrow 0} \frac{\cos x - 1}{3x^2} = L$ and consequently $\lim_{x \rightarrow 0} \frac{\sin x - x}{x^3} = L$. But from Lemma 4.1.1 on Page 105, $\lim_{x \rightarrow 0} \frac{-\sin x}{6x} = \frac{-1}{6}$. Therefore, by L'Hôpital's rule, $\lim_{x \rightarrow 0} \frac{\sin x - x}{x^3} = \frac{-1}{6}$.

Warning 5.8.5 *Be sure to check the assumptions of L'Hôpital's rule before using it.*

Example 5.8.6 *Find $\lim_{x \rightarrow 0+} \frac{\cos 2x}{x}$.*

The numerator becomes close to 1 and the denominator gets close to 0. Therefore, the assumptions of L'Hôpital's rule do not hold and so it does not apply. In fact there is no limit unless you define the limit to equal $+\infty$. Now let's try to use the conclusion of L'Hôpital's rule even though the conditions for using this rule are not verified. Take the derivative of the numerator and the denominator which yields $\frac{-2 \sin 2x}{1}$, an expression whose limit as $x \rightarrow 0+$ equals 0. This is a good illustration of the above warning.

Some people get the unfortunate idea that one can find limits by doing experiments with a calculator. If the limit is taken as x gets close to 0, these people think one can find the limit by evaluating the function at values of x which are closer and closer to 0. Theoretically, this should work although you have no way of knowing how small you need to take x to get a good estimate of the limit. In practice, the procedure may fail miserably.

Example 5.8.7 Find $\lim_{x \rightarrow 0} \frac{\ln|1+x^{10}|}{x^{10}}$.

This limit equals $\lim_{y \rightarrow 0} \frac{\ln|1+y|}{y} = \lim_{y \rightarrow 0} \frac{(\frac{1}{1+y})}{1} = 1$ where L'Hôpital's rule has been used. This is an amusing example. You should plug .001 in to the function, $\frac{\ln|1+x^{10}|}{x^{10}}$ and see what your calculator or computer gives you. If it is like mine, it will give the answer, 0 and will keep on returning the answer of 0 for smaller numbers than .001. This illustrates the folly of trying to compute limits through calculator or computer experiments. Indeed, you could say that a calculator is as useful for taking limits as a bicycle is for swimming.

There is another form of L'Hôpital's rule in which

$$\lim_{x \rightarrow b-} f(x) = \pm\infty, \quad \lim_{x \rightarrow b-} g(x) = \pm\infty.$$

Theorem 5.8.8 Let $[a, b] \subseteq [-\infty, \infty]$ and suppose f, g are functions which satisfy,

$$\lim_{x \rightarrow b-} f(x) = \pm\infty \text{ and } \lim_{x \rightarrow b-} g(x) = \pm\infty, \quad (5.18)$$

and f' and g' exist on (a, b) with $g'(x) \neq 0$ on (a, b) . Suppose also

$$\lim_{x \rightarrow b-} \frac{f'(x)}{g'(x)} = L. \quad (5.19)$$

Then

$$\lim_{x \rightarrow b-} \frac{f(x)}{g(x)} = L. \quad (5.20)$$

Proof: By the definition of limit and 5.19 there exists $c < b$ such that if $t > c$, then

$$\left| \frac{f'(t)}{g'(t)} - L \right| < \frac{\varepsilon}{2}.$$

Now pick x, y such that $c < x < y < b$. By the Cauchy mean value theorem, there exists $t \in (x, y)$ such that

$$g'(t)(f(x) - f(y)) = f'(t)(g(x) - g(y)).$$

Since $g'(s) \neq 0$ on (a, b) , it follows $g(x) - g(y) \neq 0$. Therefore,

$$\frac{f'(t)}{g'(t)} = \frac{f(x) - f(y)}{g(x) - g(y)}$$

and so, since $t > c$,

$$\left| \frac{f(x) - f(y)}{g(x) - g(y)} - L \right| < \frac{\varepsilon}{2}.$$

Now this implies

$$\left| \frac{f(y) \left(\frac{f(x)}{f(y)} - 1 \right)}{g(y) \left(\frac{g(x)}{g(y)} - 1 \right)} - L \right| < \frac{\varepsilon}{2}$$

where for all y large enough, both $\frac{f(x)}{f(y)} - 1$ and $\frac{g(x)}{g(y)} - 1$ are not equal to zero. Continuing to rewrite the above inequality yields

$$\left| \frac{f(y)}{g(y)} - L \frac{\left(\frac{g(x)}{g(y)} - 1 \right)}{\left(\frac{f(x)}{f(y)} - 1 \right)} \right| < \frac{\varepsilon}{2} \left| \frac{\left(\frac{g(x)}{g(y)} - 1 \right)}{\left(\frac{f(x)}{f(y)} - 1 \right)} \right|.$$

Therefore, for y large enough,

$$\left| \frac{f(y)}{g(y)} - L \right| \leq \left| L - L \frac{\left(\frac{g(x)}{g(y)} - 1 \right)}{\left(\frac{f(x)}{f(y)} - 1 \right)} \right| + \frac{\varepsilon}{2} \left| \frac{\left(\frac{g(x)}{g(y)} - 1 \right)}{\left(\frac{f(x)}{f(y)} - 1 \right)} \right| < \varepsilon$$

due to the assumption 5.18 which implies

$$\lim_{y \rightarrow b-} \frac{\left(\frac{g(x)}{g(y)} - 1 \right)}{\left(\frac{f(x)}{f(y)} - 1 \right)} = 1.$$

Therefore, whenever y is large enough,

$$\left| \frac{f(y)}{g(y)} - L \right| < \varepsilon$$

and this is what is meant by 5.20. This proves the theorem.

As before, there is no essential difference between the proof in the case where $x \rightarrow b-$ and the proof when $x \rightarrow a+$. This observation is stated as the next corollary.

Corollary 5.8.9 *Let $[a, b] \subseteq [-\infty, \infty]$ and suppose f, g are functions which satisfy,*

$$\lim_{x \rightarrow a+} f(x) = \pm\infty \text{ and } \lim_{x \rightarrow a+} g(x) = \pm\infty, \quad (5.21)$$

and f' and g' exist on (a, b) with $g'(x) \neq 0$ on (a, b) . Suppose also that

$$\lim_{x \rightarrow a+} \frac{f'(x)}{g'(x)} = L. \quad (5.22)$$

Then

$$\lim_{x \rightarrow a+} \frac{f(x)}{g(x)} = L. \quad (5.23)$$

Theorems 5.8.1 5.8.8 and Corollaries 5.8.2 and 5.8.9 will be referred to as L'Hôpital's rule from now on. Theorem 5.8.1 and Corollary 5.8.2 involve the notion of indeterminate forms of the form $\frac{0}{0}$. Please do not think any meaning is being assigned to the nonsense expression $\frac{0}{0}$. It is just a symbol to help remember the sort of thing described by Theorem 5.8.1 and Corollary 5.8.2. Theorem 5.8.8 and Corollary 5.8.9 deal with indeterminate forms which are of the form $\frac{\pm\infty}{\pm\infty}$. Again, this is just a symbol which is helpful in remembering the sort of thing being considered. There are other indeterminate forms which can be reduced to these forms just discussed. Don't ever try to assign meaning to such symbols.

Example 5.8.10 Find $\lim_{y \rightarrow \infty} \left(1 + \frac{x}{y}\right)^y$.

It is good to first see why this is called an indeterminate form. One might think that as $y \rightarrow \infty$, it follows $x/y \rightarrow 0$ and so $1 + \frac{x}{y} \rightarrow 1$. Now 1 raised to anything is 1 and so it would seem this limit should equal 1. On the other hand, if $x > 0$, $1 + \frac{x}{y} > 1$ and a number raised to higher and higher powers should approach ∞ . It really isn't clear what this limit should be. It is an indeterminate form which can be described as 1^∞ . By definition,

$$\left(1 + \frac{x}{y}\right)^y = \exp\left(y \ln\left(1 + \frac{x}{y}\right)\right).$$

Now using L'Hôpital's rule,

$$\begin{aligned} \lim_{y \rightarrow \infty} y \ln\left(1 + \frac{x}{y}\right) &= \lim_{y \rightarrow \infty} \frac{\ln\left(1 + \frac{x}{y}\right)}{1/y} \\ &= \lim_{y \rightarrow \infty} \frac{\frac{1}{1+(x/y)}(-x/y^2)}{(-1/y^2)} \\ &= \lim_{y \rightarrow \infty} \frac{x}{1 + (x/y)} = x \end{aligned}$$

Therefore,

$$\lim_{y \rightarrow \infty} y \ln\left(1 + \frac{x}{y}\right) = x$$

Since exp is continuous, it follows

$$\lim_{y \rightarrow \infty} \left(1 + \frac{x}{y}\right)^y = \lim_{y \rightarrow \infty} \exp\left(y \ln\left(1 + \frac{x}{y}\right)\right) = e^x.$$

5.8.1 Interest Compounded Continuously

Suppose you put money in the bank and it accrues interest at the rate of r per payment period. These terms need a little explanation. If the payment period is one month, and you started with \$100 then the amount at the end of one month would equal $100(1+r) = 100 + 100r$. In this the second term is the interest and the first is called the principal. Now you have $100(1+r)$ in the bank. This becomes the new principal. How much will you have at the end of the second month? By analogy to what was just done it would equal

$$100(1+r) + 100(1+r)r = 100(1+r)^2.$$

In general, the amount you would have at the end of n months is $100(1+r)^n$.

When a bank says they offer 6% compounded monthly, this means r , the rate per payment period equals .06/12. Consider the problem of a rate of r per year and compounding the interest n times a year and letting n increase without bound. This is what is meant by compounding continuously. The interest rate per payment period is then r/n and the number of payment periods after time t years is approximately tn . From the above the amount in the account after t years is

$$P\left(1 + \frac{r}{n}\right)^{nt} \tag{5.24}$$

Recall from Example 5.8.10 that $\lim_{y \rightarrow \infty} \left(1 + \frac{x}{y}\right)^y = e^x$. The expression in 5.24 can be written as

$$P\left[\left(1 + \frac{r}{n}\right)^n\right]^t$$

and so, taking the limit as $n \rightarrow \infty$, you get

$$Pe^{rt} = A.$$

This shows how to compound interest continuously.

Example 5.8.11 Suppose you have \$100 and you put it in a savings account which pays 6% compounded continuously. How much will you have at the end of 4 years?

From the above discussion, this would be $100e^{(.06)4} = 127.12$. Thus, in 4 years, you would gain interest of about \$27.

5.9 Exercises

1. Find the limits.

- (a) $\lim_{x \rightarrow 0} \frac{3x-4 \sin 3x}{\tan 3x}$
- (b) $\lim_{x \rightarrow \frac{\pi}{2}} (\tan x)^{x-(\pi/2)}$
- (c) $\lim_{x \rightarrow 1} \frac{\arctan(4x-4)}{\arcsin(4x-4)}$
- (d) $\lim_{x \rightarrow 0} \frac{\arctan 3x-3x}{x^3}$
- (e) $\lim_{x \rightarrow 0+} \frac{9^{\sec x-1}-1}{3^{\sec x-1}-1}$
- (f) $\lim_{x \rightarrow 0} \frac{3x+\sin 4x}{\tan 2x}$
- (g) $\lim_{x \rightarrow \pi/2} \frac{\ln(\sin x)}{x-(\pi/2)}$
- (h) $\lim_{x \rightarrow 0} \frac{\cosh 2x-1}{x^2}$
- (i) $\lim_{x \rightarrow 0} \frac{-\arctan x+x}{x^3}$
- (j) $\lim_{x \rightarrow 0} \frac{x^8 \sin \frac{1}{x}}{\sin 3x}$
- (k) $\lim_{x \rightarrow \infty} (1+5^x)^{\frac{2}{x}}$
- (l) $\lim_{x \rightarrow 0} \frac{-2x+3 \sin x}{x}$
- (m) $\lim_{x \rightarrow 1} \frac{\ln(\cos(x-1))}{(x-1)^2}$
- (n) $\lim_{x \rightarrow 0+} \sin^{\frac{1}{x}} x$
- (o) $\lim_{x \rightarrow 0} (\csc 5x - \cot 5x)$
- (p) $\lim_{x \rightarrow 0+} \frac{3^{\sin x}-1}{2^{\sin x}-1}$
- (q) $\lim_{x \rightarrow 0+} (4x)^{x^2}$
- (r) $\lim_{x \rightarrow \infty} \frac{x^{10}}{(1.01)^x}$
- (s) $\lim_{x \rightarrow 0} (\cos 4x)^{(1/x^2)}$

2. Find the following limits.

- (a) $\lim_{x \rightarrow 0+} \frac{1-\sqrt{\cos 2x}}{\sin^4(4\sqrt{x})}$
- (b) $\lim_{x \rightarrow 0} \frac{2^{x^2}-2^{5x}}{\sin\left(\frac{x^2}{5}\right)-\sin(3x)}$
- (c) $\lim_{n \rightarrow \infty} n \left(\sqrt[n]{7} - 1 \right).$

- (d) $\lim_{x \rightarrow \infty} \left(\frac{3x+2}{5x-9} \right)^{x^2}.$
- (e) $\lim_{x \rightarrow \infty} \left(\frac{3x+2}{5x-9} \right)^{1/x}.$
- (f) $\lim_{n \rightarrow \infty} \left(\cos \frac{2x}{\sqrt{n}} \right)^n.$
- (g) $\lim_{n \rightarrow \infty} \left(\cos \frac{2x}{\sqrt{5n}} \right)^n.$
- (h) $\lim_{x \rightarrow 3} \frac{x^x-27}{x-3}.$
- (i) $\lim_{n \rightarrow \infty} \cos \left(\pi \frac{\sqrt{4n^2+13n}}{n} \right).$
- (j) $\lim_{x \rightarrow \infty} \left(\sqrt[3]{x^3+7x^2} - \sqrt{x^2-11x} \right).$
- (k) $\lim_{x \rightarrow \infty} \left(\sqrt[5]{x^5+7x^4} - \sqrt[3]{x^3-11x^2} \right).$
- (l) $\lim_{x \rightarrow \infty} \left(\frac{5x^2+7}{2x^2-11} \right)^{\frac{x}{1-x}}.$
- (m) $\lim_{x \rightarrow \infty} \left(\frac{5x^2+7}{2x^2-11} \right)^{\frac{x \ln x}{1-x}}.$
- (n) $\lim_{x \rightarrow 0+} \frac{\ln(e^{2x^2}+7\sqrt{x})}{\sinh(\sqrt{x})}.$
- (o) $\lim_{x \rightarrow 0+} \frac{\sqrt[3]{x}-\sqrt[5]{x}}{\sqrt[3]{x}-\sqrt[11]{x}}.$

3. Find the following limits.

- (a) $\lim_{x \rightarrow 0+} (1+3x)^{\cot 2x}$
- (b) $\lim_{x \rightarrow 0} \frac{\sin x-x}{x^2} = 0$
- (c) $\lim_{x \rightarrow 0} \frac{\sin x-x}{x^3}$
- (d) $\lim_{x \rightarrow 0} \frac{\tan(\sin x)-\sin(\tan x)}{x^7}$
- (e) $\lim_{x \rightarrow 0} \frac{\tan(\sin 2x)-\sin(\tan 2x)}{x^7}$
- (f) $\lim_{x \rightarrow 0} \frac{\sin(x^2)-\sin^2(x)}{x^4}$
- (g) $\lim_{x \rightarrow 0} \frac{e^{-(1/x^2)}}{x}$
- (h) $\lim_{x \rightarrow 0} \left(\frac{1}{x} - \cot(x) \right)$
- (i) $\lim_{x \rightarrow 0} \frac{\cos(\sin x)-1}{x^2}$
- (j) $\lim_{x \rightarrow \infty} \left(x^2 (4x^4+7) \right)^{1/2} - 2x^4$

- (k) $\lim_{x \rightarrow 0} \frac{\cos(x) - \cos(4x)}{\tan(x^2)}$ a measure of the difference between compounding continuously and compounding daily.
- (l) $\lim_{x \rightarrow 0} \frac{\arctan(3x)}{x}$
- (m) $\lim_{x \rightarrow \infty} \left[(x^9 + 5x^6)^{1/3} - x^3 \right]$ 6. Let $a > 1$. Find
4. Suppose you want to have \$2000 saved at the end of 5 years. How much money should you place into an account which pays 7% per year compounded continuously?
5. Using a good calculator, find $e^{.06} - \left(1 + \frac{.06}{360}\right)^{360}$. Explain why this gives

$$\lim_{x \rightarrow \infty} \left(\frac{1}{x} \frac{a^x - 1}{a - 1} \right)^{1/x}.$$

This is a case of a problem which appeared on the 1956 Putnam exam. **Hint:** Consider the $\ln$ of the function and split it up. It isn't too bad if you do this.

5.10 Related Rates

Sometimes some variables are related by a formula and it is known how fast all are changing but one. The related rates problem asks for how fast the remaining variable is changing.

Example 5.10.1 A cube of ice is melting such that $\frac{dV}{dt} = -4 \text{ cm}^3/\text{sec}$ where V is the volume. How fast are the sides changing when they are equal to 5 centimeters in length?

The volume is $V = x^3$ where x is the length of a side of the cube. Therefore, the chain rule implies

$$-4 = \frac{dV}{dt} = 3x^2 \frac{dx}{dt}$$

and the problem is to find $\frac{dx}{dt}$ when $x = 5$. Therefore,

$$\frac{dx}{dt} = \frac{-4}{3(25)} = \frac{-4}{75} \text{ cm/second}$$

at this time.

Note there is no way of knowing the volume or the sides as a functions of t .

Example 5.10.2 One car travels north at 70 miles per hour and the other travels east at 60 miles per hour toward an intersection. How fast is the distance between the two cars changing when this distance equals five miles and the car heading north is at a distance of three miles from the intersection?

Let l denote the distance between the cars. Thus if x is the distance from the intersection of the car traveling east and y is the distance from the intersection of the car traveling north, $l^2 = x^2 + y^2$. When $y = 3$, it follows that $x = 4$. Therefore, at the instant described

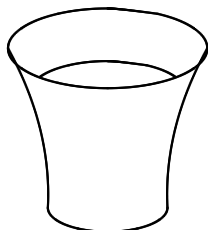
$$\begin{aligned} 2ll' &= 2xx' + 2yy' \\ 10l' &= 8(-60) + 6(-70) \end{aligned}$$

and so $l' = -90$ miles per hour at this instant.

5.11 Exercises

1. One car travels north at 70 miles per hour toward an intersection and the other travels east at 60 miles per hour away from the intersection. How fast is the distance between the two cars changing when this distance equals five miles and the car heading north is at a distance of three miles from the intersection?
2. A trash compactor compacts some trash which is in the shape of a box having a square base and a height equal to twice the length of a side of the base. Suppose each side of the base is changing at the rate of -3 inches per second. How fast is the volume changing when the side of the base equals 10 inches?
3. An isosceles triangle has two sides of equal length. Imagine such a triangle in which the two legs have length 8 inches and denote the included angle by θ and the area by A . Suppose $\frac{dA}{dt} = \sqrt{3}$ square inches per minute. How fast is θ changing when $\theta = \pi/6$ radians?
4. A point having coordinates (x, y) moves over the ellipse, $\frac{x^2}{4} + \frac{y^2}{9} = 1$. If $\frac{dy}{dt} = 2$, find $\frac{dx}{dt}$ at the point $(2, 3)$.
5. A spectator at a tennis tournament sits 10 feet from the end of the net and on the line determined by the net. He watches the ball go back and forth, and will have a very sore neck when he wakes up the next morning. How fast is the angle between his line of sight and the line determined by the net changing when the ball crosses over the net at a point 12 feet from the end assuming the ball travels at a speed of 60 miles per hour?
6. The surface area of a sphere of radius r equals $4\pi r^2$ and the volume of the ball of radius r equals $(4/3)\pi r^3$. A balloon in the shape of a ball is being inflated at the rate of 6 cubic inches per minute. How fast is the surface area changing when the volume of the ball equals 20π cubic inches?
7. * A mother cheetah attempts to fix dinner, a Thompson gazelle, for her hungry children. She moves at 100 feet per second while dinner travels at 80 feet per second. How fast is the distance between her and dinner decreasing when she is located at the point $(0, 40)$ feet and dinner is moving in the direction of the positive x axis at the point $(30, 0)$ feet? **Hint:** Let (x, y) denote the coordinates of the cheetah and let $(z, 0)$ denote the coordinates of the gazelle. Then if l is the desired distance, $l^2 = (x - z)^2 + y^2$. At the instant described, assuming the cheetah moves toward the gazelle at all times, $y'/x' = -4/3$ (why?) and also $\sqrt{(x')^2 + (y')^2} = 100$.
8. A six foot high man walks at a speed of 5 feet per second away from a light post which is 12 feet high that has the light right on the top. How fast is the end of his shadow moving when he is at a distance of 10 feet from the base of the light pole. How fast is the end of the shadow moving when he is 5 feet from the pole? (Assume he does not walk normally but instead oozes along like a giant amoeba so that his head is always exactly 6 feet above the ground.)
9. The volume of a right circular cone is $\frac{1}{3}\pi r^2 h$. Grain comes off a conveyor belt and falls to the ground making a right circular cone. It is observed that $r'(t) = .5$ feet per minute and $h'(t) = .3$ feet per minute. It is also known that the rate at which the grain falls off the conveyor belt is 100π cubic feet per minute. When the radius of the cone is 10 feet what is the height of the cone?

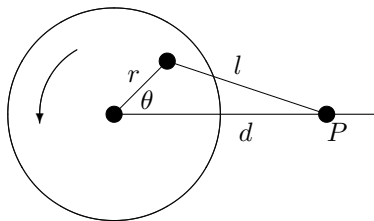
10. A hemispherical dish of radius 5 inches is sitting on a table. Soup is being poured in at the constant rate of 4 cubic inches per second. How fast is the level of soup rising when the radius of the top surface of the soup equals 3 inches? The volume of soup at depth y will be shown later to equal $V(y) = \pi \left(5y^2 - \frac{y^3}{3} \right)$.
11. A vase of water is sitting on a table.



It will be shown later that if $V(y)$ is the total volume of the vase up to height y , then $\frac{dV}{dy} = A(y)$ where $A(y)$ is the surface area of the top surface of the water at this height. (To see this is very reasonable, note that a little chunk of volume of the vase between heights y and $y + dy$ would be $dV = A(y) dy$, area times height.) Also, the rate at which the water evaporates is proportional to the surface area of the exposed water. Thus $\frac{dV}{dt} = -kA(y)$. Show $\frac{dy}{dt}$ is a constant even though the surface area of the exposed water is constantly changing in a typical vase.

12. A revolving search light at a prison is $1/4$ mile from the nearest point of a long wall. It makes one revolution per minute. How fast is the light traveling along the wall at a point which is $1/4$ mile from the nearest point? Give your answer in miles per hour.
13. A painter is on top of a 13 foot ladder which leans against a house. The base of the ladder is moving away from the house at the rate of 2 feet per second causing the top of the ladder to move down the house. How fast is the painter descending when the base of the ladder is at a distance of 5 feet from the house?
14. A rope fastened to the bow of a row boat has the other end wound around a windlass which is 4 feet above the level of the bow of the boat. The current pulls the boat away at the rate of 2 feet per second. How fast is the rope unwinding when the distance between the bow of the boat and the windlass is 10 feet?
15. A kite 100 feet above the ground is being blown away from the person holding its string in a direction parallel to the ground at the rate of 10 feet per second. At what rate must the string be let out when the length of string already let out is 200 feet?
16. A kite is moving horizontally at the rate of 10 feet per second and is 100 feet high. How fast is the angle of elevation of the kite string changing when 200 feet of string have been let out?
17. A certain volume of an ideal gas satisfies $PV = kT$ where T is the absolute temperature, P is the pressure, V is the volume and k is a constant which depends on the amount of the gas and the sort of gas in the sample. Find a formula for $\frac{dV}{dt}$ in terms of k, P, V and their derivatives.
18. A disposable cup is made in the shape of a right circular cone with height 5 inches and radius 2 inches. Water flows in to this conical cup at the rate of 4 cubic inches per minute. How fast is the water level rising when the water in the cone is three inches deep? The volume of a cone is $\frac{1}{3}\pi r^2 h$.

19. The two equal sides of an isosceles triangle have length x inches and the third leg has length y inches. Suppose $\frac{dx}{dt} = 2$ inches per minute and that the length of the other side changes in such a way that the area of the triangle is always 10 square inches. For θ the angle between the two equal sides, find $\frac{d\theta}{dt}$ when $x = 5$ inches.
20. A object is moving over the ellipse whose equation is $\frac{x^2}{9} + \frac{y^2}{4} = 1$. Near the point $\left(\frac{3\sqrt{2}}{2}, \sqrt{2}\right)$, it is observed that x is changing at the rate of 1 unit per second. How fast is y changing when the object is at this point?
21. * Consider the following diagram.



This represents a wheel which is spinning at a constant angular velocity equal to ω . Thus $\frac{d\theta}{dt} = \omega$. The circles represent axles and the two top lines joining the circles represent two rigid bars. The circle on the far right is allowed to slide on the indicated line. Thus, as the wheel turns, the point, P moves back and forth on the horizontal line shown. Find the velocity of P when $\theta = \pi/4$ in terms of r, l , and ω . **Hint:** Remember the law of cosines.

5.12 The Derivative And Optimization

There are existence theorems such as Theorem 2.15.11 on Page 79 which ensure a maximum or minimum of a function exists but in this section the goal is to give ways to find the maximum or minimum values of a function.

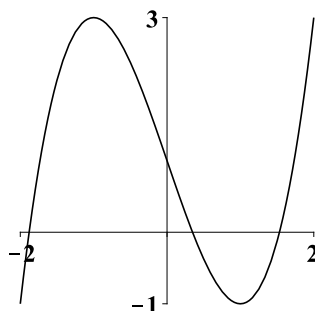
Suppose f is continuous on $[a, b]$. The minimum or maximum could occur at either end point or it could occur at a point in the open interval (a, b) . If it occurs at a point of the open interval (a, b) , say at x_0 , and if $f'(x_0)$ exists, then from Theorem 3.4.2 on Page 93 $f'(x_0) = 0$. Therefore, the following simple procedure can be used to locate the maximum or minimum of a function, f . Find all points, x , in (a, b) where $f'(x) = 0$ and all points, x , in (a, b) where $f'(x)$ does not exist. Then consider these points along with the end points of the interval. Evaluate f at the end points and at these points where the derivative is zero or does not exist. The largest must be the maximum value of f on the interval, $[a, b]$, and the smallest must be the minimum value of f on the interval, $[a, b]$. Typically, this involves checking only finitely many points.

Sometimes there are no end points. In this case, you do not necessarily know a maximum value or a minimum value of a continuous function even exists. However, if the function is differentiable and if a maximum or minimum exists, it can still be found by looking at the points where the derivative equals zero.

Example 5.12.1 Find the maximum and minimum values of the function, $f(x) = x^3 - 3x + 1$ on the interval $[-2, 2]$.

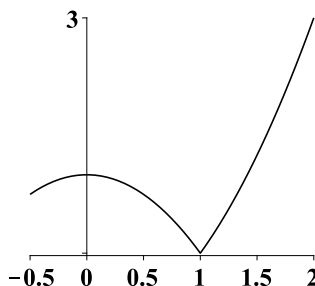
The points where $f'(x) = 3x^2 - 3 = 0$ are $x = 1$ or -1 . There are no points where the derivative does not exist. Therefore, evaluate the function at $-1, -2, 2$, and 1 . Thus $f(-1) = 3, f(1) = -1, f(-2) = -1$, and $f(2) = 3$. Therefore, the maximum value of the function occurs at the point -1 and 2 and has the value of 3 while the minimum

value of the function occurs at -1 and -2 and equals -1 . The following is a graph of this function.



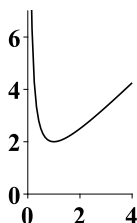
Example 5.12.2 Find the maximum and minimum values of the function $f(x) = |x^2 - 1|$ on the interval $[-.5, 2]$.

You should verify that this function fails to have a derivative at the point $x = 1$. For $x \in (-.5, 1)$ the function equals $1 - x^2$ and so its derivative equals zero at $x = 0$. For $x > 1$, the function equals $x^2 - 1$ and so there is no point larger than 1 where the derivative equals zero. Therefore, the points to look at are the end points, $-.5, 2$, the points where the derivative fails to exist, 1 and the point where the derivative equals zero, $x = 0$. Now $f(-.5) = .75$, $f(0) = 1$, $f(1) = 0$, and $f(2) = 3$. It follows the function achieves its maximum at the end point, $x = 2$ and its minimum at the point, $x = 1$ where the derivative fails to exist. The following is a graph of this function.



Example 5.12.3 Find the minimum value of the function $f(x) = x + \frac{1}{x}$ for $x \in (0, \infty)$.

The graph of this function is given below.



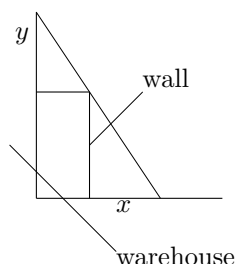
From the graph, it seems there should exist a minimum value at the bottom of the graph. To find it, take the derivative of f and set it equal to zero and then solve for the value of x . Thus

$$1 - \frac{1}{x^2} = 0$$

and so $x = 1$. The solution to the equation, $x = -1$ is of no interest because it is not greater than zero. Therefore, the minimum value of the function on the interval, $(0, \infty)$ equals $f(1) = 2$ as suggested by the graph.

Example 5.12.4 *An eight foot high wall stands one foot from a warehouse. What is the length of the shortest ladder which extends from the ground to the warehouse.*

A diagram of this situation is the following picture.



In this picture, the slanted line represents the ladder and x and y are as shown. By similar triangles, $y/1 = 8/x$. Therefore, $xy = 8$. From the Pythagorean theorem the length of the ladder equals $\sqrt{1 + y^2} + \sqrt{x^2 + 64}$. Now using the relation between x and y , the function of a single variable, x , to minimize is

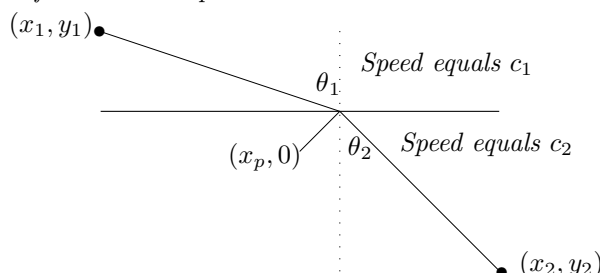
$$f(x) = \sqrt{1 + \frac{64}{x^2}} + \sqrt{x^2 + 64},$$

Clearly $x > 0$ so there are no endpoints to worry about. (Why?) Also the function is differentiable and so it suffices to consider only points where the derivative equals zero. This is a little messy but finally

$$f'(x) = \frac{1}{\sqrt{(x^2 + 64)}} \frac{-64 + x^3}{x^2}$$

and the value where this equals zero is $x = 4$. It follows the shortest ladder is of length $\sqrt{1 + \frac{64}{4^2}} + \sqrt{4^2 + 64} = 5\sqrt{5}$ feet.

Example 5.12.5 *Fermat's principle says that light travels on a path which will minimize the total time. Consider the following picture of light passing from (x_1, y_1) to (x_2, y_2) as shown. The angle θ_1 is called the angle of incidence while the angle, θ_2 is called the angle of refraction. The picture indicates a situation in which $c_1 > c_2$.*



Then define by x the quantity $x_p - x_1$. What is the relation between θ_1 and θ_2 ?

The time it takes for the light to go from (x_1, y_1) to the point $(x_p, 0)$ equals

$$\sqrt{x^2 + y_1^2}/c_1$$

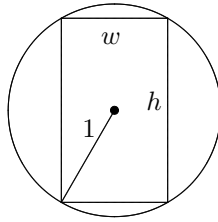
and the time it takes to go from $(x_p, 0)$ to (x_2, y_2) is $\sqrt{(x_2 - x_1 - x)^2 + y_2^2}/c_2$. Therefore, the total time is $T = \frac{\sqrt{x^2 + y_1^2}}{c_1} + \frac{\sqrt{(x_2 - x_1 - x)^2 + y_2^2}}{c_2}$. Thus T is minimized if

$$\begin{aligned}\frac{dT}{dx} &= \frac{d}{dx} \left(\frac{\sqrt{x^2 + y_1^2}}{c_1} + \frac{\sqrt{(x_2 - x_1 - x)^2 + y_2^2}}{c_2} \right) \\ &= \frac{x}{c_1 \sqrt{x^2 + y_1^2}} - \frac{(x_2 - x_1 - x)}{c_2 \sqrt{(x_2 - x_1 - x)^2 + y_2^2}} \\ &= \frac{\sin(\theta_1)}{c_1} - \frac{\sin(\theta_2)}{c_2} = 0\end{aligned}$$

at this point. Therefore, this yields the desired relation between θ_1 and θ_2 . This is called Snell's law.

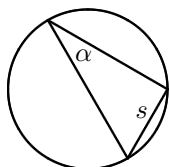
5.13 Exercises

- Find the maximum and minimum values for the following functions defined on the given intervals.
 - $x^3 - 3x^2 + x - 7$, $[0, 4]$
 - $\ln(x^2 - x + 2)$, $[0, 2]$
 - $x^3 + 3x$, $[-1, 10]$
 - $\frac{x^2 + 1 + 3x^3}{3x^2 + 5}$, $[-1, 1]$
 - $\sin(x^3 - x)$, $[-1, 1]$
 - $x^2 - x \tan x$, $[-1, 1]$
 - $1 - 2x^2 + x^4$, $[-2, 2]$
 - $\ln(2 - 2x^2 + x^4)$, $[-1, 2]$
 - $x^2 + 4x - 8$, $[-4, 2]$
 - $x^2 - 3x + 6$, $[-2, 4]$
 - $-x^2 + 3x$, $[-4, 2]$
 - $x + \frac{1}{x}$, $(0, \infty)$
- A cylindrical can is to be constructed to hold 30 cubic inches. The top and bottom of the can are constructed of a material costing one cent per square inch and the sides are constructed of a material costing 2 cents per square inch. Find the minimum cost for such a can.
- Two positive numbers sum to 8. Find the numbers if their product is to be as large as possible.
- The ordered pair, (x, y) is on the ellipse, $x^2 + 4y^2 = 4$. Form the rectangle which has (x, y) as one end of a diagonal and $(0, 0)$ at the other end. Find the rectangle of this sort which has the largest possible area.
- A rectangle is inscribed in a circle of radius r . Find the formula for the rectangle of this sort which has the largest possible area.
- A point is picked on the ellipse, $x^2 + 4y^2 = 4$ which is in the first quadrant. Then a line tangent to this point is drawn which intersects the x axis at a point, x_1 and the y axis at the point y_1 . The area of the triangle formed by the y axis, the x axis, and the line just drawn is thus $\frac{x_1 y_1}{2}$. Out of all possible triangles formed in this way, find the one with smallest area.
- Find maximum and minimum values if they exist for the function, $f(x) = \frac{\ln x}{x}$ for $x > 0$.
- Describe how you would find the maximum value of the function, $f(x) = \frac{\ln x}{2 + \sin x}$ for $x \in (0, 6)$ if it exists. **Hint:** You might want to use a calculator to graph this and get an idea what is going on.
- A rectangular beam of height h and width w is to be sawed from a circular log of radius 1 foot. Find the dimensions of the strongest such beam assuming the strength is of the form kh^2w . Here k is some constant which depends on the type of wood used.



10. A farmer has 600 feet of fence with which to enclose a rectangular piece of land that borders a river. If he can use the river as one side, what is the largest area that he can enclose.
11. An open box is to be made by cutting out little squares at the corners of a rectangular piece of cardboard which is 20 inches wide and 40 inches long, and then folding up the rectangular tabs which result. What is the largest possible volume which can be obtained?
12. A feeding trough is to be made from a rectangular piece of metal which is 3 feet wide and 12 feet long by folding up two rectangular pieces of dimension one foot by 12 feet. What is the best angle for this fold?
13. Find the dimensions of the right circular cone which has the smallest area given the volume is 30π cubic inches. The volume of the right circular cone is $(1/3)\pi r^2 h$ and the area of the cone is $\pi r\sqrt{h^2 + r^2}$.
14. A wire of length 10 inches is cut into two pieces, one of length x and the other of length $10 - x$. One piece is bent into the shape of a square and the other piece is bent into the shape of a circle. Find the two lengths such that the sum of the areas of the circle and the square is as large as possible. What are the lengths if the sum of the two areas is to be as small as possible.
15. A hiker begins to walk to a cabin in a dense forest. He is walking on a road which runs from East to West and the cabin is located exactly one mile north of a point two miles down the road. He walks 5 miles per hour on the road but only 3 miles per hour in the woods. Find the path which will minimize the time it takes for him to get to the cabin.
16. A park ranger needs to get to a fire observation tower which is one mile from a long straight road in a dense forest. The point on the road closest to the observation tower is 10 miles down the road on which the park ranger is standing. Knowing that he can walk at 4 miles per hour on the road but only one mile per hour in the forest, how far down the road should he walk before entering the forest, in order to minimize the travel time?
17. A refinery is on a straight shore line. Oil needs to flow from a mooring place for oil tankers to this refinery. Suppose the mooring place is two miles off shore from a point on the shore 8 miles away from the refinery which is also on the shore and that it costs five times as much to lay pipe under water than above the ground. Describe the most economical route for a pipeline from the mooring place to the refinery.
18. Two hallways, one 5 feet wide and the other 6 feet wide meet. It is desired to carry a ladder horizontally around the corner. What is the longest ladder which can be carried in this way? **Hint:** Consider a line through the inside corner which extends to the opposite walls. The shortest such line will be the length of the longest ladder. You might also consider Example 4.1.3 on Page 107.
19. A triangle is inscribed in a circle in such a way that one side of the triangle is always the same length, s . Show that out of all such triangles the maximum area is obtained when the triangle is an isosceles triangle. **Hint:** From theorems in plane geometry, the angle opposite the side having fixed length is a constant, no

matter how you draw the triangle. (See Problem 32 on Page 30 to see why this is if the geometry interests you.) Use the law of sines and this fact. In the following picture, α is a constant.

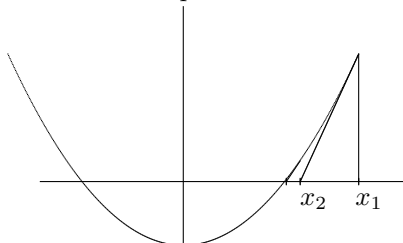


20. A window is to be constructed for the wall of a church which is to consist of a rectangle of height b surmounted by a half circle of radius a . Suppose the total perimeter of the window is to be no more than $4\pi + 8$ feet. Find the dimensions of the window which will admit the most light.
21. You know $\lim_{x \rightarrow \infty} \ln x = \infty$. Show that if $\alpha > 0$, then $\lim_{x \rightarrow \infty} \frac{\ln x}{x^\alpha} = 0$.
22. * A parabola opens down. The vertex is at the point, $(0, a)$ and the parabola intercepts the x axis at the points $(-b, 0)$ and $(b, 0)$. A tangent line to the parabola is drawn in the first quadrant which has the property that the triangle formed by this tangent line and the x and y axes has smallest possible area. Find a relationship between a and b such that the normal line to the point of tangency passes through $(0, 0)$. Also determine what kind of triangle this is.
23. * Suppose p and q are two positive numbers larger than 1 which satisfy $\frac{1}{p} + \frac{1}{q} = 1$. Now let a and b be two positive numbers and consider $f(t) = \frac{1}{p}(at)^p + \frac{1}{q}\left(\frac{b}{t}\right)^q$ for $t > 0$. Show the minimum value of f is ab . Prove the important inequality,
- $$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$
24. * Using Problem 23 establish the following magnificent inequality which is a case of Holder's inequality. For $\frac{1}{p} + \frac{1}{q} = 1$, and a_i, b_i nonnegative numbers,
- $$\sum_{i=1}^n a_i b_i \leq \left(\sum_{i=1}^n a_i^p \right)^{1/p} \left(\sum_{i=1}^n b_i^q \right)^{1/q}.$$
- Hint:** You might consider using the inequality of Problem 23 on the following expression.
- $$\sum_{i=1}^n \frac{a_i}{\left(\sum_{i=1}^n a_i^p \right)^{1/p}} \frac{b_i}{\left(\sum_{i=1}^n b_i^q \right)^{1/q}}$$
25. * Using the inequality of Problem 24 establish the following inequality for $p > 1$ which is called the triangle inequality.
- $$\left(\sum_{i=1}^n |a_i + b_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^n |a_i|^p \right)^{1/p} + \left(\sum_{i=1}^n |b_i|^p \right)^{1/p}.$$
- The reason this inequality is important is that people sometimes define the length of a vector in $\mathbb{R}^n$, $\mathbf{a} \equiv (a_1, \dots, a_n)$ in the following way.
- $$\|\mathbf{a}\|_p \equiv \left(\sum_{k=1}^n |a_k|^p \right)^{1/p}$$
- and the triangle inequality says
- $$\|\mathbf{a} + \mathbf{b}\|_p \leq \|\mathbf{a}\|_p + \|\mathbf{b}\|_p$$
- which turns out to be one of the axioms for a norm.

5.14 The Newton Raphson Method

The Newton Raphson method is a way to get approximations of solutions to various equations. For example, suppose you want to find $\sqrt{2}$. The existence of $\sqrt{2}$ is not difficult to establish by considering the continuous function, $f(x) = x^2 - 2$ which is

negative at $x = 0$ and positive at $x = 2$. Therefore, by the intermediate value theorem, there exists $x \in (0, 2)$ such that $f(x) = 0$ and this x must equal $\sqrt{2}$. The problem consists of how to find this number, not just to prove it exists. The following picture illustrates the procedure of the Newton Raphson method.



In this picture, a first approximation, denoted in the picture as x_1 is chosen and then the tangent line to the curve $y = f(x)$ at the point $(x_1, f(x_1))$ is obtained. The equation of this tangent line is

$$y - f(x_1) = f'(x_1)(x - x_1).$$

Then extend this tangent line to find where it intersects the x axis. In other words, set $y = 0$ and solve for x . This value of x is denoted by x_2 . Thus

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

This second point, x_2 is the second approximation and the same process is done for x_2 that was done for x_1 in order to get the third approximation, x_3 . Thus

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}.$$

Continuing this way, yields a sequence of points, $\{x_n\}$ given by

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \quad (5.25)$$

which hopefully has the property that $\lim_{n \rightarrow \infty} x_n = x$ where $f(x) = 0$. You can see from the above picture that this must work out in the case of $f(x) = x^2 - 2$.

Now carry out the computations in the above case for $x_1 = 2$ and $f(x) = x^2 - 2$. From 5.25,

$$x_2 = 2 - \frac{2}{4} = 1.5.$$

Then

$$x_3 = 1.5 - \frac{(1.5)^2 - 2}{2(1.5)} \leq 1.417,$$

$$x_4 = 1.417 - \frac{(1.417)^2 - 2}{2(1.417)} = 1.414216302046577,$$

What is the true value of $\sqrt{2}$? To several decimal places this is $\sqrt{2} = 1.414213562373095$, showing that the Newton Raphson method has yielded a very good approximation after only a few iterations, even starting with an initial approximation, 2, which was not very good.

This method does not always work. For example, suppose you wanted to find the solution to $f(x) = 0$ where $f(x) = x^{1/3}$. You should check that the sequence of iterates which results does not converge. This is because, starting with x_1 the above procedure

yields $x_2 = -2x_1$ and so as the iteration continues, the sequence oscillates between positive and negative values as its absolute value gets larger and larger.

However, if $f(x_0) = 0$ and $f''(x) > 0$ for x near x_0 , you can draw a picture to show that the method will yield a sequence which converges to x_0 provided the first approximation, x_1 is taken sufficiently close to x_0 . Similarly, if $f''(x) < 0$ for x near x_0 , then the method produces a sequence which converges to x_0 provided x_1 is close enough to x_0 .

5.15 Exercises

1. By drawing representative pictures, show convergence of the Newton Raphson method in the cases described above where $f''(x) > 0$ near x_0 or $f''(x) < 0$ near x_0 .
2. Draw some graphs to illustrate the Newton Raphson method does not yield a convergent sequence in the case where $f(x) = x^{1/3}$.
3. Use the Newton Raphson method to approximate the first positive solution of $x - \tan x = 0$. **Hint:** You may need to use a calculator to deal with $\tan x$.
4. Use the Newton Raphson method to compute an approximation to $\sqrt{3}$ which is within 10^{-6} of the true value. Explain how you know you are this close.
5. Using the Newton Raphson method and an appropriate picture, discuss the convergence of the recursively defined sequence $x_{n+1} = ((p-1)x_n + cx_n^{1-p})/p$ where $x_1, c > 0$ and $p > 1$.

Antiderivatives

6.0.1 Outcomes

1. Understand and describe the concept of an initial value problem.
2. Understand and use the methods of substitution, integration by parts, trig. substitutions and partial fractions to find solutions to initial value problems.
3. Find areas between curves.
4. Find volumes using the method of cross sections and shells.
5. Find lengths of curves.
6. Find areas of surfaces of revolution.
7. Understand and use methods for solving linear and separable differential equations and understand some examples of these.
8. Understand the concept of fluid pressure of an incompressible fluid and the force it produces.
9. Be able to do simple problems involving work.

A differential equation is an equation which involves an unknown function and its derivatives. Differential equations are the unifying idea in this chapter. Many interesting problems may be solved by formulating them as solutions of a suitable differential equation with initial condition called an initial value problem.

6.1 Initial Value Problems

The initial value problem is to find a function, $y(x)$ for $x \in [a, b]$ such that

$$y'(x) = f(x, y(x)), \quad y(a) = y_0.$$

Various assumptions are made on $f(x, y)$. At this time it is assumed that f does not depend on y and f is a given continuous defined $[a, b]$. Thus the initial value problem of interest here is one of the form

$$y'(x) = f(x), \quad y(a) = y_0. \tag{6.1}$$

As an example of an application of an initial value problem, recall the discussion which led to the derivative on Page 85. There $r(t)$ was the x coordinate of a point moving on the x axis at time t and it was shown there that the velocity of this object was $r'(t)$. When this is positive, the object is moving to the right on the x axis and

when it is negative, the object is moving to the left. Suppose $r'(t)$ was known, say $r'(t) = f(t)$ along with $r(0) = x_0$ and you wanted to find $r(t)$. Then you are really asking for the solution to an initial value problem of the form

$$r'(t) = f(t), \quad r(0) = x_0.$$

Theorem 6.1.1 *There is at most one solution to the initial value problem, 6.1 which is continuous on $[a, b]$.*

Proof: Suppose both $A(x)$ and $B(x)$ are solutions to this initial value problem for $x \in (a, b)$. Then letting $H(x) \equiv A(x) - B(x)$, it follows that $H(a) = 0$ and $H'(x) = A'(x) - B'(x) = f(x) - f(x) = 0$. Therefore, from Corollary 3.6.4 on Page 99, it follows $H(x)$ equals a constant on (a, b) . By continuity of H , this constant must equal $H(a) = 0$.

The main difficulty in solving these initial value problems like 6.1 is in finding a function whose derivative equals the given function. This is in general a very hard problem, although techniques for doing this are presented later which will cover many cases of interest. The functions whose derivatives equal a given function, $f(x)$, are called antiderivatives and there is a special notation used to denote them.

Definition 6.1.2 *Let f be a function. $\int f(x) dx$ denotes the set of antiderivatives of f . Thus $F \in \int f(x) dx$ means $F'(x) = f(x)$. It is customary to refer to $f(x)$ as the integrand. This symbol is also called the indefinite integral and sometimes is referred to as an integral although this last usage is not correct.*

The reason this last usage is not correct is that the integral of a function is a single number not a whole set of functions. Nevertheless, you can't escape the fact that it is common usage to call that symbol an integral.

Lemma 6.1.3 *Suppose $F, G \in \int f(x) dx$ for $x \in (a, b)$. Then there exists a constant, C such that for all $x \in (a, b)$, $F(x) = G(x) + C$.*

Proof:

$$F'(x) - G'(x) = f(x) - f(x) = 0$$

for all $x \in (a, b)$. Consequently, by Corollary 3.6.4 on Page 99, $F(x) - G(x) = C$. This proves the lemma.

There is another simple lemma about antiderivatives.

Lemma 6.1.4 *If a and b are nonzero real numbers, and if $\int f(x) dx$ and $\int g(x) dx$ are nonempty, then*

$$\int (af(x) + bg(x)) dx = a \int f(x) dx + b \int g(x) dx$$

Proof: The symbols on the two sides of the equation denote sets of functions. It is necessary to verify the two sets of functions are the same. Suppose then that $F \in \int f(x) dx$ and $G \in \int g(x) dx$. Then $aF + bG$ is a typical function of the right side of the equation. Taking the derivative of this function, yields $af(x) + bg(x)$ and so this shows the set of functions on the right side is a subset of the set of functions on the left.

Now take $H \in \int (af(x) + bg(x)) dx$ and pick $F \in \int f(x) dx$. Then $aF \in a \int f(x) dx$ and

$$(H(x) - aF(x))' = af(x) + bg(x) - af(x) = bg(x)$$

showing that

$$H - aF \in \int bg(x) dx = b \int g(x) dx$$

because $b \neq 0$. Therefore

$$H \in aF + b \int g(x) dx \subseteq a \int f(x) dx + b \int g(x) dx.$$

This has shown the two sets of functions are the same and proves the lemma.

From Lemma 6.1.3 it follows that if $F(x) \in \int f(x) dx$, then every other function in $\int f(x) dx$ is of the form $F(x) + C$ for a suitable constant, C . Thus it is customary to write

$$\int f(x) dx = F(x) + C$$

where it is understood that C is an arbitrary constant, called a constant of integration.

From the formulas for derivatives presented earlier, the following table of antiderivatives follows.

$f(x)$	$\int f(x) dx$
$x^n, n \neq -1$	$\frac{x^n}{n+1} + C$
x^{-1}	$\ln x + C$
$\cos(x)$	$\sin(x) + C$
$\sin(x)$	$-\cos(x) + C$
$\sec^2(x)$	$\tan(x) + C$
e^x	$e^x + C$
$\cosh(x)$	$\sinh(x) + C$
$\sinh(x)$	$\cosh(x) + C$
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin(x) + C$
$\frac{1}{\sqrt{1+x^2}}$	$\operatorname{arcsinh} x + C$
$\frac{1}{x\sqrt{x^2-1}}$	$\operatorname{arcsec} x + C$
$\frac{1}{1+x^2}$	$\arctan(x) + C$

The above table is a good starting point for other antiderivatives. For example,

Proposition 6.1.5 *Let $\sum_{k=0}^n a_k x^k$ be a polynomial. Then*

$$\int \sum_{k=0}^n a_k x^k dx = \sum_{k=0}^n a_k \frac{x^{k+1}}{k+1} + C$$

Proof: This follows from the above table and Lemma 6.1.4.

Example 6.1.6 *Suppose the velocity of an object moving on the x axis is given by the function $\cos(t)$ and when $t = 0$, the object is at 0. Find the position of the object.*

As explained above, you need to solve

$$r'(t) = \cos(t), \quad r(0) = 0.$$

To do this, note that from the differential equation, $r(t) = \sin(t) + C$ and it only remains to find the constant, C such that $r(0) = 0$. Thus $0 = r(0) = \sin(0) + C = C$ and so $r(t) = \sin(t)$.

Example 6.1.7 Consider the same problem as in Example 6.1.6 but this time let $r(1) = 5$. What is the answer in this case?

From the differential equation it is still the case that $r(t) = \sin(t) + C$ but now C needs to be chosen such that $r(1) = 5$. Thus $5 = r(1) = \sin(1) + C$ and so $C = 5 - \sin(1)$. Therefore, the solution to the initial value problem,

$$r'(t) = \cos(t), \quad r(1) = 5$$

is $r(t) = \sin(t) + (5 - \sin(1))$.

They are all like this. You find an antiderivative. The answer you want is this antiderivative added to an appropriate constant chosen to satisfy the given initial condition. The problem is in finding the antiderivative. In general, this is a very hard problem but there are techniques for solving it in some cases. The next section is one such techniques.

6.2 The Method Of Substitution

The method of substitution is based on the following formula which is merely a restatement of the chain rule.

$$\int f(g(x)) g'(x) dx = F(x) + C, \quad (6.2)$$

where $F'(y) = f(y)$. Here are some examples of the method of substitution.

Example 6.2.1 Find $\int \sin(x) \cos(x) \sqrt{3 + 2^{-1} \sin^2(x)} dx$

Note it is of the form given in 6.2 with $g(x) = 2^{-1} \sin^2(x)$ and $F(u) = \frac{2}{3} \left(\sqrt{3+u} \right)^3$. Therefore,

$$\int \sin(x) \cos(x) \sqrt{3 + 2^{-1} \sin^2(x)} dx = \frac{1}{12} \left(\sqrt{12 + 2 \sin^2(x)} \right)^3 + C.$$

Example 6.2.2 Find $\int (1 + x^2)^6 2x dx$.

This is a special case of 6.2 when $g(x) = 1 + x^2$ and $F(u) = u^7/7$. Therefore, the answer is

$$\int (1 + x^2)^6 2x dx = (1 + x^2)^7 / 7 + C.$$

Example 6.2.3 Find $\int (1 + x^2)^6 x dx$

This equals

$$\frac{1}{2} \int (1 + x^2)^6 2x dx = \frac{1}{2} \frac{(1 + x^2)^7}{7} + C = \frac{(1 + x^2)^7}{14} + C.$$

Actually, it is not necessary to recall 6.2 and massage things to get them in that form. There is a trick based on the Leibniz notation for the derivative which is very useful and illustrated in the following example.

Example 6.2.4 Find $\int \cos(2x) \sin^2(2x) dx$.

Let $u = \sin(2x)$. Then $\frac{du}{dx} = 2 \cos(2x)$. Now formally

$$\frac{du}{2} = \cos(2x) dx.$$

Thus

$$\begin{aligned} \int \cos(2x) \sin^2(2x) dx &= \frac{1}{2} \int u^2 du = \frac{u^3}{6} + C \\ &= \frac{(\sin(2x))^3}{6} + C \end{aligned}$$

The expression $(1/2) du$ replaced the expression $\cos(2x) dx$ which occurs in the original problem and the resulting problem in terms of u was much easier. This was solved and finally the original variable was replaced. When using this method, it is a good idea to check your answer to be sure you have not made a mistake. Thus in this example, the chain rule implies $\left(\frac{(\sin(2x))^3}{6}\right)' = \cos(2x) \sin^2(2x)$ which verifies the answer is right. Here is another example.

Example 6.2.5 Find $\int \sqrt[3]{2x+7} dx$.

In this example $u = 2x + 7$ so that $du = 2dx$. Then

$$\begin{aligned} \int \sqrt[3]{2x+7} dx &= \int \sqrt[3]{\overbrace{u-7}^x \overbrace{\frac{1}{2}}^{dx}} du \\ &= \int \left(\frac{1}{4} u^{4/3} - \frac{7}{4} u^{1/3} \right) du \\ &= \frac{3}{28} u^{7/3} - \frac{21}{16} u^{4/3} + C \\ &= \frac{3}{28} (2x+7)^{7/3} - \frac{21}{16} (2x+7)^{4/3} + C \end{aligned}$$

Example 6.2.6 Find $\int x 3^{x^2} dx$

Let $u = 3^{x^2}$ so that $\frac{du}{dx} = 2x \ln(3) 3^{x^2}$ and $\frac{du}{2 \ln(3)} = x 3^{x^2} dx$. Thus

$$\begin{aligned} \int x 3^{x^2} dx &= \frac{1}{2 \ln(3)} \int du = \frac{1}{2 \ln(3)} [u + C] \\ &= \frac{1}{2 \ln(3)} 3^{x^2} + \left(\frac{1}{2 \ln(3)} \right) C \end{aligned}$$

Since the constant is an arbitrary constant, this is written as

$$\frac{1}{2 \ln(3)} 3^{x^2} + C.$$

Example 6.2.7 Find $\int \cos^2(x) dx$

Recall that $\cos(2x) = \cos^2(x) - \sin^2(x)$ and $1 = \cos^2(x) + \sin^2(x)$. Then subtracting and solving for $\cos^2(x)$,

$$\cos^2(x) = \frac{1 + \cos(2x)}{2}.$$

Therefore,

$$\int \cos^2(x) \, dx = \int \frac{1 + \cos(2x)}{2} \, dx$$

Now letting $u = 2x$, $du = 2dx$ and so

$$\begin{aligned} \int \cos^2(x) \, dx &= \int \frac{1 + \cos(u)}{4} \, du \\ &= \frac{1}{4}u + \frac{1}{4} \sin u + C \\ &= \frac{1}{4}(2x + \sin(2x)) + C. \end{aligned}$$

Also

$$\int \sin^2(x) \, dx = -\frac{1}{2} \cos x \sin x + \frac{1}{2}x + C$$

which is left as an exercise.

Example 6.2.8 Find $\int \tan(x) \, dx$

Let $u = \cos x$ so that $du = -\sin(x) \, dx$. Then writing the antiderivative in terms of u , this becomes $\int \frac{-1}{u} \, du$. At this point, recall that $(\ln |u|)' = 1/u$. Thus this antiderivative is $-\ln |u| + C = \ln |u^{-1}| + C$ and so $\int \tan(x) \, dx = \ln |\sec x| + C$.

This illustrates a general procedure.

Procedure 6.2.9 $\int \frac{f'(x)}{f(x)} \, dx = \ln |f(x)| + C$.

This follows from the chain rule.

Example 6.2.10 Find $\int \sec(x) \, dx$.

This is usually done by a trick. You write as

$$\int \frac{\sec(x)(\sec(x) + \tan(x))}{(\sec(x) + \tan(x))} \, dx$$

and note that the numerator of the integrand is the derivative of the denominator. Thus

$$\int \sec(x) \, dx = \ln |\sec(x) + \tan(x)| + C.$$

Example 6.2.11 Find $\int \csc(x) \, dx$.

This is done like the antiderivatives for the secant.

$$\frac{d}{dx} \csc(x) = -\csc(x) \cot(x)$$

and $\frac{d}{dx} \cot(x) = -\csc^2(x)$. Write the integral as

$$-\int \frac{-\csc(x)(\cot(x) + \csc(x))}{(\cot(x) + \csc(x))} \, dx = -\ln |\cot(x) + \csc(x)| + C.$$

6.3 Exercises

- Find the indicated antiderivatives.

(a) $\int \frac{x}{\sqrt{2x-3}} \, dx$

(b) $\int x(3x^2 + 6)^5 \, dx$

(c) $\int x \sin(x^2) \, dx$

- (d) $\int \sin^3(2x) \cos(2x) dx$
- (e) $\int \frac{1}{\sqrt{1+4x^2}} dx$ **Hint:** Remember the $\sinh^{-1}$ function and its derivative.
2. Solve the initial value problems.
- (a) $\frac{dy}{dx} = \frac{x}{\sqrt{2x-3}}, y(2) = 1$
- (b) $\frac{dy}{dx} = 5x(3x^2 + 6)^5, y(0) = 3$
- (c) $\frac{dy}{dx} = 3x^2 \sin(2x^3), y(1) = 1$
- (d) $y'(x) = \frac{1}{\sqrt{1+3x^2}}, y(1) = 1$
- (e) $y'(x) = \sec(x), y(0) = 3$
- (f) $y'(x) = x \csc(x^2), y(1) = 1$
3. An object moves on the x axis having velocity equal to $\frac{3t^3}{7+t^4}$. Find the position of the object given that at $t = 1$, it is at the point 2.
4. An object moves on the x axis having velocity equal to $t \sin(2t^2)$. Find the position of the object given that at $t = 1$, it is at the point 1.
5. An object moves on the x axis having velocity equal to $\sec(t)$. Find the position of the object given that at $t = 1$, it is at the point -2 .
6. Find the indicated antiderivatives.
- (a) $\int \sec(3x) dx$
- (b) $\int \sec^2(3x) \tan(3x) dx$
- (c) $\int \frac{1}{3+5x^2} dx$
- (d) $\int \frac{1}{\sqrt{5-4x^2}} dx$
- (e) $\int \frac{3}{x\sqrt{4x^2-5}} dx$
7. Find the indicated antiderivatives.
- (a) $\int x \cosh(x^2 + 1) dx$
- (b) $\int x^3 5^{x^4} dx$
- (c) $\int \sin(x) 7^{\cos(x)} dx$
- (d) $\int x \sin(x^2) dx$
- (e) $\int x^5 \sqrt{2x^2 + 1} dx$ **Hint:** Let $u = 2x^2 + 1$.
8. Find $\int \sin^2(x) dx$. **Hint:** Derive and use $\sin^2(x) = \frac{1 - \cos(2x)}{2}$.
9. Find the indicated antiderivatives.
- (a) $\int \frac{\ln x}{x} dx$
- (b) $\int \frac{x^3}{3+x^4} dx$
- (c) $\int \frac{1}{x^2+2x+2} dx$ **Hint:** Complete the square in the denominator and then let $u = x + 1$.
- (d) $\int \frac{1}{\sqrt{4-x^2}} dx$
- (e) $\int \frac{1}{x\sqrt{x^2-9}} dx$ **Hint:** Let $x = 3u$.
- (f) $\int \frac{\ln(x^2)}{x} dx$
- (g) Find $\int \frac{x^3}{\sqrt{6x^2+5}} dx$
- (h) Find $\int x \sqrt[3]{6x+4} dx$
10. Find the indicated antiderivatives.
- (a) $\int x \sqrt{2x+4} dx$
- (b) $\int x \sqrt{3x+2} dx$
- (c) $\int \frac{1}{\sqrt{36-25x^2}} dx$
- (d) $\int \frac{1}{\sqrt{9-4x^2}} dx$
- (e) $\int \frac{1}{\sqrt{1+4x^2}} dx$
- (f) $\int \frac{x}{\sqrt{(3x-1)}} dx$
- (g) $\int \frac{x}{\sqrt{5x+1}} dx$
- (h) $\int \frac{1}{x\sqrt{9x^2-4}} dx$
- (i) $\int \frac{1}{\sqrt{9+4x^2}} dx$
11. Find $\int \frac{1}{x^{1/3}+x^{1/2}} dx$. **Hint:** Try letting $x = u^6$ and use long division.
12. Suppose f is a function defined on $\mathbb{R}$ and it satisfies the functional equation, $f(a+b) = f(a) + f(b)$. Suppose also $f'(0) = k$. Find $f(x)$.
13. Suppose f is a function defined on $\mathbb{R}$ having values in $(0, \infty)$ and it satisfies the functional equation, $f(a+b) = f(a)f(b)$. Suppose also $f'(0) = k$. Find $f(x)$.
14. Suppose f is a function defined on $(0, \infty)$ having values in $\mathbb{R}$ and it satisfies the functional equation, $f(ab) = f(a) + f(b)$. Suppose also $f'(1) = k$. Find $f(x)$.
15. Suppose f is a function defined on $\mathbb{R}$ and it satisfies the functional equation, $f(a+b) = f(a) + f(b) + 3ab$. Suppose also that $\lim_{h \rightarrow 0} \frac{f(h)}{h} = 7$. Find $f(x)$ if possible.

6.4 Integration By Parts

Another technique for finding antiderivatives is called integration by parts and is based on the product rule. Recall the product rule. If u' and v' exist, then

$$(uv)'(x) = u'(x)v(x) + u(x)v'(x). \quad (6.3)$$

Therefore,

$$(uv)'(x) - u'(x)v(x) = u(x)v'(x)$$

Proposition 6.4.1 *Let u and v be differentiable functions for which $\int u(x)v'(x) dx$ and $\int u'(x)v(x) dx$ are nonempty. Then*

$$uv - \int u'(x)v(x) dx = \int u(x)v'(x) dx. \quad (6.4)$$

Proof: Let $F \in \int u'(x)v(x) dx$. Then

$$(uv - F)' = (uv)' - F' = (uv)' - u'v = uv'$$

by the chain rule. Therefore every function from the left in 6.4 is a function found in the right side of 6.4. Now let $G \in \int u(x)v'(x) dx$. Then $(uv - G)' = -uv' + (uv)' = u'v$ by the product rule. It follows that $uv - G \in \int u'(x)v(x) dx$ and so $G \in uv - \int u'(x)v(x) dx$. Thus every function from the right in 6.4 is a function from the left. This proves the proposition.

Example 6.4.2 Find $\int x \sin(x) dx$

Let $u(x) = x$ and $v'(x) = \sin(x)$. Then applying 6.4,

$$\begin{aligned} \int x \sin(x) dx &= (-\cos(x))x - \int (-\cos(x)) dx \\ &= -x \cos(x) + \sin(x) + C. \end{aligned}$$

Example 6.4.3 Find $\int x \ln(x) dx$

Let $u(x) = \ln(x)$ and $v'(x) = x$. Then from 6.4,

$$\begin{aligned} \int x \ln(x) dx &= \frac{x^2}{2} \ln(x) - \int \frac{x^2}{2} \left(\frac{1}{x} \right) \\ &= \frac{x^2}{2} \ln(x) - \int \frac{x}{2} \\ &= \frac{x^2}{2} \ln(x) - \frac{1}{4}x^2 + C \end{aligned}$$

Example 6.4.4 Find $\int \arctan(x) dx$

Let $u(x) = \arctan(x)$ and $v'(x) = 1$. Then from 6.4,

$$\begin{aligned} \int \arctan(x) dx &= x \arctan(x) - \int x \left(\frac{1}{1+x^2} \right) dx \\ &= x \arctan(x) - \frac{1}{2} \int \frac{2x}{1+x^2} dx \\ &= x \arctan(x) - \frac{1}{2} \ln(1+x^2) + C. \end{aligned}$$

Sometimes you want to find antiderivatives for something like $\int f g dx$ where $f^{(m)} = 0$ for some positive integer, m . For example, $\int x^5 \sin x dx$. If you do integration by parts repeatedly, what do you get? Let $G'_1 = g, G'_2 = G_1, G'_3 = G_2$ etc. Then the first application of integration by parts yields $f G_1 - \int G_1 f' dx$. The next application of integration by parts yields $f G_1 - G_2 f' + \int G_2 f'' dx$. Yet another application of integration by parts yields $f G_1 - G_2 f' + G_3 f'' - \int G_3 f''' dx$. Eventually the process will stop because a high enough derivative of f equals zero. This justifies the following procedure for finding antiderivatives in this case.

Procedure 6.4.5 Suppose $f^{(m)} = 0$ for some m a positive integer and let $G'_k = G_{k-1}$ for all k and $G_0 = g$. Then

$$\int f g dx = f G_1 - f' G_2 + f'' G_3 - f''' G_4 + \cdots$$

Just keep writing these terms, alternating signs until the process yields a zero. Then add on an arbitrary constant of integration and stop. Sometimes people remember this in the form of a table.

		g
f	$\xrightarrow{+}$	G_1
f'	$\xrightarrow{-}$	G_2
f''	$\xrightarrow{+}$	G_3
f'''	$\xrightarrow{-}$	G_4

Thus you fill in the table until the left column ends in a 0 and then do the arrows, $f G_1 - f' G_2 + f'' G_3 \cdots$ till the process ends. Then add C , a constant of integration.

Example 6.4.6 Find $\int x^5 \sin x dx$.

From the above procedure and letting $f(x) = x^5$, this equals

$$\begin{aligned} & x^5 (-\cos(x)) - 5x^4 (-\sin(x)) + 20x^3 (\cos(x)) - 60x^2 (\sin(x)) \\ & + 120x (-\cos(x)) - 120 (-\sin(x)) + C. \end{aligned}$$

6.5 Exercises

(e) $\int x^3 \cos(x^2) dx$

1. Find the following antiderivatives.

(a) $\int x^3 e^{-3x} dx$

(b) $\int x^4 \cos x dx$

(c) $\int x^5 e^x dx$

(d) $\int x^6 \sin(2x) dx$

(e) $\int x^3 \cos(x^2) dx$

2. Find the following antiderivatives.

(a) $\int x e^{-3x} dx$

(b) $\int \frac{1}{x(\ln(|x|))^2} dx$

(c) $\int x \sqrt{2-x} dx$

(d) $\int (\ln|x|)^2 dx$ **Hint:** Let $u(x) = (\ln|x|)^2$ and $v'(x) = 1$.

3. Show that $\int \sec^3(x) dx =$

$$\frac{1}{2} \tan(x) \sec(x)$$

$$+ \frac{1}{2} \ln |\sec x + \tan x| + C.$$

4. Find $\int \frac{x e^x}{(1+x)^2} dx$.

5. Consider the following argument. Integrate by parts, letting $u(x) = x$ and $v'(x) = \frac{1}{x^2}$ to get

$$\int \frac{1}{x} dx = \int x \left(\frac{1}{x^2} \right) dx$$

$$= \left(-\frac{1}{x} \right) x + \int \frac{1}{x} dx$$

$$= -1 + \int \frac{1}{x} dx.$$

Now subtracting $\int \frac{1}{x} dx$ from both sides, $0 = -1$. Is there anything wrong here? If so, what?

6. Find the following antiderivatives.

- (a) $\int x^3 \arctan(x) dx$
- (b) $\int x^3 \ln(x) dx$
- (c) $\int x^2 \sin(x) dx$
- (d) $\int x^2 \cos(x) dx$
- (e) $\int x \arcsin(x) dx$
- (f) $\int \cos(2x) \sin(3x) dx$
- (g) $\int x^3 e^{x^2} dx$
- (h) $\int x^3 \cos(x^2) dx$

7. Find the antiderivatives

- (a) $\int x^2 \sin x dx$
- (b) $\int x^3 \sin x dx$
- (c) $\int x^3 7^x dx$
- (d) $\int x^2 \ln x dx$
- (e) $\int (x+2)^2 e^x dx$
- (f) $\int x^3 2^x dx$
- (g) $\int \sec^3(2x) \tan(2x) dx$
- (h) $\int x^2 7^x dx$

8. Solve the initial value problem $y'(x) = f(x)$, $\lim_{x \rightarrow 0^+} y(x) = 1$ where $f(x)$ is each of the integrands in Problem 7.

9. Solve the initial value problem $y'(x) = f(x)$, $\lim_{x \rightarrow 0^+} y(x) = 2$ where $f(x)$ is each of the integrands in Problem 6.

10. Try doing $\int \sin^2 x dx$ the obvious way. If you don't make any mistakes, the process will go in circles. Now do it by taking

$$\begin{aligned} \int \sin^2 x dx &= x \sin^2 x \\ &\quad - 2 \int x \sin x \cos x dx \\ &= x \sin^2 x - \int x \sin(2x) dx. \end{aligned}$$

11. An object moves on the x axis having velocity equal to $t \sin t$. Find the position of the object given that at $t = 1$, it is at the point 2.

12. An object moves on the x axis having velocity equal to $\sec^3(t)$. Find the position of the object given that at $t = 0$, it is at the point 2. **Hint:** You might want to use Problem 3.

6.6 Trig. Substitutions

Certain antiderivatives are easily obtained by making an auspicious substitution involving a trig. function. The technique will be illustrated by presenting examples.

Example 6.6.1 Find $\int \frac{1}{(x^2+2x+2)^2} dx$.

Complete the square as before and write

$$\int \frac{1}{(x^2+2x+2)^2} dx = \int \frac{1}{((x+1)^2+1)^2} dx$$

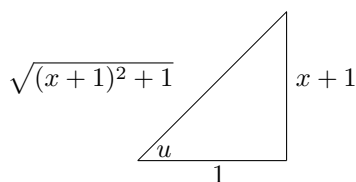
Use the following substitution next.

$$x+1 = \tan u \tag{6.5}$$

so $dx = (\sec^2 u) du$. Therefore, this last indefinite integral becomes

$$\begin{aligned} \int \frac{\sec^2 u}{(\tan^2 u + 1)^2} du &= \int (\cos^2 u) du \\ &= \int \frac{1 + \cos 2u}{2} du \\ &= \frac{u}{2} + \frac{\sin 2u}{4} + C \\ &= \frac{u}{2} + \frac{2 \sin u \cos u}{4} + C \end{aligned}$$

Next write this in terms of x using the following device based on the following picture.



In this picture which is descriptive of 6.5, $\sin u = \frac{x+1}{\sqrt{(x+1)^2 + 1}}$ and

$$\cos u = \frac{1}{\sqrt{(x+1)^2 + 1}}.$$

Therefore, putting in this information to change back to the x variable,

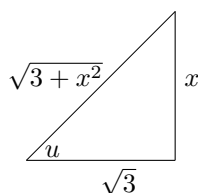
$$\begin{aligned} &\int \frac{1}{(x^2 + 2x + 2)^2} dx \\ &= \frac{1}{2} \arctan(x+1) + \frac{1}{2} \frac{x+1}{\sqrt{(x+1)^2 + 1}} \frac{1}{\sqrt{(x+1)^2 + 1}} + C \\ &= \frac{1}{2} \arctan(x+1) + \frac{1}{2} \frac{x+1}{(x+1)^2 + 1} + C. \end{aligned}$$

Example 6.6.2 Find $\int \frac{1}{\sqrt{x^2+3}} dx$.

Let $x = \sqrt{3} \tan u$ so $dx = \sqrt{3} (\sec^2 u) du$. Making the substitution, consider

$$\begin{aligned} &\int \frac{1}{\sqrt{3} \sqrt{\tan^2 u + 1}} \sqrt{3} (\sec^2 u) du \\ &= \int (\sec u) du = \ln |\sec u + \tan u| + C \end{aligned}$$

Now the following diagram is descriptive of the above transformation.



Using the above diagram, $\sec u = \frac{\sqrt{3+x^2}}{\sqrt{3}}$ and $\tan u = \frac{x}{\sqrt{3}}$. Therefore, restoring the x variable,

$$\begin{aligned}\int \frac{1}{\sqrt{x^2+3}} dx &= \ln \left| \frac{\sqrt{3+x^2}}{\sqrt{3}} + \frac{x}{\sqrt{3}} \right| + C \\ &= \ln \left| \sqrt{3+x^2} + x \right| + C.\end{aligned}$$

Note the constant C changed in going from the top to the bottom line. It is $C - \ln \sqrt{3}$ but it is customary to simply write this as C because C is arbitrary.

Example 6.6.3 Find $\int (4x^2 + 3)^{1/2} dx$.

Let $2x = \sqrt{3} \tan u$ so $2dx = \sqrt{3} \sec^2(u) du$. Then making the substitution,

$$\begin{aligned}& \sqrt{3} \int (\tan^2 u + 1)^{1/2} \frac{\sqrt{3}}{2} \sec^2(u) du \\ &= \frac{3}{2} \int \sec^3(u) du.\end{aligned}\tag{6.6}$$

Now use integration by parts to obtain

$$\begin{aligned}\int \sec^3(u) du &= \int \sec^2(u) \sec(u) du = \\ &= \tan(u) \sec(u) - \int \tan^2(u) \sec(u) du \\ &= \tan(u) \sec(u) - \int (\sec^2(u) - 1) \sec(u) du \\ &= \tan(u) \sec(u) + \int \sec(u) du - \int \sec^3(u) du \\ &= \tan(u) \sec(u) + \ln |\sec(u) + \tan(u)| - \int \sec^3(u) du\end{aligned}$$

Therefore,

$$2 \int \sec^3(u) du = \tan(u) \sec(u) + \ln |\sec(u) + \tan(u)| + C$$

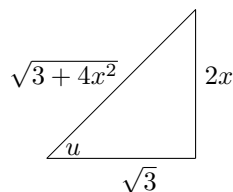
and so

$$\int \sec^3(u) du = \frac{1}{2} [\tan(u) \sec(u) + \ln |\sec(u) + \tan(u)|] + C.\tag{6.7}$$

Now it follows from 6.6 that in terms of u the set of antiderivatives is given by

$$\frac{3}{4} [\tan(u) \sec(u) + \ln |\sec(u) + \tan(u)|] + C$$

Use the following diagram to change back to the variable, x .



From the diagram, $\tan(u) = \frac{2x}{\sqrt{3}}$ and $\sec(u) = \frac{\sqrt{3+4x^2}}{\sqrt{3}}$. Therefore,

$$\begin{aligned} \int (4x^2 + 3)^{1/2} dx &= \\ &= \frac{3}{4} \left[\frac{2x}{\sqrt{3}} \frac{\sqrt{3+4x^2}}{\sqrt{3}} + \ln \left| \frac{\sqrt{3+4x^2}}{\sqrt{3}} + \frac{2x}{\sqrt{3}} \right| \right] + C \\ &= \frac{3}{4} \left[\frac{2x}{\sqrt{3}} \frac{\sqrt{3+4x^2}}{\sqrt{3}} + \ln \left| \frac{\sqrt{3+4x^2}}{\sqrt{3}} + \frac{2x}{\sqrt{3}} \right| \right] + C \\ &= \frac{1}{2} x \sqrt{3+4x^2} + \frac{3}{4} \ln \left| \sqrt{3+4x^2} + 2x \right| + C \end{aligned}$$

Note that these examples involved something of the form $(a^2 + (bx)^2)$ and the trig substitution,

$$bx = a \tan u$$

was the right one to use. This is the auspicious substitution which often simplifies these sorts of problems. However, there is a possibly better way to do these kinds.

Example 6.6.4 Find $\int (4x^2 + 3)^{1/2} dx$ another way.

Let $2x = \sqrt{3} \sinh u$ and so $2dx = \sqrt{3} \cosh(u) du$. Then substituting in the integral leads to

$$\begin{aligned} & \int \sqrt{3} \sqrt{1 + \sinh^2(u)} \frac{\sqrt{3}}{2} \cosh(u) du \\ &= \frac{3}{2} \int \cosh^2(u) du + C \\ &= \frac{3}{4} \cosh(u) \sinh(u) + \frac{3}{4} u + C \\ &= \frac{3}{4} \sqrt{1 + \sinh^2(u)} \sinh(u) + \frac{3}{4} u + C \\ &= \frac{1}{2} x \sqrt{3 + 4x^2} + \frac{3}{4} \sinh^{-1} \left(\frac{2x}{\sqrt{3}} \right) + C \end{aligned}$$

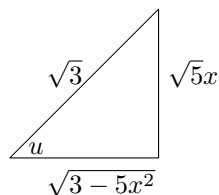
This other way is often used by computer algebra systems. If you solve for $\sinh^{-1} x$ in terms of $\ln$, you get the same set of antiderivatives. The function $\sinh^{-1}$ is also written as $\operatorname{arcsinh}$ by analogy to the trig. functions.

Example 6.6.5 Find $\int \sqrt{3 - 5x^2} dx$

In this example, let $\sqrt{5}x = \sqrt{3} \sin(u)$ so $\sqrt{5}dx = \sqrt{3} \cos(u) du$. The reason this might be a good idea is that it will get rid of the square root sign as shown below. Making the substitution,

$$\begin{aligned} \int \sqrt{3 - 5x^2} dx &= \sqrt{3} \int \sqrt{1 - \sin^2(u)} \frac{\sqrt{3}}{\sqrt{5}} \cos(u) du \\ &= \frac{3}{\sqrt{5}} \int \cos^2(u) du \\ &= \frac{3}{\sqrt{5}} \int \frac{1 + \cos 2u}{2} du \\ &= \frac{3}{\sqrt{5}} \left(\frac{u}{2} + \frac{\sin 2u}{4} \right) + C \\ &= \frac{3}{2\sqrt{5}} u + \frac{3}{2\sqrt{5}} \sin u \cos u + C \end{aligned}$$

The appropriate diagram is the following.



From the diagram, $\sin(u) = \frac{\sqrt{5}x}{\sqrt{3}}$ and $\cos(u) = \frac{\sqrt{3-5x^2}}{\sqrt{3}}$. Therefore, changing back to x ,

$$\begin{aligned} \int \sqrt{3-5x^2} dx &= \\ &= \frac{3}{2\sqrt{5}} \arcsin\left(\frac{\sqrt{5}x}{\sqrt{3}}\right) + \frac{3}{2\sqrt{5}} \frac{\sqrt{5}x}{\sqrt{3}} \frac{\sqrt{3-5x^2}}{\sqrt{3}} + C \\ &= \frac{3}{10} \sqrt{5} \arcsin\left(\frac{1}{3} \sqrt{15}x\right) + \frac{1}{2} x \sqrt{3-5x^2} + C \end{aligned}$$

Example 6.6.6 Find $\int \sqrt{5x^2 - 3} dx$

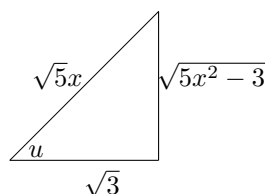
In this example, let $\sqrt{5}x = \sqrt{3} \sec(u)$ so $\sqrt{5}dx = \sqrt{3} \sec(u) \tan(u) du$. Then changing the variable, consider

$$\begin{aligned} & \sqrt{3} \int \sqrt{\sec^2(u) - 1} \frac{\sqrt{3}}{\sqrt{5}} \sec(u) \tan(u) du \\ &= \frac{3}{\sqrt{5}} \int \tan^2(u) \sec(u) du \\ &= \frac{3}{\sqrt{5}} \left[\int \sec^3(u) du - \int \sec(u) du \right]. \end{aligned}$$

Now from 6.7, this equals

$$\begin{aligned} & \frac{3}{\sqrt{5}} \left[\frac{1}{2} [\tan(u) \sec(u) + \ln |\sec(u) + \tan(u)|] - \ln |\tan(u) + \sec(u)| \right] + C \\ &= \frac{3}{2\sqrt{5}} \tan(u) \sec(u) - \frac{3}{2\sqrt{5}} \ln |\sec(u) + \tan(u)| + C. \end{aligned}$$

Now it is necessary to change back to x . The diagram is as follows.



Therefore, $\tan(u) = \frac{\sqrt{5x^2-3}}{\sqrt{3}}$ and $\sec(u) = \frac{\sqrt{5}x}{\sqrt{3}}$ and so

$$\begin{aligned} \int \sqrt{5x^2 - 3} dx &= \\ &= \frac{3}{2\sqrt{5}} \frac{\sqrt{5x^2-3}}{\sqrt{3}} \frac{\sqrt{5}x}{\sqrt{3}} - \frac{3}{2\sqrt{5}} \ln \left| \frac{\sqrt{5}x}{\sqrt{3}} + \frac{\sqrt{5x^2-3}}{\sqrt{3}} \right| + C \\ &= \frac{1}{2} (\sqrt{5x^2-3}) x - \frac{3}{10} \sqrt{5} \ln \left| \sqrt{5}x + \sqrt{(-3+5x^2)} \right| + C \end{aligned}$$

To summarize, here is a short table of auspicious substitutions corresponding to certain expressions.

Table Of Auspicious Substitutions

Expression	$a^2 + b^2x^2$	$a^2 - b^2x^2$	$a^2x^2 - b^2$
Trig. substitution	$bx = a \tan(u)$	$bx = a \sin(u)$	$ax = b \sec(u)$
Hyperbolic substitution	$bx = a \sinh(u)$		

Of course there are no “magic bullets” but these substitutions will often simplify an expression enough to allow you to find an antiderivative. These substitutions are often especially useful when the expression is enclosed in a square root.

6.7 Exercises

1. Find the antiderivatives.

- $\int \frac{x}{\sqrt{4-x^2}} dx$
- $\int \frac{3}{\sqrt{36-25x^2}} dx$
- $\int \frac{3}{\sqrt{16-25x^2}} dx$
- $\int \frac{1}{\sqrt{4-9x^2}} dx$
- $\int \frac{1}{\sqrt{36-x^2}} dx$
- $\int (\sqrt{9-16x^2})^3 dx$
- $\int (\sqrt{16-x^2})^5 dx$
- $\int \sqrt{25-36x^2} dx$
- $\int (\sqrt{4-9x^2})^3 dx$
- $\int \sqrt{1-9x^2} dx$

2. Find the antiderivatives.

- $\int \sqrt{36x^2-25} dx$
- $\int \sqrt{x^2-4} dx$
- $\int (\sqrt{16x^2-9})^3 dx$
- $\int \sqrt{25x^2-16} dx$

3. Find the antiderivatives.

- $\int \frac{1}{26+x^2-2x} dx$ **Hint:** Complete the square.

- $\int \sqrt{x^2+9} dx$
- $\int \sqrt{4x^2+25} dx$
- $\int x\sqrt{4x^4+9} dx$
- $\int x^3\sqrt{4x^4+9} dx$
- $\int \frac{1}{(16+25(x-3)^2)^2} dx$
- $\int \frac{1}{261+25x^2-150x} dx$ **Hint:** Complete the square.
- $\int (\sqrt{25x^2+9})^3 dx$
- $\int \frac{1}{25+16x^2} dx$

4. Find the antiderivatives. **Hint:** Complete the square.

- $\int \sqrt{4x^2+16x+15} dx$
- $\int \sqrt{x^2+6x} dx$
- $\int \frac{3}{\sqrt{-32-9x^2-36x}} dx$
- $\int \frac{3}{\sqrt{-5-x^2-6x}} dx$
- $\int \frac{1}{\sqrt{9-16x^2-32x}} dx$
- $\int \sqrt{4x^2+16x+7} dx$

6.8 Partial Fractions

The main technique for finding antiderivatives in the case $f(x) = \frac{p(x)}{q(x)}$ for p and q polynomials is the technique of partial fractions. Before presenting this technique, a few more examples are presented.

Example 6.8.1 Find $\int \frac{1}{x^2+2x+2} dx$.

To do this, complete the square in the denominator to write

$$\int \frac{1}{x^2 + 2x + 2} dx = \int \frac{1}{(x+1)^2 + 1} dx$$

Now change the variable, letting $u = x + 1$, so that $du = dx$. Then the last indefinite integral reduces to

$$\int \frac{1}{u^2 + 1} du = \arctan u + C$$

and so

$$\int \frac{1}{x^2 + 2x + 2} dx = \arctan(x+1) + C.$$

Example 6.8.2 Find $\int \frac{1}{3x+5} dx$.

Let $u = 3x + 5$ so $du = 3dx$ and changing the variable,

$$\frac{1}{3} \int \frac{1}{u} du = \frac{1}{3} \ln|u| + C.$$

Therefore,

$$\int \frac{1}{3x+5} dx = \frac{1}{3} \ln|3x+5| + C.$$

Example 6.8.3 Find $\int \frac{3x+2}{x^2+x+1} dx$.

First complete the square in the denominator.

$$\begin{aligned} \int \frac{3x+2}{x^2+x+1} dx &= \int \frac{3x+2}{x^2+x+\frac{1}{4}+\frac{3}{4}} dx \\ &= \int \frac{3x+2}{\left(x+\frac{1}{2}\right)^2+\frac{3}{4}} dx. \end{aligned}$$

Now let

$$\left(x + \frac{1}{2}\right)^2 = \frac{3}{4}u^2$$

so that $x + \frac{1}{2} = \frac{\sqrt{3}}{2}u$. Therefore, $dx = \frac{\sqrt{3}}{2}du$ and changing the variable,

$$\begin{aligned} &\frac{4}{3} \int \frac{3\left(\frac{\sqrt{3}}{2}u - \frac{1}{2}\right) + 2}{u^2 + 1} \frac{\sqrt{3}}{2} du \\ &= \frac{\sqrt{3}}{2} \left(2\sqrt{3} \int \frac{u}{u^2 + 1} du - \frac{2}{3} \int \frac{1}{u^2 + 1} du \right) \\ &= \frac{\sqrt{3}}{2} \left(\sqrt{3} \int \frac{2u}{u^2 + 1} du - \frac{2}{3} \int \frac{1}{u^2 + 1} du \right) \\ &= \frac{3}{2} \ln(u^2 + 1) - \frac{\sqrt{3}}{3} \arctan u + C \end{aligned}$$

Therefore,

$$\begin{aligned} &\int \frac{3x+2}{x^2+x+1} dx = \\ &\frac{3}{2} \ln \left(\left(\frac{2}{\sqrt{3}} \left(x + \frac{1}{2} \right) \right)^2 + 1 \right) - \frac{\sqrt{3}}{3} \arctan \left(\frac{2}{\sqrt{3}} \left(x + \frac{1}{2} \right) \right) + C. \end{aligned}$$

The method of partial fractions splits rational functions into a sum of functions which are like those which were just done successfully. In using this method it is essential that in the rational function the degree of the numerator is smaller than the degree of the denominator. The following simple but important Lemma shows that you can always reduce to this case.

Lemma 6.8.4 *Let $f(x)$ and $g(x)$ be polynomials. Then there exists a polynomial, $q(x)$ such that*

$$f(x) = q(x)g(x) + r(x)$$

where the degree of $r(x)$ is less than the degree of $g(x)$ or $r(x) = 0$.

Proof: Consider the polynomials of the form $f(x) - g(x)l(x)$ and out of all these polynomials, pick one which has the smallest degree. This can be done because of the well ordering of the natural numbers. Let this take place when $l(x) = q_1(x)$ and let

$$r(x) = f(x) - g(x)q_1(x).$$

It is required to show degree of $r(x) < \text{degree of } g(x)$ or else $r(x) = 0$.

Suppose $f(x) - g(x)l(x)$ is never equal to zero for any $l(x)$. Then $r(x) \neq 0$. It is required to show the degree of $r(x)$ is smaller than the degree of $g(x)$. If this doesn't happen, then the degree of $r \geq$ the degree of g . Let

$$\begin{aligned} r(x) &= b_mx^m + \cdots + b_1x + b_0 \\ g(x) &= a_nx^n + \cdots + a_1x + a_0 \end{aligned}$$

where $m \geq n$ and b_m and a_n are nonzero. Then let $r_1(x)$ be given by

$$\begin{aligned} r_1(x) &= r(x) - \frac{x^{m-n}b_m}{a_n}g(x) \\ &= (b_mx^m + \cdots + b_1x + b_0) - \frac{x^{m-n}b_m}{a_n}(a_nx^n + \cdots + a_1x + a_0) \end{aligned}$$

which has smaller degree than m , the degree of $r(x)$. But

$$\begin{aligned} r_1(x) &= \overbrace{f(x) - g(x)q_1(x)}^{r(x)} - \frac{x^{m-n}b_m}{a_n}g(x) \\ &= f(x) - g(x)\left(q_1(x) + \frac{x^{m-n}b_m}{a_n}\right), \end{aligned}$$

and this is not zero by the assumption that $f(x) - g(x)l(x)$ is never equal to zero for any $l(x)$ yet has smaller degree than $r(x)$ which is a contradiction to the choice of $r(x)$. This proves the lemma.

Corollary 6.8.5 *Let $f(x)$ and $g(x)$ be polynomials. Then there exists a polynomial, $r(x)$ such that the degree of $r(x) < \text{degree of } g(x)$ and a polynomial, $q(x)$ such that*

$$\frac{f(x)}{g(x)} = q(x) + \frac{r(x)}{g(x)}.$$

Here is an example where the degree of the numerator exceeds the degree of the denominator.

Example 6.8.6 Find $\int \frac{3x^5+7}{x^2-1} dx$.

In this case the degree of the numerator is larger than the degree of the denominator and so long division must first be used. Thus

$$\frac{3x^5 + 7}{x^2 - 1} = 3x^3 + 3x + \frac{7 + 3x}{x^2 - 1}$$

Now look for a partial fractions expansion of the form

$$\frac{7 + 3x}{x^2 - 1} = \frac{a}{(x - 1)} + \frac{b}{(x + 1)}.$$

Therefore,

$$7 + 3x = a(x + 1) + b(x - 1).$$

Letting $x = 1$, $a = 5$. Then letting $x = -1$, it follows $b = -2$. Therefore,

$$\frac{7 + 3x}{x^2 - 1} = \frac{5}{x - 1} - \frac{2}{x + 1}$$

and so

$$\frac{3x^5 + 7}{x^2 - 1} = 3x^3 + 3x + \frac{5}{x - 1} - \frac{2}{x + 1}.$$

therefore,

$$\int \frac{3x^5 + 7}{x^2 - 1} dx = \frac{3}{4}x^4 + \frac{3}{2}x^2 + 5 \ln(x - 1) - 2 \ln(x + 1) + C.$$

Here is another example.

Example 6.8.7 Find $\int \frac{-x^3 + 11x^2 + 24x + 14}{(2x + 3)(x + 5)(x^2 + x + 1)} dx$.

In this problem, first check to see if the degree of the numerator in the integrand is less than the degree of the denominator. In this case, this is so. If it is not so, use long division to write the integrand as the sum of a polynomial with a rational function in which the degree of the numerator is less than the degree of the denominator. See the preceding corollary which guarantees this can be done. Now look for a partial fractions expansion for the integrand which is in the following form.

$$\frac{a}{2x + 3} + \frac{b}{x + 5} + \frac{cx + d}{x^2 + x + 1}$$

and try to find constants, a, b, c , and d so that the above rational functions sum to the integrand. The reason $cx + d$ is used in the numerator of the last expression is that $x^2 + x + 1$ cannot be factored using real polynomials. Thus the problem involves finding a, b, c, d , such that

$$\frac{-x^3 + 11x^2 + 24x + 14}{(2x + 3)(x + 5)(x^2 + x + 1)} = \frac{a}{2x + 3} + \frac{b}{x + 5} + \frac{cx + d}{x^2 + x + 1}$$

and so

$$\begin{aligned} -x^3 + 11x^2 + 24x + 14 &= a(x + 5)(x^2 + x + 1) + \\ &+ b(2x + 3)(x^2 + x + 1) + (cx + d)(2x + 3)(x + 5). \end{aligned} \quad (6.8)$$

Now these are two polynomials which are supposed to be equal. Therefore, they have the same coefficients. Multiplying the right side out and collecting the terms,

$$-x^3 + 11x^2 + 24x + 14 =$$

$$(2b + 2c + a)x^3 + (6a + 5b + 13c + 2d)x^2 + (6a + 13d + 5b + 15c)x + 15d + 5a + 3b$$

and therefore, it is necessary to solve the equations,

$$\begin{aligned} 2b + 2c + a &= -1 \\ 6a + 5b + 13c + 2d &= 11 \\ 6a + 13d + 5b + 15c &= 24 \\ 15d + 5a + 3b &= 14 \end{aligned}$$

The solution is $c = 1, a = 1, b = -2, d = 1$. Therefore,

$$\frac{-x^3 + 11x^2 + 24x + 14}{(2x + 3)(x + 5)(x^2 + x + 1)} = \frac{1}{2x + 3} - \frac{2}{x + 5} + \frac{1 + x}{x^2 + x + 1}.$$

This may look like a fairly formidable problem. In reality it is not that bad. First let $x = -5$ in 6.8 and obtain a simple equation for finding b . Next let $x = -3/2$ to get a simple equation for a . This reduces the above system to a more manageable size. Anyway, it is now possible to find an antiderivative of the given function.

$$\begin{aligned} \int \frac{-x^3 + 11x^2 + 24x + 14}{(2x + 3)(x + 5)(x^2 + x + 1)} dx &= \\ \int \frac{1}{2x + 3} dx - \int \frac{2}{x + 5} dx + \int \frac{1 + x}{x^2 + x + 1} dx. \end{aligned}$$

Each of these indefinite integrals can be found using the techniques given above. Thus the antiderivative is

$$\begin{aligned} \frac{1}{2} \ln |2x + 3| - 2 \ln |x + 5| + \\ \frac{1}{2} \ln (x^2 + x + 1) + \frac{1}{3} \sqrt{3} \arctan \left(\frac{\sqrt{3}}{3} (2x + 1) \right) + C. \end{aligned}$$

What is done when the factors are repeated?

Example 6.8.8 Find $\int \frac{3x+7}{(x+2)^2(x+3)} dx$.

First observe that the degree of the numerator is less than the degree of the denominator. In this case the correct form of the partial fraction expansion is

$$\frac{a}{(x + 2)} + \frac{b}{(x + 2)^2} + \frac{c}{(x + 3)}.$$

The reason there are two terms devoted to $(x + 2)$ is that this is squared. Computing the constants yields

$$\frac{3x + 7}{(x + 2)^2(x + 3)} = \frac{1}{(x + 2)^2} + \frac{2}{x + 2} - \frac{2}{x + 3}$$

and therefore,

$$\int \frac{3x + 7}{(x + 2)^2(x + 3)} dx = -\frac{1}{x + 2} + 2 \ln (x + 2) - 2 \ln (x + 3) + C.$$

Example 6.8.9 Find the proper form for the partial fractions expansion of

$$\frac{x^3 + 7x + 9}{(x^2 + 2x + 2)^3 (x + 2)^2 (x + 1) (x^2 + 1)}.$$

First check to see if the degree of the numerator is smaller than the degree of the denominator. Since this is the case, look for a partial fractions decomposition in the following form.

$$\frac{ax+b}{(x^2+2x+2)} + \frac{cx+d}{(x^2+2x+2)^2} + \frac{ex+f}{(x^2+2x+2)^3} +$$

$$\frac{A}{(x+2)} + \frac{B}{(x+2)^2} + \frac{D}{(x+1)} + \frac{gx+h}{x^2+1}.$$

These examples illustrate what to do when using the method of partial fractions. You first check to be sure the degree of the numerator is less than the degree of the denominator. If this is not so, do a long division. Then you factor the denominator into a product of factors, some linear of the form $ax+b$ and others quadratic, ax^2+bx+c which cannot be factored further. Next follow the procedure illustrated in the above examples and summarized below.

Warning: When you use partial fractions, **be sure you look for something which is of the right form.** Otherwise you may be looking for something which is not there. The rules are summarized next.

Rules For Finding Partial Fractions Expansion Of A Rational Function

1. Check to see if the numerator has smaller degree than the denominator. If this is not so, correct the situation by doing long division.
2. Factor the denominator into a product of linear factors, (Things like $(ax+b)$) and irreducible quadratic factors, (Things like (ax^2+bx+c) where $b^2-4ac < 0$.)¹
3. Let m, n be positive integers. Corresponding to $(ax+b)^m$ in the denominator, you should have a sum of the form $\sum_{i=1}^m \frac{c_i}{(ax+b)^i}$ in the partial fractions expansion. Here the c_i are the constants to be found. Corresponding to $(ax^2+bx+c)^n$ in the denominator where $b^2-4ac < 0$, you should have a sum of the form $\sum_{i=1}^n \frac{p_i x + q_i}{(ax^2+bx+c)^i}$ in the partial fractions expansion. Here the p_i and q_i are to be found.
4. Find the constants, c_i , p_i , and q_i . Use whatever method you like. You might see if you can make up new ways to do this if you like. If you have followed steps 1 - 3 correctly, it will work out.

The above technique for finding the coefficients is fine but some people like to do it other ways. It really doesn't matter how you do it. Here is another example.

Example 6.8.10 Find the partial fractions expansion for $\frac{15x^4+44x^3+71x^2+64x+28+2x^5}{(x+2)^2(x^2+2x+2)^2}$.

The degree of the top is 4 and the degree of the bottom is 6 so you don't need to do long division. You do have to look for the right thing however. The correct form for the partial fractions expansion is

$$\frac{a}{x+2} + \frac{b}{(x+2)^2} + \frac{cx+d}{x^2+2x+2} + \frac{ex+f}{(x^2+2x+2)^2}$$

$$= \frac{15x^4 + 44x^3 + 71x^2 + 64x + 28 + 2x^5}{(x+2)^2(x^2+2x+2)^2}$$

¹Of course this factoring of the denominator is easier said than done. In general you can't do it at all. Of course there are big theorems which guarantee the existence of such a factorization but these theorems do not tell how to find it. This is an example of the gap between theory and practice which permeates mathematics.

Multiply by the denominator of the right side. This yields

$$\begin{aligned} & 15x^4 + 44x^3 + 71x^2 + 64x + 28 + 2x^5 \\ = & a(x+2)(x^2+2x+2)^2 + b(x^2+2x+2)^2 \\ & + (cx+d)(x+2)^2(x^2+2x+2) + (ex+f)(x+2)^2 \end{aligned}$$

You could multiply out the right side and match the coefficients but you could also do other things. First set $x = -2$ because this will cause many of the terms on the right to equal zero. The left side equals

$$15(-2)^4 + 44(-2)^3 + 71(-2)^2 + 64(-2) + 28 + 2(-2)^5 = 8.$$

the right side equals

$$b\left((-2)^2 + 2(-2) + 2\right)^2 = 4b$$

and so $b = 2$. Progress has been made. Now subtract the term $2(x^2 + 2x + 2)^2$ from both sides and then divide both sides by $(x + 2)$. First the subtraction:

$$\begin{aligned} & 15x^4 + 44x^3 + 71x^2 + 64x + 28 + 2x^5 - 2(x^2 + 2x + 2)^2 \\ = & a(x+2)(x^2+2x+2)^2 + (cx+d)(x+2)^2(x^2+2x+2) + (ex+f)(x+2)^2. \end{aligned}$$

Simplifying this yields

$$\begin{aligned} & 13x^4 + 36x^3 + 55x^2 + 48x + 20 + 2x^5 \\ = & a(x+2)(x^2+2x+2)^2 + (cx+d)(x+2)^2(x^2+2x+2) + (ex+f)(x+2)^2. \end{aligned}$$

Now $(x + 2)$ divides the right side so it must also divide the left. Do the division by whatever method you like best.

$$\begin{aligned} & 2x^4 + 9x^3 + 18x^2 + 19x + 10 \\ = & a(x^2 + 2x + 2)^2 + (cx + d)(x + 2)(x^2 + 2x + 2) + (ex + f)(x + 2). \end{aligned} \quad (6.9)$$

Now let $x = -2$ again. This yields

$$2(-2)^4 + 9(-2)^3 + 18(-2)^2 + 19(-2) + 10 = a\left((-2)^2 + 2(-2) + 2\right)^2$$

and so

$$4 = a(4)$$

which shows $a = 1$. Now use this information in 6.9. Thus

$$\begin{aligned} & 2x^4 + 9x^3 + 18x^2 + 19x + 10 - (x^2 + 2x + 2)^2 \\ = & (cx + d)(x + 2)(x^2 + 2x + 2) + (ex + f)(x + 2). \end{aligned}$$

It looks like you can divide both sides by $(x + 2)$ again. Lets do so.

$$x^3 + 3x^2 + 4x + 3 = (cx + d)(x^2 + 2x + 2) + (ex + f).$$

Now finally, lets expand both sides.

$$x^3 + 3x^2 + 4x + 3 = cx^3 + (2c + d)x^2 + (2d + 2c + e)x + 2d + f. \quad (6.10)$$

At this point, you could match the coefficients and get a system of linear equations or you could be more clever. Lets be clever. Differentiate both sides. Then

$$3x^2 + 6x + 4 = 3cx^2 + 4cx + 2dx + 2d + 2c + e. \quad (6.11)$$

Now let's differentiate both sides again.

$$6x + 6 = 6cx + 4c + 2d. \quad (6.12)$$

Differentiate yet again.

$$6 = 6c.$$

Thus $c = 1$. Now let $x = 0$ in 6.12 to get

$$6 = 4 + 2d$$

so $d = 1$. Now let $x = 0$ in 6.11 to get

$$4 = 2 + 2 + e$$

showing $e = 0$. Now it only remains to find f . Let $x = 0$ in 6.10 and obtain

$$3 = 2 + f$$

so $f = 1$. It follows the partial fractions expansion is

$$\frac{1}{x+2} + \frac{2}{(x+2)^2} + \frac{x+1}{x^2+2x+2} + \frac{1}{(x^2+2x+2)^2}.$$

One other thing should be mentioned. Suppose you wanted to find the integral in this example. The first three terms are by now routine. How about the last one?

Example 6.8.11 Find $\int \frac{1}{(x^2+2x+2)^2} dx$

First complete the square to write this as $\int \frac{1}{((x+1)^2+1)^2} dx$. Now do a trig. substitution. You should let $x+1 = \tan \theta$. Then the integral becomes

$$\begin{aligned} \int \frac{1}{\sec^4 \theta} \sec^2 \theta d\theta &= \int \cos^2 \theta d\theta = \int \frac{1 + \cos(2\theta)}{2} d\theta \\ &= \frac{\theta}{2} + \frac{2 \sin \theta \cos \theta}{4} + C. \end{aligned}$$

Setting up a little triangle as in the section on trig. substitutions, you can restore the original variables to obtain

$$\frac{1}{2} \arctan(x+1) + \frac{1}{2} \left(\frac{x+1}{x^2+2x+2} \right) + C.$$

6.9 Rational Functions Of Trig. Functions

There is a technique which reduces certain kinds of integrals involving trig. functions to the technique of partial fractions. This is illustrated in the following example.

Example 6.9.1 Find $\int \frac{\cos \theta}{1+\cos \theta} d\theta$.

The integrand is an example of a rational function of cosines and sines. When such a thing occurs there is a substitution which will reduce the integrand to a rational function like those above which can then be integrated using partial fractions. The substitution is $u = \tan\left(\frac{\theta}{2}\right)$. Thus in this example, $du = \left(1 + \tan^2\left(\frac{\theta}{2}\right)\right) \frac{1}{2} d\theta$ and so in terms of this new variable, the indefinite integral is

$$\int \frac{2 \cos(2 \arctan u)}{(1 + \cos(2 \arctan u))(1 + u^2)} du.$$

You can evaluate $\cos(2 \arctan u)$ exactly. This equals $2 \cos^2(\arctan u) - 1$. Setting up a little triangle as above, $\cos(\arctan u)$ equals $1/\sqrt{1+u^2}$ and so the integrand reduces to

$$\frac{2 \left(2 \left(1/\sqrt{1+u^2} \right)^2 - 1 \right)}{\left(1 + \left(2 \left(1/\sqrt{1+u^2} \right)^2 - 1 \right) \right) (1+u^2)} = \frac{1-u^2}{1+u^2} = -1 + \frac{2}{1+u^2}$$

therefore, in terms of u the antiderivative equals $-u + 2 \arctan u$. Now replace u to obtain

$$-\tan\left(\frac{\theta}{2}\right) + 2 \arctan\left(\tan\left(\frac{\theta}{2}\right)\right) + C.$$

This procedure can be expected to work in general. Suppose you want to find

$$\int \frac{p(\cos \theta, \sin \theta)}{q(\cos \theta, \sin \theta)} d\theta$$

where p and q are polynomials in each argument. Make the substitution $u = \tan \frac{\theta}{2}$. As above this means

$$du = \left(1 + \tan^2 \left(\frac{\theta}{2} \right) \right) \frac{1}{2} d\theta = \frac{1}{2} (1+u^2) d\theta.$$

It remains to substitute for $\sin \theta$ and $\cos \theta$. Recall that $\sin\left(\frac{\theta}{2}\right) = \pm \sqrt{\frac{1-\cos \theta}{2}}$ and $\cos\left(\frac{\theta}{2}\right) = \pm \sqrt{\frac{1+\cos \theta}{2}}$. Thus,

$$\tan\left(\frac{\theta}{2}\right) = \frac{\pm \sqrt{1-\cos \theta}}{\sqrt{1+\cos \theta}}$$

and so

$$u^2 = \tan^2\left(\frac{\theta}{2}\right) = \frac{1-\cos \theta}{1+\cos \theta}$$

and solving this for $\cos \theta$ and $\sin \theta$ yields

$$\cos \theta = \frac{1-u^2}{1+u^2}, \quad \sin \theta = \pm \frac{2u}{1+u^2}.$$

It follows that in terms of u the integral becomes

$$\int \frac{p\left(\frac{1-u^2}{1+u^2}, \pm \frac{2u}{1+u^2}\right)}{q\left(\frac{1-u^2}{1+u^2}, \pm \frac{2u}{1+u^2}\right)} \frac{2du}{1+u^2}$$

which is a rational function of u and so in theory, you might be able to find the integral from the method of partial fractions.

6.10 Exercises

of the following rational functions.

1. Give a condition on a, b , and c such that ax^2+bx+c cannot be factored as a product of two polynomials which have real coefficients.
2. Find the partial fractions expansion

- (a) $\frac{2x+7}{(x+1)^2(x+2)}$
- (b) $\frac{5x+1}{(x^2+1)(2x+3)}$
- (c) $\frac{5x+1}{(x^2+1)^2(2x+3)}$
- (d) $\frac{5x^4+10x^2+3+4x^3+6x}{(x+1)(x^2+1)^2}$

3. Find the antiderivatives

(a) $\int \frac{x^5 + 4x^4 + 5x^3 + 2x^2 + 2x + 7}{(x+1)^2(x+2)} dx$

(b) $\int \frac{5x+1}{(x^2+1)(2x+3)} dx$

(c) $\int \frac{5x+1}{(x^2+1)^2(2x+3)} dx$

4. Each of $\cot \theta$, $\tan \theta$, $\sec \theta$, and $\csc \theta$ is a rational function of $\cos \theta$ and $\sin \theta$. Use the technique of substituting $u = \tan(\frac{\theta}{2})$ to find antiderivatives for each of these.

5. Find $\int \frac{\sin \theta}{1 + \sin \theta} d\theta$. **Hint:** Use the above procedure of letting $u = \tan(\frac{\theta}{2})$ and then multiply both the top and the bottom by $(1 - \sin \theta)$ to see another way of doing it.

6. Find $\int \frac{\cos \theta + 1}{\cos \theta + 2} d\theta$ using the substitution $u = \tan(\frac{\theta}{2})$.

7. In finding $\int \sec(x) dx$, try the substitution $u = \sin(x)$.

8. In finding $\int \csc(x) dx$ try the substitution $u = \cos(x)$.

9. Find the antiderivatives.

(a) $\int \frac{17x-3}{(6x+1)(x-1)} dx$

(b) $\int \frac{50x^4 - 95x^3 - 20x^2 - 3x + 7}{(5x+3)(x-2)(2x-1)} dx$

Hint: Notice the degree of the numerator is larger than the degree of the denominator.

(c) $\int \frac{8x^2 + x - 5}{(3x+1)(x-1)(2x-1)} dx$

(d) $\int \frac{3x+2}{(5x+3)(x+1)} dx$

10. Find the antiderivatives

(a) $\int \frac{52x^2 + 68x + 46 + 15x^3}{(x+1)^2(5x^2 + 10x + 8)} dx$

(b) $\int \frac{9x^2 - 42x + 38}{(3x+2)(3x^2 - 12x + 14)} dx$

(c) $\int \frac{9x^2 - 6x + 19}{(3x+1)(3x^2 - 6x + 5)} dx$

11. Solve the initial value problem, $y' = f(x)$, $y(0) = 1$ for $f(x)$ equal to each of the integrands in Problem 10.

12. *Find the antiderivatives. You will need to complete the square and then make a trig. substitution. The answers are given.

(a) $\int \frac{1}{(3x^2 + 12x + 13)^2} dx =$

$$\frac{1}{12} \frac{6x + 12}{3x^2 + 12x + 13} +$$

$$\frac{1}{6} \sqrt{3} \arctan\left(\frac{1}{6} (6x + 12) \sqrt{3}\right) + C$$

(b) $\int \frac{1}{(5x^2 + 10x + 7)^2} dx =$

$$\frac{1}{40} \frac{10x + 10}{5x^2 + 10x + 7} +$$

$$\frac{1}{40} \sqrt{10} \arctan \frac{1}{20} \cdot$$

$$(10x + 10) \sqrt{10} + C$$

(c) $\int \frac{1}{(5x^2 - 20x + 23)^2} dx =$

$$\frac{1}{60} \frac{-20 + 10x}{5x^2 - 20x + 23}$$

$$+ \frac{1}{90} \sqrt{15} \cdot$$

$$\arctan\left(\frac{1}{30} (-20 + 10x) \cdot\right.$$

$$\left. \sqrt{15}\right) + C$$

13. Solve the initial value problem, $y' = f(x)$, $y(0) = 1$ for $f(x)$ equal to each of the integrands in Problem 12.

6.11 Practice Problems For Antiderivatives

The process of finding antiderivatives has absolutely nothing to do with mathematics. However, it is fun and it is good to become adept at doing it. This can only be accomplished through working lots of problems. Here are lots of practice problems for finding antiderivatives. Some of these are very hard but you don't have to do all of them if you don't want to. However, the more you do, the better you will be at taking antiderivatives. Most of these problems are modifications of problems I found in a Russian calculus book. This book had some which were even harder. I shall give answers

to these problems so you can see whether you have it right. Beware that sometimes you may get it right even though it looks different than the answer given. Also, there is no guarantee that my answers are right.

1. Find $\int \frac{\sqrt{2x+1}}{x} dx$.

Answer:

$$\int \frac{\sqrt{2x+1}}{x} dx = 2\sqrt{2}\sqrt{x} + \ln x + C$$

2. Find $\int \frac{te^t}{(t+1)^2} dt$ **Hint:** Write this as

$$\int \left(\frac{(1+t)e^t}{(t+1)^2} - \frac{e^t}{(1+t)^2} \right) dt$$

$$= \int \left(\frac{e^t}{(t+1)} - \frac{e^t}{(1+t)^2} \right) dt.$$

Answer: $\frac{e^t}{t+1} + C$

3. Find $\int \frac{5-2x^2}{5+2x^2} dx$. **Hint:** $\frac{5-2x^2}{5+2x^2} = -1 + \frac{10}{5+2x^2}$.

Answer:

$$\int \frac{5-2x^2}{5+2x^2} dx = -x + \sqrt{10} \arctan \frac{1}{5}x\sqrt{10} + C$$

4. Find $\int (3-x^5)^2 dx$.

Answer:

$$\int (3-x^5)^2 dx = \frac{1}{11}x^{11} - x^6 + 9x + C$$

5. Find $\int (2x+3)^{-25} dx$. **Hint:** Let $u = 2x+3$

Answer:

$$\int (2x+3)^{-25} dx = -\frac{1}{48(2x+3)^{24}} + C$$

6. Find $\int 5^x dx$.

Answer:

$$\int 5^x dx = \frac{1}{\ln 5} 5^x + C$$

7. Find $\int \cosh^2 8x dx$. **Hint:** Try integration by parts.

Answer:

$$\int \cosh^2 8x dx = \frac{1}{16} \cosh 8x \sinh 8x + \frac{1}{2}x + C$$

8. Find $\int \tanh^2 2x dx$.

Answer:

$$\int \tanh^2 2x dx =$$

$$-\frac{1}{2} \tanh 2x - \frac{1}{4} \ln(-1 + \tanh 2x)$$

$$+ \frac{1}{4} \ln(1 + \tanh 2x) + C$$

9. Find $\int \cosh(3x+3) dx$.

Answer:

$$\int \cosh(3x+3) dx = \frac{1}{3} \sinh(3x+3) + C$$

10. Find $\int 8^{1+x} dx$. **Hint:** This equals

$$8 \int e^{x \ln 8} dx.$$

Answer:

$$\int 8^{1+x} dx = \frac{8}{\ln 8} 8^x + C$$

11. Find $\int \frac{\sqrt{(36+x^2)} + \sqrt{(36-x^2)}}{\sqrt{(1296-x^4)}} dx$.

Hint: The integrand equals $\frac{\sqrt{(36+x^2)} + \sqrt{(36-x^2)}}{\sqrt{(36-x^2)}\sqrt{(36+x^2)}}$

$$= \frac{1}{\sqrt{36-x^2}} + \frac{1}{\sqrt{36+x^2}}.$$

Answer:

$$\int \frac{\sqrt{36+x^2} + \sqrt{36-x^2}}{\sqrt{1296-x^4}} dx = \int \frac{1}{\sqrt{36-x^2}} dx +$$

$$\int \frac{1}{\sqrt{(36+x^2)}} dx = \arcsin \frac{1}{6}x +$$

$$\ln(x + \sqrt{36+x^2}) + C$$

12. Find $\int (7x+1)^{30} dx$.

Answer:

$$\int (7x+1)^{30} dx = \frac{1}{217} (7x+1)^{31} + C$$

13. Find $\int \sqrt{1+\sin 3x} dx$. **Hint:** The integrand equals $\frac{\cos(3x)}{\sqrt{1-\sin(3x)}}$.

Answer:

$$\int \sqrt{1+\sin 3x} dx =$$

$$\frac{2}{3} (\sin 3x - 1) \frac{\sqrt{(1+\sin 3x)}}{\cos 3x} + C$$

14. Find $\int (\sqrt{3+2x})^5 dx$.

Answer:

$$\int (\sqrt{3+2x})^5 dx = \frac{1}{7} (\sqrt{3+2x})^7 + C$$

15. Find $\int \frac{1}{49+4x^2} dx$.

Answer:

$$\int \frac{1}{49+4x^2} dx = \frac{1}{14} \arctan \frac{2}{7}x + C$$

16. Find $\int \frac{1}{\sqrt{4x^2-9}} dx$.

Answer:

$$\int \frac{1}{\sqrt{4x^2-9}} dx =$$

$$\frac{1}{2} \ln(2x + \sqrt{4x^2-9}) + C$$

17. Find $\int \frac{1}{\sin^2(2x+3)} dx$.

Answer:

$$\int \frac{1}{\sin^2(2x+3)} dx = -\frac{1}{2\sin(2x+3)} \cos(2x+3) + C$$

18. Find $\int \frac{1}{1+\cos 6x} dx$. **Hint:** You could let $u = 6x$ and then use the technique for rational functions of $\cos x$ and $\sin x$.

Answer:

$$\int \frac{1}{1+\cos(6x)} dx = \frac{1}{6} \tan 3x + C.$$

19. Find $\int \frac{1}{1+\sin(3x)} dx$.

Answer:

$$\int \frac{1}{1+\sin 3x} dx = -\frac{2}{3(1+\tan \frac{3}{2}x)} + C$$

20. Find $\int x^2 \sqrt{4x^3+2} dx$. **Hint:** Let $u = 4x^3+2$.

Answer:

$$\int x^2 \sqrt{4x^3+2} dx = \frac{1}{18} (\sqrt{4x^3+2})^3 + C$$

21. Find $\int x^8 (\sqrt{3x^9+2})^5 dx$.

Answer:

$$\int x^8 (\sqrt{3x^9+2})^5 dx = \frac{2}{189} (\sqrt{3x^9+2})^7 + C$$

22. Find $\int \frac{x}{3+8x^4} dx$. **Hint:** Try $u = x^2$.

Answer:

$$\int \frac{x}{3+8x^4} dx = \frac{1}{24} \sqrt{6} \arctan \left(\frac{2}{3} x^2 \sqrt{6} \right) + C$$

23. Find $\int \frac{1}{3(2+x)\sqrt{x}} dx$. **Hint:** Try $x = u^2$.

Answer:

$$\int \frac{1}{3(2+x)\sqrt{x}} dx = \frac{1}{9} \sqrt{3} \sqrt{6} \arctan \left(\frac{1}{6} \sqrt{3} \sqrt{x} \sqrt{6} \right) + C$$

24. Find $\int \frac{1}{x\sqrt{(25x^2-9)}} dx$.

Answer:

You could let $3 \sec u = 5x$ so

$(3 \sec u \tan u) du = 5 dx$ and then

$$\int \frac{1}{x\sqrt{(25x^2-9)}} dx = \frac{1}{3} \int du = \frac{u}{3} + C.$$

Now restoring the original variables, this yields $\frac{1}{3} \operatorname{arcsec} \frac{5}{3} |x| + C$.

25. Find $\int \frac{1}{\sqrt{x(3x-5)}} dx$. **Hint:** You might try completing the square in $x(3x-5)$ and then changing the variable in an appropriate manner.

Answer:

$$\int \frac{dx}{\sqrt{x(3x-5)}} = \frac{1}{3} \sqrt{3} \ln \left(\sqrt{3} \left(x - \frac{5}{6} \right) + \sqrt{3x^2 - 5x} \right) + C$$

26. Find $\int \frac{1}{x \cos(\ln 4x)} dx$. **Hint:** You might try letting $u = \ln(4x)$.

Answer:

$$\int \frac{1}{x \cos(\ln 4x)} dx = \ln(\sec(\ln 4x) + \tan(\ln 4x)) + C$$

27. Find $\int x^7 e^{x^8} dx$.

Answer:

$$\int x^7 e^{x^8} dx = \frac{1}{8} e^{x^8} + C$$

28. Find $\int \frac{\ln^4 x}{x} dx$.

Answer:

$$\int \frac{\ln^4 x}{x} dx = \frac{1}{5} \ln^5 x + C$$

29. Find $\int \frac{1}{x(2+\ln 2x)} dx$.

Answer:

$$\int \frac{1}{x(2+\ln 2x)} dx = \ln(2 + \ln 2x) + C$$

30. Find $\int \frac{\sin 7x + \cos 7x}{\sqrt{\sin 7x - \cos 7x}} dx$. **Hint:** Try $u = \sin 7x - \cos 7x$.

Answer:

$$\int \frac{\sin 7x + \cos 7x}{\sqrt{(\sin 7x - \cos 7x)}} dx = \frac{2}{7} \sqrt{(\sin 7x - \cos 7x)} + C$$

31. Find $\int \csc 4x dx$.

Answer:

$$\int \csc 4x dx = \frac{1}{4} \ln |\csc 4x - \cot 4x| + C$$

32. Find $\int \frac{\arctan 5x}{1+25x^2} dx$. **Hint:** Try $u = \arctan(5x)$.

Answer:

$$\int \frac{\arctan 5x}{1+25x^2} dx = \frac{1}{10} \arctan^2 5x + C$$

33. Find $\int \frac{18}{(6+2x)(6-x)} \cos \left(\ln \frac{6+2x}{6-x} \right) dx$.

Hint: It might help to first let

$u = \ln \frac{6+2x}{6-x}$ and see if it simplifies.

Answer:

$$\int \frac{18}{(6+2x)(6-x)} \cos \left(\ln \frac{6+2x}{6-x} \right) dx =$$

$$\sin \left(\ln \frac{6+2x}{6-x} \right) + C$$

34. Find $\int x^{23} (2 - 6x^{12})^{10} dx$

Hint: Maybe let $u = 2 - 6x^{12}$.

Answer:

$$\int x^{23} (2 - 6x^{12})^{10} dx =$$

$$\frac{1}{5184} (2 - 6x^{12})^{11} - \frac{1}{2376} (2 - 6x^{12})^{11} + C$$

35. Find $\int \frac{x^5}{\sqrt{6-3x^2}} dx$.

Answer:

$$\int \frac{x^5}{\sqrt{6-3x^2}} dx =$$

$$-\frac{1}{15} x^4 \sqrt{6-3x^2} - \frac{8}{45} x^2 \sqrt{6-3x^2}$$

$$- \frac{32}{45} \sqrt{6-3x^2} + C$$

36. Find $\int \cos^3(3x) \sin^{\frac{1}{2}}(3x) dx$.

Answer:

$$\int \cos^3 3x \sin^{\frac{1}{2}} 3x dx =$$

$$\int \left(\cos 3x (1 - \sin^2 3x) \sin^{\frac{1}{2}} 3x \right) dx$$

$$= -\frac{2}{21} \sin^{\frac{7}{2}} 3x + \frac{2}{9} \sin^{\frac{3}{2}} 3x + C$$

37. Find $\int \frac{1}{e^{2x} + e^x} dx$. **Hint:** Try $u = e^x$.

Answer:

$$\int \frac{1}{e^{2x} + e^x} dx = \frac{-1 - xe^x}{e^x} + \ln(e^x + 1) + C$$

38. Find $\int \frac{1}{\sqrt{e^x + 1}} dx$.

Answer:

Let $u^2 = e^x$ so $2udu = e^x dx = au^2 dx$. In terms of u this is

$$2 \int \frac{1}{u\sqrt{1+u^2}} du. \text{ Now let}$$

$u = \tan \theta$ so $du = (\sec^2 \theta) d\theta$. Then the indefinite integral becomes

$$2 \int \frac{\sec^2 \theta}{\tan \theta \sec(\theta)} d\theta = 2 \int \csc \theta d\theta =$$

$$2 \ln |\csc \theta - \cot \theta| + C.$$

In terms of u this is

$$2 \ln \left| \frac{\sqrt{u^2+1}}{u} - \frac{1}{u} \right| + C \text{ and in terms of } x \text{ this is}$$

$$2 \ln \left| \frac{\sqrt{(e^x+1)}}{e^{\frac{1}{2}x}} - e^{-\frac{1}{2}x} \right| + C.$$

39. Find $\int \frac{\arctan \sqrt{x}}{\sqrt{x}(1+x)} dx$.

Answer:

$$\int \frac{\arctan \sqrt{x}}{\sqrt{x}(1+x)} = \arctan^2 \sqrt{x} + C$$

40. Find $\int \sqrt{\frac{1+x}{1-x}} dx$.

Answer:

Multiply the fraction on the top and bottom by $1+x$ to get

$$\int \frac{1+x}{\sqrt{1-x^2}} dx. \text{ Now let } x = \sin \theta \text{ so } dx = \cos \theta d\theta. \text{ Then this is}$$

$$\int \frac{1+\sin \theta}{\sqrt{1-\sin^2 \theta}} \cos \theta d\theta =$$

$$\int (1 + \sin \theta) d\theta = \theta - \cos \theta + C.$$

In terms of x this gives

$$\arcsin x - \sqrt{1-x^2} + C.$$

41. Find $\int \sqrt{\frac{x-2}{x+2}} dx$.

Answer:

Multiply the fraction on the top and bottom by $x+2$ to get $\int \frac{\sqrt{x^2-4}}{x+2} dx$

Now let $x = 2 \sec \theta$ so $dx = 2 \sec \theta \tan \theta d\theta$ and in terms of θ the indefinite integral is

$$2 \int \frac{\sqrt{\sec^2(\theta)-1}}{\sec \theta + 1} \sec \theta \tan \theta d\theta =$$

$$2 \int \frac{\tan^2 \theta \sec(\theta)}{1+\sec \theta} d\theta$$

$$= 2 \int \frac{\tan^2 \theta}{1+\cos \theta} = 2 \int \sec^2 \theta d\theta -$$

$$2 \int \sec \theta d\theta =$$

$2 \tan \theta - 2 \ln (\sec \theta + \tan \theta) + C$. Now in terms of x this is

$$\sqrt{(x^2-4)} -$$

$$2 \ln \left(\frac{1}{2}x + \frac{1}{2}\sqrt{(x^2-4)} \right) + C.$$

42. Find $\int \frac{x^2}{\sqrt{(4+9x^2)}} dx$.

Answer:

$$\int \frac{x^2}{\sqrt{4+9x^2}} dx =$$

$$\frac{1}{18} x \sqrt{4+9x^2} -$$

$$\frac{2}{27} \ln (3x + \sqrt{4+9x^2}) + C$$

43. Find $\int \frac{1}{\sqrt{x^2+49}} dx$.

Answer:

$$\int \frac{1}{\sqrt{x^2+49}} dx = \ln(x + \sqrt{x^2 + 49}) + C$$

44. Find $\int \frac{1}{\sqrt{x^2-36}} dx$.

Answer:

$$\int \frac{1}{\sqrt{x^2-36}} dx = \ln(x + \sqrt{x^2 - 36}) + C$$

45. Find $\int x \ln(3x) dx$.

Answer:

$$\int x \ln(3x) dx = \frac{1}{2}x^2 \ln 3x - \frac{1}{4}x^2 + C$$

46. Find $\int x \ln^2(6x) dx$.

Answer:

$$\int x \ln^2 6x dx = \frac{1}{2}x^2 \ln^2 6x - \frac{1}{2}x^2 \ln 6x + \frac{1}{4}x^2 + C$$

47. Find $\int x^3 e^{x^2} dx$.

Answer:

$$\int x^3 e^{x^2} dx = \frac{1}{2}x^2 e^{x^2} - \frac{1}{2}e^{x^2} + C$$

48. Find $\int x^2 \sin 8x dx$

Answer:

$$\begin{aligned} \int x^2 \sin 8x dx &= \\ -\frac{1}{8}x^2 \cos 8x + \frac{1}{256} \cos 8x + \frac{1}{32}x \sin 8x + C \end{aligned}$$

49. Find $\int \arcsin x dx$

Answer:

$$\begin{aligned} \int \arcsin x dx &= \\ x \arcsin x + \sqrt{1-x^2} + C \end{aligned}$$

50. Find $\int \arctan x dx$

Answer:

$$\begin{aligned} \int \arctan x dx &= \\ x \arctan x - \frac{1}{2} \ln(1+x^2) + C \end{aligned}$$

51. Find $\int \frac{\arctan 6x}{x^3} dx$. **Hint:** Do integration by parts on this one. This will get things started. Then recall partial fractions.

Answer:

$$\begin{aligned} \int \frac{\arctan 6x}{x^3} dx &= \\ -\frac{1}{2x^2} \arctan 6x - \frac{3}{x} - 18 \arctan 6x + C \end{aligned}$$

52. Find $\int \sin(4x) \ln(\tan 4x) dx$

Answer:

Integration by parts gives

$$\begin{aligned} &-\frac{1}{4} \cos 4x \ln(\tan 4x) + \\ &\frac{1}{4} \ln(\csc 4x - \cot 4x) + C \end{aligned}$$

53. Find $\int e^{2x} \sqrt{e^{4x} + 1} dx$. **Hint:** Try $u = e^{2x}$. This will yield something which will look a little different than the answer I have given below.

Answer:

$$\begin{aligned} &\frac{1}{2} \int \sqrt{u^2 + 1} du \\ &\int e^{2x} \sqrt{(e^{4x} + 1)} dx \\ &= \frac{1}{4} e^{2x} \sqrt{(e^{4x} + 1)} + \frac{1}{4} \operatorname{arcsinh}(e^{2x}) \\ &+ C \end{aligned}$$

54. Find $\int \cos(\ln 7x) dx$.

Answer:

$$\begin{aligned} \int \cos(\ln 7x) dx &= x \cos(\ln 7x) \\ &+ \int x \sin(\ln(7x)) \left(\frac{1}{x}\right) dx \\ &= x \cos(\ln 7x) + \\ &[x \sin(\ln 7x) - \int \cos(\ln(7x)) dx] \\ &\text{and so} \\ \int \cos(\ln 7x) dx &= \\ \frac{1}{2} [x \cos(\ln 7x) + x \sin(\ln 7x)] + C \end{aligned}$$

55. Find $\int \sin(\ln 7x) dx$.

Answer:

$$\begin{aligned} \int \sin(\ln 7x) dx &= \\ \frac{1}{2} [x \sin(\ln 7x) - x \cos(\ln 7x)] + C \end{aligned}$$

56. Find $\int e^{5x} \sin 3x dx$.

Answer:

$$\begin{aligned} \int e^{5x} \sin 3x dx &= \\ -\frac{3}{34} e^{5x} \cos 3x + \frac{5}{34} e^{5x} \sin 3x + C \end{aligned}$$

57. Find $\int e^{3x} \sin^2 2x dx$.

Answer:

$$\begin{aligned} \int e^{3x} \sin^2 2x dx &= \\ \frac{1}{6} e^{3x} - \frac{3}{25} e^{3x} (\cos 2)x - \frac{4}{25} e^{3x} (\sin 2) \\ &x + C \end{aligned}$$

58. Find $\int \frac{x+4}{(x+3)^2} dx$.

Answer:

$$\begin{aligned} \int \frac{x+4}{(x+3)^2} dx &= \\ \ln(x+3) - \frac{1}{x+3} + C \end{aligned}$$

59. Find $\int \frac{7x^2+100x+347}{(x+2)(x+7)^2} dx$.

Answer:

$$\int \frac{7x^2+100x+347}{(x+2)(x+7)^2} dx$$

$$= 7 \ln(x+2) - \frac{2}{x+7} + C$$

60. Find $\int \frac{3x^2+7x+8}{(x+2)(x^2+2x+2)} dx$.

Answer:

$$\int \frac{3x^2+7x+8}{(x+2)(x^2+2x+2)} dx$$

$$= 3 \ln(x+2) + \arctan(1+x) + C$$

61. Find $\int \frac{4x^2+65x+266}{(x+2)(x^2+16x+66)} dx$.

Answer:

$$\int \frac{4x^2+65x+266}{(x+2)(x^2+16x+66)} dx$$

$$= 4 \ln(2+x)$$

$$+ \frac{1}{2} \sqrt{2} \arctan \frac{1}{4} (2x+16) \sqrt{2}$$

62. Find $\int \frac{1}{x^3+8} dx$. **Hint:** You need to first factor x^3+8 and then use partial fractions.

Answer:

$$\int \frac{1}{x^3+8} dx =$$

$$\frac{1}{12} \ln(x+2) - \frac{1}{24} \ln(x^2-2x+4)$$

$$+ \frac{1}{12} \sqrt{3} \arctan \frac{1}{6} (2x-2) \sqrt{3} + C$$

63. Find $\int \frac{1}{x^4+x^2+1} dx$.

Answer:

$$\int \frac{1}{x^4+x^2+1} dx =$$

$$\frac{1}{4} \ln(x^2+x+1) +$$

$$\frac{1}{6} \sqrt{3} \arctan \frac{1}{3} (2x+1) \sqrt{3}$$

$$- \frac{1}{4} \ln(x^2-x+1) +$$

$$\frac{1}{6} \sqrt{3} \arctan \frac{1}{3} (2x-1) \sqrt{3} + C$$

64. Find $\int \frac{1}{x^4+81} dx$. **Hint:**

$$x^4+81 =$$

$$(x^2-3x\sqrt{2}+9)(x^2+3x\sqrt{2}+9).$$

Answer:

Now you can use partial fractions to find

$$\int \frac{1}{x^4+81} dx =$$

$$\frac{1}{216} \sqrt{2} \ln \frac{x^2+3x\sqrt{2}+3}{x^2-3x\sqrt{2}+3} +$$

$$\frac{1}{108} \sqrt{2} \arctan \left(\frac{1}{3} x \sqrt{2} + 1 \right)$$

$$+ \frac{1}{108} \sqrt{2} \arctan \left(\frac{1}{3} x \sqrt{2} - 1 \right) + C$$

65. Find $\int \frac{1}{x^6+64} dx$.

Answer:

First factor the denominator.

$$x^6+64 =$$

$$(x^2+4) \left((x-\sqrt{3})^2+1 \right)$$

$$\cdot \left((x+\sqrt{3})^2+1 \right).$$

$$\frac{1}{x^6+64} = \frac{Ax+B}{x^2+4} + \frac{Cx+D}{(x-\sqrt{3})^2+1}$$

$$+ \frac{Ex+F}{(x+\sqrt{3})^2+1} \text{ and so after wading}$$

through much affliction, the partial fractions decomposition is

$$\frac{16}{3x^2+12} + \frac{-\frac{1}{192}\sqrt{3}x+\frac{1}{48}}{(x-\sqrt{3})^2+1} + \frac{\frac{1}{192}\sqrt{3}x+\frac{1}{48}}{(x+\sqrt{3})^2+1}.$$

Therefore, the indefinite integral is

$$\int \left(\frac{16}{3x^2+12} + \frac{-\frac{1}{192}\sqrt{3}x+\frac{1}{48}}{(x-\sqrt{3})^2+1} + \right.$$

$$\left. + \frac{\frac{1}{192}\sqrt{3}x+\frac{1}{48}}{(x+\sqrt{3})^2+1} \right) dx =$$

$$\frac{1}{96} \arctan \frac{1}{2} x - \frac{1}{384} \sqrt{3} \ln(x^2-2\sqrt{3}x+4)$$

$$+ \frac{1}{192} \arctan(x-\sqrt{3}) +$$

$$\frac{1}{384} \sqrt{3} \ln(x^2+2\sqrt{3}x+4) +$$

$$\frac{1}{192} \arctan(x+\sqrt{3}) + C$$

66. Find $\int \frac{x^2}{(x-3)^{100}} dx$.

Hint: You ought to let $u = x - 3$.

Answer:

$$\int \frac{x^2}{(x-3)^{100}} dx$$

$$= \frac{1}{11(3-x)^{99}} - \frac{3}{49(3-x)^{98}} + \frac{1}{97(3-x)^{97}} + C$$

67. Find $\int 2 \frac{4+4x+3x^2+x^3}{(x^2+1)(x^2+2x+5)} dx$.

Answer:

$$\int 2 \frac{4+4x+3x^2+x^3}{(x^2+1)(x^2+2x+5)} dx$$

$$= \frac{1}{2} \ln(x^2+1) + \arctan x +$$

$$\frac{1}{2} \ln(x^2+2x+5) + \arctan\left(\frac{1}{2} + \frac{1}{2}x\right)$$

$$+ C$$

68. Find $\int \frac{1-x^7}{x(1+x^7)} dx$

Answer:

$$\int \frac{1-x^7}{x(1+x^7)} dx$$

$$= \ln x - \frac{2}{7} \ln(1+x^7) + C$$

69. Find $\int \frac{x^2+1}{x^4-2x^2+1} dx$

Answer:

$$\int \frac{x^2+1}{x^4-2x^2+1} dx = -\frac{1}{2(x-1)} - \frac{1}{2(1+x)} + C$$

70. Find $\int \frac{x^5}{x^8+1} dx$

Answer:

Let $u = x^2$. $\int \frac{x^5}{x^8+1} dx$
 $= \int \frac{u^2}{2u^4+2} du$. Now use partial fractions.
 $\int \frac{x^5}{x^8+1} dx = \frac{1}{16} \sqrt{2} \ln \frac{x^4-x^2\sqrt{2}+1}{x^4+x^2\sqrt{2}+1}$
 $+ \frac{1}{8} \sqrt{2} \arctan(x^2\sqrt{2}+1)$
 $+ \frac{1}{8} \sqrt{2} \arctan(x^2\sqrt{2}-1) + C$

71. Find $\int \frac{x^2+4}{x^6+64} dx$.

Hint: $x^6 + 64 = (x^2 + 4) \cdot (x^4 - 4x^2 + 16)$.

Answer:

$$\int \frac{x^2+4}{x^6+64} dx = -\frac{1}{96} \sqrt{3} \ln(x^2 - 2\sqrt{3}x + 4) + \frac{1}{16} \arctan(x - \sqrt{3}) + \frac{1}{96} \sqrt{3} \ln(x^2 + 2\sqrt{3}x + 4) + \frac{1}{16} \arctan(x + \sqrt{3}) + C$$

72. Find $\int \frac{dx}{x(2+3\sqrt{x}+\sqrt[3]{x})}$.

Answer:

$$\int \frac{dx}{x(2+3\sqrt{x}+\sqrt[3]{x})} = \int \frac{6}{u(2+3u^3+u^2)} du = 3 \ln u - \frac{6}{7} \ln(1+u) - \frac{15}{14} \ln(3u^2 - 2u + 2) - \frac{3}{35} \sqrt{5} \arctan \frac{1}{10} (6u - 2) \sqrt{5} + C$$

$$= 3 \ln x^{1/6} - \frac{6}{7} \ln(1 + x^{1/6}) - \frac{15}{14} \ln(3x^{1/3} - 2x^{1/6} + 2) - \frac{3}{35} \sqrt{5} \arctan \frac{1}{10} (6x^{1/6} - 2) \sqrt{5} + C$$

73. Find $\int \frac{\sqrt{x+3}-\sqrt{x-3}}{\sqrt{x+3}+\sqrt{x-3}} dx$.

Hint: Show the integrand equals $\frac{1}{3}x - \frac{1}{3}\sqrt{x^2-9}$.

Answer:

$$\int \frac{\sqrt{x+3}-\sqrt{x-3}}{\sqrt{x+3}+\sqrt{x-3}} dx = \int \left(\frac{1}{3}x - \frac{1}{3}\sqrt{x^2-9} \right) dx = \frac{1}{6}x^2 - \frac{1}{6}x\sqrt{x^2-9} + \frac{3}{2} \ln(x + \sqrt{x^2-9}) + C$$

74. Find $\int \frac{x^2}{\sqrt{x^2+6x+13}} dx$. **Hint:** Complete the square in the denominator.

Answer:

$$\int \frac{x^2}{\sqrt{x^2+6x+13}} dx = \frac{1}{2}x\sqrt{x^2+6x+13} - \frac{9}{2}\sqrt{x^2+6x+13} + 7 \ln(x + 3 + \sqrt{x^2+6x+13}) + C$$

75. Find $\int \frac{\sqrt{x^2+4x+5}}{x} dx$.

Answer:

$$\int \frac{\sqrt{x^2+4x+5}}{x} dx = \sqrt{x^2+4x+5} + 2 \operatorname{arcsinh}(x+2) - \sqrt{5} \operatorname{arctanh} \frac{1}{5} (5+2x) \frac{\sqrt{5}}{\sqrt{x^2+4x+5}} + C$$

You might try letting $u = \sinh^{-1}(x+2)$.

76. Find $\int \frac{1}{x^3\sqrt{x^2+9}} dx$.

Answer:

$$\int \frac{1}{x^3\sqrt{x^2+9}} dx = -\frac{1}{18x^2} \sqrt{x^2+9} + \frac{1}{54} \operatorname{arctanh} \frac{3}{\sqrt{x^2+9}} + C$$

You might try letting $u = \frac{1}{\sqrt{x^2+9}}$.

77. Find $\int \frac{dx}{x^4\sqrt{x^2-9}}$.

Answer:

$$\int \frac{dx}{x^4\sqrt{x^2-9}} = \frac{1}{27x^3} \sqrt{(x^2-9)} + \frac{2}{243x} \sqrt{(x^2-9)} + C$$

78. Find $\int \sin^6(3x) dx$.

Answer:

$$\int \sin^6(3x) dx = -\frac{1}{2} \sin^5 x \cos x - \frac{5}{8} \sin^3 x \cos x - \frac{15}{16} \cos x \sin x + \frac{5}{16} x + C$$

79. Find $\int \frac{\sin^3(x)}{\cos^4(x)} dx$.

Answer:

$$\int \frac{\sin^3(x)}{\cos^4(x)} dx = \frac{1}{3} \frac{\sin^4 x}{\cos^3 x} - \frac{1}{3} \frac{\sin^4 x}{\cos x} - \frac{1}{3} \sin^2 x \cos x - \frac{2}{3} \cos x + C$$

80. Find $\int \tan^5(2x) dx$.

Answer:

$$\int \tan^5(2x) dx = -\frac{1}{4} \tan^2 2x + \frac{1}{8} \tan^4 2x + \frac{1}{4} \ln(2 + 2 \tan^2 2x) + C$$

81. Find $\int \frac{dx}{\sqrt{\tan(2x)}}$.

Answer:

$$\int \frac{dx}{\sqrt{\tan(2x)}} =$$

$$2 \frac{\tan^{\frac{1}{2}} x}{1+2 \tan^2 x} + 2 \frac{\tan^{\frac{5}{2}} x}{1+2 \tan^2 x} - 2 \tan^{\frac{1}{2}} x + \frac{1}{4} \sqrt{2} \arctan 2\sqrt{2} \frac{\tan^{\frac{1}{2}} x}{1-2 \tan x} + \frac{1}{4} \sqrt{2} \ln \frac{2 \tan x + 2\sqrt{2} \tan^{\frac{1}{2}} x + 1}{\sqrt{(1+2 \tan^2 x)}} + C$$

You might try the substitution, $u = \tan(x)$.

82. Find $\int \frac{dx}{\cos x + 3 \sin x + 4}$.

Answer:

$$\int \frac{dx}{\cos x + 3 \sin x + 4} = \frac{1}{3} \sqrt{6} \arctan \frac{1}{12} (6 \tan \frac{1}{2} x + 6) \sqrt{6} + C$$

Try the substitution, $u = \tan(\frac{x}{2})$.

6.12 Computers And Antiderivatives

It should be fairly clear by now that there are standard techniques for finding antiderivatives which sometimes work. All the main methods have been presented but there are many functions which are considered important enough that they are given names even though they are not given names in the calculus books. Sometimes the desired antiderivative is expressible in terms of one of these functions. Computer algebra systems know about these functions even though you have not encountered them. Also, the routine mechanical process of searching for an antiderivative can often be done by a computer more quickly than you could do it by hand. For example to find $\int \frac{\sin x}{x} dx$ you use the maple input

$$> \text{int}(\sin(x)/x, x);$$

and then press return. This gives $\text{Si}(x)$ for the answer. Of course you need to add in an arbitrary constant of integration for it to be correct. Maple doesn't bother to do this for you. This function, Si is one not usually studied in calculus but it is known to the computer algebra system. Sometimes the integrals involve elementary functions but they are very hard to do. For example suppose you wanted to find $\int \frac{1}{(x^4+1)} dx$. In Maple this would involve the command

$$> \text{int}(1/(1+x^4), x);$$

You would then press return and it will give $\frac{1}{8} \sqrt{2} \ln \frac{x^2+x\sqrt{2}+1}{x^2-x\sqrt{2}+1} + \frac{1}{4} \sqrt{2} \arctan(x\sqrt{2}+1) + \frac{1}{4} \sqrt{2} \arctan(x\sqrt{2}-1)$. Then you add in the arbitrary constant of integration and you are done with the wretched task of finding the antiderivative. There are many versions of computer algebra systems right now other than Maple. Scientific workplace, the software used to type this book, contains one very easy to use system which is not as good as the full version of Maple or Mathematica but which is entirely adequate for finding antiderivatives. In fact, this system is what actually computed the above answers. All that was necessary to find $\int \frac{x}{(x^4+4)} dx$ was to enter the preceding expression in math mode and then to press evaluate. This yields $\int \frac{x}{(x^4+4)} dx = \frac{1}{4} \arctan \frac{1}{2} x^2$. Then to make it right, you need to add in an arbitrary constant of integration.

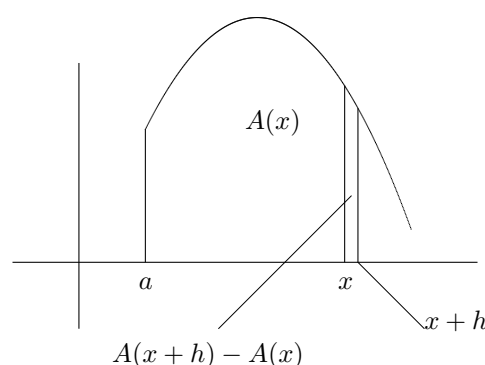
It is not the purpose of this book to feature computer algebra systems at length but it is important that you realize these things exist and make finding antiderivatives fairly

routine when they can be found. If you have access to a computer algebra system, you should experiment with using it to find antiderivatives. It is fun to have a machine do laborious computations for you. However, it is essential you acquire some facility with the various techniques independent of a machine.

Applications Of Antiderivatives

7.1 Areas

Consider the problem of finding the area between the graph of a function of one variable and the x axis as illustrated in the following picture.



The curved line on the top represents the graph of the function $y = f(x)$ and the symbol, $A(x)$ represents the area between this curve and the x axis between the point, a and the point x as shown. The vertical line from the point, $x + h$ up to the curve and the vertical line from x up to the curve define the area, $A(x + h) - A(x)$ as indicated. You can see that this area is between $hf(x)$ and $hf(x + h)$. This happens because the function is decreasing near x . In general, for continuous functions, f , Theorem 2.15.11 on Page 79 implies there exists $x_M, x_m \in [x, x + h]$ with the properties

$$f(x_M) \equiv \max \{f(x) : x \in [x, x + h]\}$$

and

$$f(x_m) \equiv \min \{f(x) : x \in [x, x + h]\}$$

Then,

$$f(x_m) = \frac{hf(x_m)}{h} \leq \frac{A(x + h) - A(x)}{h} \leq \frac{hf(x_M)}{h} = f(x_M).$$

Therefore, using the squeezing theorem, Theorem 2.9.5, and the continuity of f ,

$$A'(x) \equiv \lim_{h \rightarrow 0} \frac{A(x + h) - A(x)}{h} = f(x).$$

The consideration of $h < 0$ is also straightforward. This discussion implies the following theorem.

Theorem 7.1.1 *Let $a < b$ and let $f : [a, b] \rightarrow [0, \infty)$ be continuous. Then letting $A(x)$ denote the area between a, x , the graph of the function, and the x axis,*

$$A'(x) = f(x) \text{ for } x \in (a, b), \quad A(a) = 0. \quad (7.1)$$

Also, A is continuous on $[a, b]$.

The problem for A described in the above theorem is called an initial value problem and the equation, $A'(x) = f(x)$ is a differential equation. It is called this because it is an equation for an unknown function, $A(x)$ written in terms of the derivative of this unknown function. The assertion that A should be continuous on $[a, b]$ follows from the fact that it has to be continuous on (a, b) because of the existence of its derivative and the above argument can also be used to obtain one sided derivatives for A at the end points, a and b , which yields continuity on $[a, b]$.

Example 7.1.2 *Let $f(x) = x^2$ for $x \in [1, 2]$. Find the area between the graph of the function, the points 1 and 2, and the x axis.*

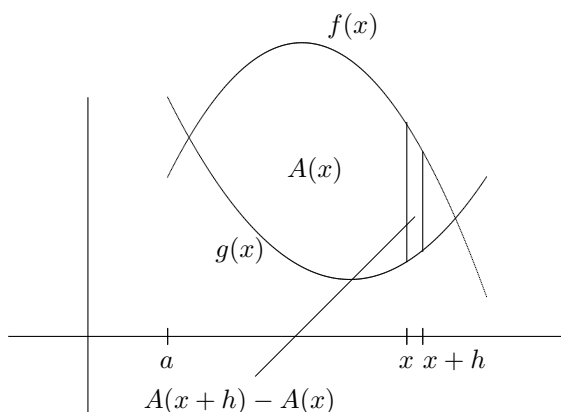
The function, $\frac{x^3}{3} + C$ has the property that its derivative gives x^2 . This is true for any C . It only remains to choose C in such a way that the function equals zero at $x = 1$. Thus $C = -\frac{1}{3}$. It follows that $A(x) = \frac{x^3}{3} - \frac{1}{3}$. Therefore, the area described equals $A(2) = \frac{8}{3} - \frac{1}{3} = \frac{7}{3}$ square units.

Example 7.1.3 *Find the area between the graph of the function $y = 1/x^2$ and the x axis for x between $1/2$ and 3 .*

The function $-\frac{1}{x} + C$ has the property that its derivative equals $1/x^2$. Letting $C = 2$, $A(x) = -\frac{1}{x} + 2$ satisfies the appropriate initial value problem and so the area equals $A(3) = \frac{5}{3}$.

7.2 Area Between Graphs

It is a minor generalization to consider the area between the graphs of two functions. Consider the following picture.



You see that sometimes the function, $f(x)$ is on top and sometimes the function, $g(x)$ is on top. It is the length of the vertical line joining the two graphs which is

of importance and this length is always $|f(x) - g(x)|$ regardless of which function is larger. By Theorem 2.15.11 on Page 79 there exist $x_M, x_m \in [x, x+h]$ satisfying

$$|f(x_M) - g(x_M)| \equiv \max \{|f(x) - g(x)| : x \in [x, x+h]\}$$

and

$$|f(x_m) - g(x_m)| \equiv \min \{|f(x) - g(x)| : x \in [x, x+h]\}.$$

Then

$$\frac{|f(x_m) - g(x_m)|h}{h} \leq \frac{A(x+h) - A(x)}{h} \leq \frac{|f(x_M) - g(x_M)|h}{h},$$

and using the squeezing theorem, Theorem 2.9.5 on Page 60, as $h \rightarrow 0$

$$A'(x) = |f(x) - g(x)|$$

Also $A(a) = 0$ as before. This yields the following theorem which generalizes the one presented earlier because the x axis is the graph of the function, $y = 0$.

Theorem 7.2.1 *Let $a < b$ and let $f, g : [a, b] \rightarrow \mathbb{R}$ be continuous. Then letting $A(t)$ denote the area between the graphs of the two functions for $x \in [a, t]$,*

$$A'(t) = |f(t) - g(t)| \text{ for } t \in (a, b), \quad A(a) = 0. \quad (7.2)$$

Also, A is continuous on $[a, b]$.

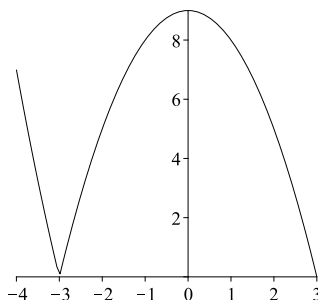
This theorem provides the justification for the following procedure for finding the area between the graphs of two functions which also applies to more general situations than described in the above theorem.

Procedure 7.2.2 *To find the area between the graphs of the functions $y = f(x)$ and $y = g(x)$ for $x \in [a, b]$, split the interval into non overlapping subintervals, $I_1, I_2, \dots, I_k$ which have the property that on I_i , $|f(x) - g(x)|$ equals either $f(x) - g(x)$ or $g(x) - f(x)$. If $I_i = [p_i, q_i]$, take an antiderivative of $|f(x) - g(x)|$ on $[p_i, q_i]$, H_i . (This might be a reasonable problem because on this interval, you won't need to write in absolute value signs.) Then the area between the curves for $x \in I_i$ is $H_i(q_i) - H_i(p_i)$. The desired area is the sum of these.*

Proof: Consider the area between the curves for $x \in [p_i, q_i]$. You need $A(q_i)$ where $A' = |f(x) - g(x)|$ and $A(p_i) = 0$. Let H be any antiderivative. Then from Lemma 6.1.3, $A(x) = H(x) + C$ where C is some constant. Thus $C = -H(p_i)$ and so $A(q_i) = H(q_i) - H(p_i)$.

Example 7.2.3 *Let $f(x) = 8 - \frac{x^2}{2}$ and $g(x) = \frac{x^2}{2} - 1$. Find the area between the graphs of the two functions for $x \in [-4, 3]$.*

The following graph is of the absolute value of the difference of the two functions.



You should always draw such a graph to determine what is going on. $|f(x) - g(x)| =$

$$|9 - x^2| = \begin{cases} x^2 - 9 & \text{if } x \in [-4, -3] \\ 9 - x^2 & \text{if } x \in [-3, 3] \end{cases}$$

It follows that on $(-4, -3)$, an antiderivative is $H(x) = \frac{x^3}{3} - 9x$. Therefore, the area between the curves for $x \in [-4, -3]$ is

$$\left(\frac{(-3)^3}{3} - 9(-3) \right) - \left(\frac{(-4)^3}{3} - 9(-4) \right) = \frac{10}{3}$$

Now consider $x \in [-3, 3]$. On this interval, an antiderivative is $H(x) = 9x - \frac{x^3}{3}$ and so the area between the curves for x in this interval is

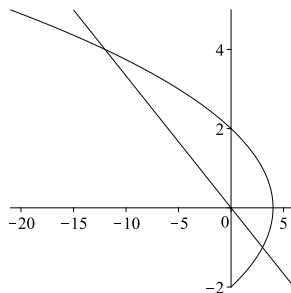
$$\left(9(3) - \frac{(3)^3}{3} \right) - \left(9(-3) - \frac{(-3)^3}{3} \right) = 36$$

The total area is the sum of these. Thus the total area is $36 + \frac{10}{3} = \frac{118}{3}$.

A similar procedure holds for finding the area between two functions which are of the form $x = f(y)$ and $x = g(y)$ for $y \in [c, d]$. You just let y play the role of x in the above.

Example 7.2.4 Find the area between $x = 4 - y^2$ and $x = -3y$.

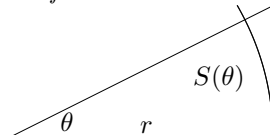
Here is a graph of these two curves.



First find where the two graphs intersect. $4 - y^2 = -3y$. The solution is $y = -1$ and 4 . For y in this interval, you can verify that $4 - y^2 > -3y$ and so $|4 - y^2 - (-3y)| = 4 - y^2 + 3y$. An antiderivative is $4y - \frac{y^3}{3} + \frac{3y^2}{2}$ and so the desired area is

$$4(4) - \frac{(4)^3}{3} + \frac{3(4)^2}{2} - \left(4(-1) - \frac{(-1)^3}{3} + \frac{3(-1)^2}{2} \right) = \frac{125}{6}.$$

Exercise 7.2.5 Find the area of the circular sector shown below.



This is interesting because it is necessary to split the integral into two pieces corresponding to the two different functions which are on the top. The first of these is $y = x \tan(\theta)$ for $x \in [0, r \cos(\theta)]$ and the second is $y = \sqrt{r^2 - x^2}$ for $x \in [r \cos(\theta), r]$. Thus to find the area, consider the two initial value problems

$$A'(x) = x \tan(\theta), \quad A(0) = 0$$

and

$$B'(x) = \sqrt{r^2 - x^2}, \quad B(r \cos(\theta)) = 0$$

Then the total area is

$$A(r \cos(\theta)) + B(r)$$

It follows

$$A(x) = \tan(\theta) \frac{x^2}{2}.$$

To find $B(x)$, use the trig substitution $x = r \sin(u)$. Thus you need to find

$$\begin{aligned} r^2 \int \cos^2(u) du &= r^2 \left(\int \frac{1 + \cos(2u)}{2} \right) \\ &= r^2 \left(\frac{u}{2} + \frac{\sin(2u)}{4} + C \right) = r^2 \left(\frac{u}{2} + \frac{\sin(u) \cos(u)}{2} + C \right) \end{aligned}$$

Using an appropriate triangle to change back to x as described above, this yields

$$\begin{aligned} B(x) &= r^2 \left(\frac{1}{2} \arcsin\left(\frac{x}{r}\right) + \frac{(x/r)(\sqrt{r^2 - x^2}/r)}{2} \right) + C \\ &= \frac{1}{2} \left(\arcsin \frac{x}{r} \right) r^2 + \frac{1}{2} x \sqrt{r^2 - x^2} + C \end{aligned}$$

Since $B(r \cos(\theta)) = 0$,

$$0 = \frac{1}{2} (\arcsin(\cos(\theta))) r^2 + \frac{1}{2} (r \cos(\theta)) \sqrt{r^2 - r^2 \cos^2(\theta)} + C$$

It follows after some simplification that the area of the sector is

$$\frac{r^2}{2} \left(\frac{\pi}{2} - \arcsin(\cos(\theta)) \right)$$

Now from the trig. identities

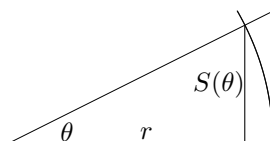
$$\begin{aligned} &\cos\left(\frac{\pi}{2} - \arcsin(\cos(\theta))\right) \\ &= \cos\left(\frac{\pi}{2}\right) \cos(\arcsin(\cos(\theta))) + \sin\left(\frac{\pi}{2}\right) \sin(\arcsin(\cos(\theta))) \\ &= \cos(\theta) \end{aligned}$$

and

$$\begin{aligned} &\sin\left(\frac{\pi}{2} - \arcsin(\cos(\theta))\right) \\ &= \sin\left(\frac{\pi}{2}\right) \cos(\arcsin(\cos(\theta))) + \cos\left(\frac{\pi}{2}\right) \sin(\arcsin(\cos(\theta))) \\ &= \sin\left(\frac{\pi}{2}\right) \cos(\arccos(\sin(\theta))) = \sin(\theta) \end{aligned}$$

and so $\frac{\pi}{2} - \arcsin(\cos(\theta)) = \theta$. Thus the area of this sector is $\frac{\theta}{2} r^2$.

Of course this is entirely the wrong way to go about it! The sensible way is to consider first principles. Instead of rectangles, use little triangles. Consider the following sector of a circle in which θ will be **small**.



Then the area of this sector, $A(\theta)$ lies between the area of the two triangles. Thus

$$\frac{1}{2}r^2 \sin(\theta) \leq A(\theta) \leq \frac{1}{2}r^2 \tan(\theta)$$

Replacing θ by $d\theta$, a very small change in θ and $A(\theta)$ by dA and noting that as θ gets smaller $\sin(\theta)$ and $\tan(\theta)$ are approximately the same and equal to θ ,

$$dA = \frac{1}{2}r^2 d\theta$$

so

$$\frac{dA}{d\theta} = \frac{r^2}{2}.$$

Example 7.2.6 Find the area of the circular sector illustrated in Example 7.2.5

You need to have

$$\frac{dA}{d\theta} = \frac{r^2}{2}, A(0) = 0.$$

The solution is now obviously

$$A(\theta) = \frac{r^2}{2}\theta$$

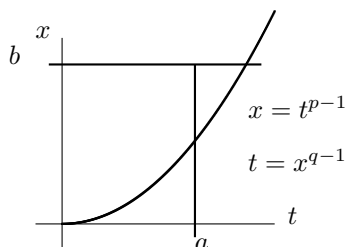
7.3 Exercises

- Find the area between the graphs of the functions, $y = x^2 + 1$ and $y = 3x + 5$.
- Find the area between the graphs of the functions, $y = 2x^2$ and $y = 6x + 8$.
- Find the area between the graphs of $y = 5x + 14$ and $y = x^2$.
- Find the area between the graphs of the functions $y = x + 1$ and $y = 2x$ for $x \in [0, 3]$.
- Find the area between $y = |x|$ and the x axis for $x \in [-2, 2]$.
- Find the area between the graphs of $y = x$ and $y = \sin x$ for $x \in [-\frac{\pi}{2}, \pi]$.
- Find the area between the graphs of $x = y^2$ and $y = 2 - x$.
- Find the area between the x axis and the graph of the function $2x\sqrt{1+x^2}$ for $x \in [0, 2]$. **Hint:** Recall the chain rule for derivatives.
- Show that the area of a right triangle equals one half the product of two sides which are not the hypotenuse.
- Let A denote the region between the x axis and the graph of the function, $f(x) = x - x^2$. For $k \in (0, 1)$, the line $y = kx$ divides this region into two pieces. Explain why there exists a number, k such that the area of these two pieces is exactly equal. **Hint:** This will likely involve the intermediate value theorem. Write an equation satisfied by k and then find an approximate value for k .
- Find the area between the graph of $f(x) = 1/x$ for $x \in [1, 2]$ and the x axis in terms of known functions.
- Find the area between the graph of $f(x) = 1/x^2$ for $x \in [1, 2]$ and the x axis in terms of known functions.

13. Find the area between $y = \sin x$ and $y = \cos x$ for $x \in [0, \frac{\pi}{4}]$.
14. Find the area between e^x and $\cos x$ for $x \in [0, 2\pi]$.
15. Find the area between the graphs of $y = \sin(2x)$ and $y = \cos(2x)$ for $x \in [0, 2\pi]$.
16. Find the area between the graphs of $y = \sin^2 x$, and the x axis, for $x \in [0, 2\pi]$.
17. Find the area between the graphs of $y = \cos^2 x$ and $y = \sin^2 x$ for $x \in [0, \pi/2]$.
18. Find the area between the graph of $f(x) = x^3/(2+3x^4)$ and the x axis for $x \in [0, 4]$.
19. Find the area between the graph of $f(x) = x^3 \sin(x^2) + 30$ and the x axis for $x \in [0, \pi]$.
20. Find the area between the graph of $f(x) = x \sin 2x$ and $x \cos x - 4$ for $x \in [0, \pi]$.
21. Find the area between $y = \frac{3x^3+9x^2+10x+6}{(x+1)^2(x^2+2x+2)}$ and the x axis for $x \in [0, 4]$.
22. Find the area between $y = \frac{5x^2+4+8x+2x^4+3x^3}{(2x+3)(x^2+1)}$ and the x axis for $x \in [0, 2]$.
23. Find the area between $\arctan(x)$ and $\ln x$ for $x \in [0, b]$ where $\arctan(b) - \ln(b) = 0$ and b is the first positive number for which this is so. $\arctan(b) - \ln(b) = 0$. You will need to use Newton's method or graphing on a calculator to find b .
24. Find the area between $y = e^x$ and $y = 2x + 1$ for $x > 0$. In order to do this, you have to find a solution to $e^x = 2x + 1$ and this will require a numerical procedure such as Newton's method or graphing and zooming on your calculator.
25. Find the area between $y = \ln x$ and $y = \sin x$ and $x \in [1, b]$ where b is such that the two functions are equal there. In order to do this, you have to find a solution to $\ln x = \sin x$ and this will require a numerical procedure such as Newton's method or graphing and zooming on your calculator.
26. A ten foot ladder leans against a wall. The bottom of the ladder starts out six feet from the wall when the bottom of the ladder begins to move away from the wall eventually causing the ladder to crash to the ground. A point on the ladder which is two feet down from the top has coordinates $(x, f(x))$. Find $f(x)$. Next find the area between the resulting curve traced out by this point and the ground.
27. Let $p > 1$. An inequality which is of major importance is

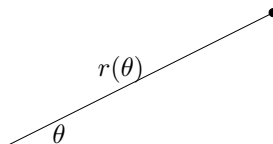
$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

where here q is defined by $1/p + 1/q = 1$. Establish this inequality by adding up areas in the following picture.



In the picture the right side of the inequality represents the sum of all the areas and the left side is the area of the rectangle determined by $(a, 0)$ and $(0, b)$.

28. Refer to Example 7.2.6 and the discussion before it for this problem. Suppose θ is an angle measured from the positive x axis and you draw the ray from $(0, 0)$ out a distance of $r(\theta)$. Thus as θ varies, the distance of the point varies also.



Suppose θ takes values in some interval $[a, b] \subseteq [0, 2\pi]$. Show the area between the curve traced out by this point and the two rays corresponding to $\theta = a$ and $\theta = b$ is given by

$$\frac{1}{2} \int_a^b (r(\theta))^2 d\theta$$

29. In Problem 28 suppose $r(\theta) = 1 - \cos(\theta)$ where $\theta \in [0, 2\pi]$. Sketch a graph of the region described in this problem and find its area.
30. Find the area between the graph of $y = 1/(1+x^2)$ and the x axis for $x \in [-R, R]$. Now take a limit of what you just obtained as $R \rightarrow \infty$.

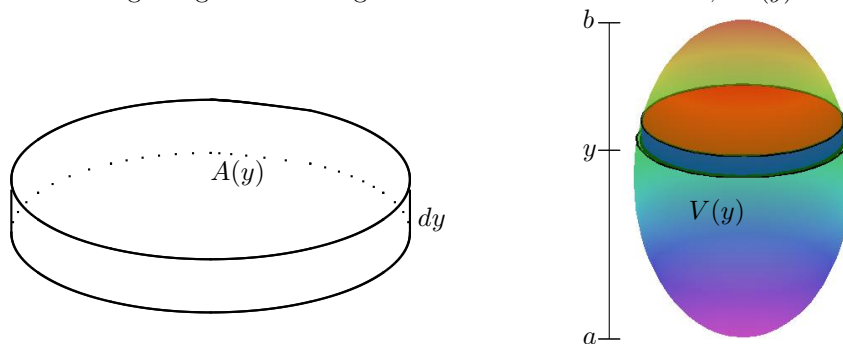
7.4 Volumes

7.4.1 Volumes Using Cross Sections

Imagine a line next to a three dimensional solid as shown in the next picture. For each y between a and b , let $A(y)$ denote the area of the cross section of the solid obtained by intersecting this solid with a plane through y perpendicular to the indicated line. Then

$$\frac{V(y+h) - V(y)}{h} \approx \frac{hA(y)}{h} = A(y)$$

the approximation getting better as h gets smaller. Thus in the limit, $V'(y) = A(y)$



and so the total volume of the solid between a and y , $V(y)$, satisfies the initial value problem,

$$\frac{dV}{dy} = A(y), \quad V(a) = 0.$$

The volume of the solid is $V(b)$. Here is a simple procedure which is based on this.

Procedure 7.4.1 Given a three dimensional solid S and a line as shown in the above picture, the volume of S which lies between $y = a$ and $y = b$ is obtained by finding

$$V \in \int_a^b A(y) dy$$

and computing $H(b) - H(a)$ where $A(y)$ is the area of the cross section at y .

Here is why. As described above, this volume is $V(b)$ where V is the solution to the initial value problem

$$V'(y) = A(y), \quad V(a) = 0.$$

Thus from Corollary 3.6.4 $V(y) = H(y) + C$ for some constant C and so

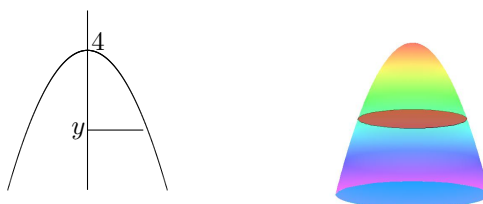
$$V(b) - V(a) = H(b) + C - (H(a) + C) = H(b) - H(a).$$

Example 7.4.2 The parabola $y = 4 - x^2$ for $y \geq 0$ is revolved about the y axis to form a solid of revolution. Find the volume of this solid of revolution.

As suggested above, one finds the area of a cross section at y , $A(y)$ and then solves $dV/dy = A(y)$, $V(0) = 0$. Then the volume of the shape is $V(4)$. At height y a cross section is a circle of radius $\sqrt{4 - y}$ and so the area of this cross section is $\pi(4 - y)$. Thus the initial value problem to solve is

$$\frac{dV}{dy} = \pi(4 - y), \quad V(0) = 0$$

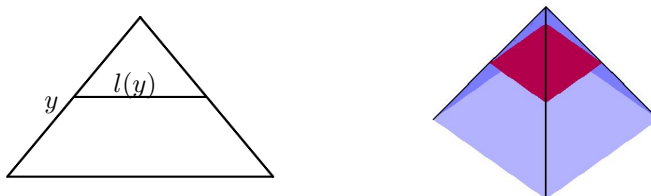
and the answer desired is $V(4)$. Below is a picture of the solid and a cross section of this solid.



Thus $V(y) = \pi(4y - \frac{1}{2}y^2)$ and so the volume is $V(4) = 8\pi$. Alternatively, you could simply find an antiderivative to $\pi(4 - y)$, $H(y) = \pi(4y - \frac{1}{2}y^2)$ and take $H(4) - H(0)$. If you wanted the volume of the region for $y \in [2, 3]$, this would be

$$H(3) - H(2) = \pi\left(4(3) - \frac{1}{2}(3)^2\right) - \pi\left(4(2) - \frac{1}{2}(2)^2\right) = \frac{3}{2}\pi$$

Example 7.4.3 Consider a pyramid which sits on a square base of length 500 feet and suppose the pyramid has height 300 feet. Find the volume of the pyramid in cubic feet.



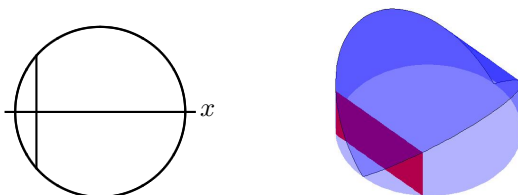
At height, y , it follows from similar triangles the length of one of the sides, $l(y)$, would satisfy $\frac{l(y)}{300 - y} = \frac{500}{300} = \frac{5}{3}$ and so $l(y) = \frac{5}{3}(300 - y)$. Therefore, the cross section at y has area $(\frac{5}{3}(300 - y))^2$ and so

$$\frac{dV}{dy} = \left(\frac{5}{3}(300 - y)\right)^2, \quad V(0) = 0,$$

and so $V(y) = -\frac{1}{5} \left(500 - \frac{5}{3}y\right)^3 + C$. Since $V(0) = 0$, the constant must equal $\frac{1}{5} (500)^3$. Therefore, the total volume of the pyramid is

$$-\frac{1}{5} \left(500 - \frac{5}{3}300\right)^3 + \frac{1}{5} (500)^3 = 25,000,000 \text{ cubic feet.}$$

Example 7.4.4 *The base of a solid is a circle of radius 10 meters in the xy plane. When this solid is cut with a plane which is perpendicular to the xy plane and the x axis, the result is a rectangle with height equal to one half the length of the side in the base. Find the volume of the resulting solid.*



Note the cross section in the solid intersects the base in the vertical line shown.

Reasoning as above for the building and letting $V(x)$ denote the volume between -10 and x , yields

$$\frac{dV}{dx} = A(x), \quad V(-10) = 0$$

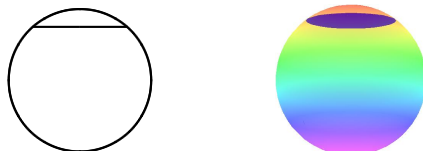
as an appropriate differential equation for this volume. Here $A(x)$ is the area of the surface resulting from the intersection of the plane with the solid. The length of the side of this surface in the base is $2\sqrt{100 - x^2}$ because the equation of the circle which bounds the base is $x^2 + y^2 = 100$. The height of the rectangle is $\sqrt{100 - x^2}$. Therefore, $A(x) = 2(100 - x^2)$ and so $V(x) = 200x - \frac{2}{3}x^3 + C$. To find C , use $V(-10) = 0$ and so $0 = 200(-10) + \frac{2}{3}1000 + C$. so $C = \frac{4000}{3}$. Therefore, the total volume is

$$V(10) = 200(10) - \frac{2}{3}(10)^3 + \frac{4000}{3} = \frac{8000}{3}$$

Alternatively, you could use the above procedure. The solid extends from $x = -10$ to $x = 10$. You need a function in $\int 2(100 - x^2) dx$. On such function is $200x - \frac{2}{3}x^3$. Then the answer is

$$200(10) - \frac{2}{3}(10)^3 - \left(200(-10) - \frac{2}{3}(-10)^3\right) = \frac{8000}{3}$$

Example 7.4.5 *Find the volume of a sphere of radius R .*



The sphere is obtained by revolving a disk of radius R about the y axis. Thus the radius of the cross section at height y would be $\sqrt{R^2 - y^2}$ and so the area of this cross section is $\pi(R^2 - y^2)$. Therefore,

$$\frac{dV}{dy} = \pi(R^2 - y^2), \quad V(-R) = 0$$

and so $V(y) = \pi \left(R^2 y - \frac{1}{3} y^3 \right) + C$. Now since $V(-R) = 0$, $C = \pi \left(-\frac{R^3}{3} + R^3 \right)$ and so

$$V(R) = \pi \left(R^3 - \frac{1}{3} (R)^3 \right) + \pi \left(-\frac{R^3}{3} + R^3 \right) = \frac{4}{3} \pi R^3$$

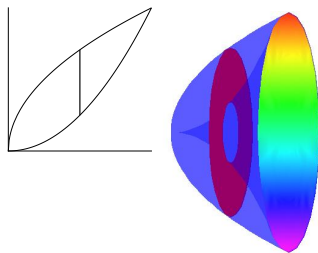
Using the procedure,

$$\pi \left(R^2 y - \frac{1}{3} y^3 \right) \in \int \pi (R^2 - y^2) dy$$

and so the answer is

$$\pi \left(R^2 (R) - \frac{1}{3} (R)^3 \right) - \pi \left(R^2 (-R) - \frac{1}{3} (-R)^3 \right) = \frac{4}{3} \pi R^3.$$

Example 7.4.6 The bounded region in the first quadrant ($x, y \geq 0$) determined by the graphs of $y = x^2$ and $y = \sqrt{x}$ is revolved about the x axis to form a solid in three dimensions. Find the volume of this solid.



As illustrated in the above picture, for each x , the cross sections consist of the area between two concentric circles, the smaller one having radius x^2 and the larger having radius $\sqrt{x}$. Note that for $0 < x < 1$, $x^2 < \sqrt{x}$. The area of this “washer” is $\pi (\sqrt{x})^2 - \pi x^4 = \pi (x - x^4) = A(x)$. Therefore, the appropriate differential equation and initial value is

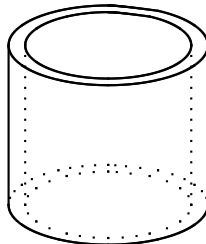
$$\frac{dV}{dx} = \pi (x - x^4), \quad V(0) = 0$$

and the answer is $V(1)$. From the differential equation and initial condition, $V(x) = \pi \left(\frac{1}{2} x^2 - \frac{1}{5} x^5 \right) + C$ since $V(0) = 0$, the constant $C = 0$. Therefore, the volume is

$$V(1) = \pi \left(\frac{1}{2} - \frac{1}{5} \right) = \frac{3}{10} \pi$$

7.4.2 Volumes Using Shells

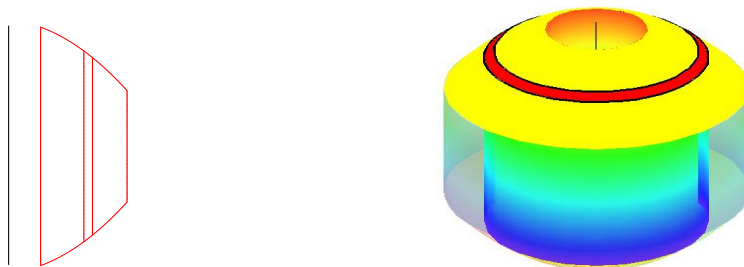
There is another way to find some volumes without using cross sections. This method involves the notion of shells. Consider the following picture of a circular shell.



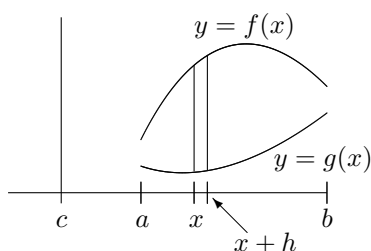
In this picture the radius of the inner circle will be r and the radius of the outer circle will be $r + \Delta r$ while the height of the shell is H . Therefore, the volume of the shell is the difference in the volumes of the two cylinders,

$$\pi(r + \Delta r)^2 H - \pi r^2 h = 2\pi H r (\Delta r) + \pi H (\Delta r)^2. \quad (7.3)$$

In the following picture, there is a 2 dimensional region in the xy plane which is revolved about a line. That which results is on the right. Notice the way the thin approximate rectangle on the left results in a shell in the solid of revolution on the right. Also note how the hole in the solid results from the gap between the line and the region in the plane.



To be more specific, consider the problem of revolving the region between $y = f(x)$ and $y = g(x)$ for $x \in [a, b]$ about the line $x = c$ for $c < a$. The following picture is descriptive of the situation.



Let $V(x)$ denote the volume of the solid which results from revolving the region between the graphs of f and g above the interval, $[a, x]$ about the line $x = c$. Thus $V(x + h) - V(x)$ equals the volume which results from revolving the region between the graphs of f and g which is also between the two vertical lines shown in the above picture. This results in a solid which is very nearly a circular shell like the one shown above and the approximation gets better as h decreases to zero. Therefore,

$$\begin{aligned} V'(x) &= \lim_{h \rightarrow 0} \frac{V(x + h) - V(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{2\pi |f(x) - g(x)| (x - c) h + \pi |f(x) - g(x)| h^2}{h} \\ &= 2\pi |f(x) - g(x)| (x - c). \end{aligned} \quad (7.4)$$

Also, $V(a) = 0$ and this means that to find the volume of revolution it suffices to solve the initial value problem,

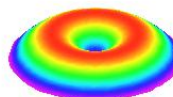
$$\frac{dV}{dx} = 2\pi |f(x) - g(x)| (x - c), \quad V(a) = 0 \quad (7.5)$$

and the volume of revolution will equal $V(b)$. Note that in the above formula, it is not necessary to worry about which is larger, $f(x)$ or $g(x)$ because it is expressed in terms of the absolute value of their difference. However, in doing the computations necessary to solve a given problem, you typically will have to worry about which is larger. Always

draw a picture of the two dimensional region and the line to get some understanding for what is going on.

Procedure 7.4.7 To find the volume of the region between $y = f(x)$, $y = g(x)$ for x between a and b which is obtained by revolving about the line $x = c$ which is either to the left or to the right of the entire two dimensional region, find $H \in \int 2\pi |f(x) - g(x)| (x - c) dx$. Then compute $H(b) - H(a)$.

Example 7.4.8 Find the volume of the solid formed by revolving the region between $y = \sin(x)$ and the x axis for $x \in [0, \pi]$ about the y axis.



In this example, $c = 0$ and since $\sin(x) \geq 0$ for $x \in [0, \pi]$, the initial value problem is

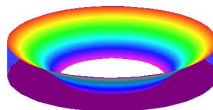
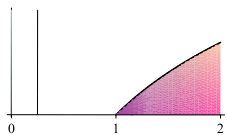
$$\frac{dV}{dx} = 2\pi x \sin(x), \quad V(0) = 0.$$

Then using integration by parts,

$$V(x) = 2\pi (\sin x - x \cos x) + C$$

and from the initial condition, $C = 0$. Therefore, the volume is $V(\pi) = 2\pi^2$. Alternatively, just find $H \in \int 2\pi x \sin(x)$ and compute $H(\pi) - H(0)$.

Example 7.4.9 The region between $y = \ln x$ and the x axis which lies between $x = 1$ and $x = 2$ is revolved about the line $x = 1/4$. Find the volume of the resulting solid of revolution.



Using the procedure described above, $\int 2\pi (x - \frac{1}{4}) \ln(x) dx = \pi x^2 \ln x - \frac{1}{2} \pi x^2 - \frac{1}{2} \pi x \ln x + \frac{1}{2} \pi x + C$. Then since $\ln 1 = 0$, the volume is

$$\pi (2)^2 \ln 2 - \frac{1}{2} \pi (2)^2 - \frac{1}{2} \pi 2 \ln 2 + \frac{1}{2} \pi 2 - \left(-\frac{1}{2} \pi + \frac{1}{2} \pi \right) = 3\pi \ln 2 - \pi$$

Example 7.4.10 Find the volume of the solid formed by revolving the region between $y = \sin x$, $y = \cos x$ for $x \in [0, \pi/4]$ about the line $x = -4$



In this example, $\cos x > \sin x$ for $x \in [0, \pi/4]$ and so using the above procedure,

$$\int 2\pi (x + 4) (\cos x - \sin x) = 2 (5 \cos x + x \sin x + 3 \sin x + x \cos x) \pi + C$$

and so the volume is

$$\begin{aligned} & 2 \left(5 \cos \left(\frac{\pi}{4} \right) + \frac{\pi}{4} \sin \left(\frac{\pi}{4} \right) + 3 \sin \left(\frac{\pi}{4} \right) + \frac{\pi}{4} \cos \left(\frac{\pi}{4} \right) \right) \pi - 2(5) \pi \\ &= 2 \left(4\sqrt{2} + \frac{1}{4} \pi \sqrt{2} \right) \pi - 10\pi \end{aligned}$$

Example 7.4.11 Find the volume of a ball of radius R using the method of shells.

You draw the picture. In this case the ball is obtained by revolving the region between $y = \sqrt{R^2 - x^2}$, $y = -\sqrt{R^2 - x^2}$ for $x \in [0, R]$ about the y axis. Thus, from the procedure,

$$\int 2\pi x \left(\sqrt{R^2 - x^2} - \left(-\sqrt{R^2 - x^2} \right) \right) dx = -\frac{4}{3} \left(\sqrt{R^2 - x^2} \right)^3 \pi$$

and so the volume is

$$-\frac{4}{3} \left(\sqrt{R^2 - R^2} \right)^3 \pi - \left(-\frac{4}{3} \left(\sqrt{R^2 - 0^2} \right)^3 \pi \right) = \frac{4}{3} R^3 \pi$$

the same formula obtained earlier using the method of cross sections.

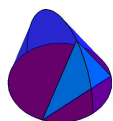
7.5 Exercises

1. The equation of an ellipse is $\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$. Sketch the graph of this in the case where $a = 2$ and $b = 3$. Now find the volume of the solid obtained by revolving this ellipse about the y axis. What is the volume if it is revolved about the x axis? What is the volume of the solid obtained by revolving it about the line $x = -2a$?
2. A sphere of radius R has a hole drilled through a diameter which is centered at the diameter and of radius $r < R$. Find the volume of what is left after the hole has been drilled. What is the volume of the material which was taken out?
3. Show the volume of a right circular cone is $(1/3) \times \text{area of the base} \times \text{height}$.



4. Let R be a region in the xy plane of area A and consider the cone formed by fixing a point in space h units above R and taking the union of all lines starting at this point which end in R . Show that under these general conditions, the volume of the cone is $(1/3) \times A \times h$. **Hint:** The cross sections at height y look just like R but shrunk. Argue that the area at height y , denoted by $A(y)$ is simply $A(y) = A \frac{(h-y)^2}{h^2}$.

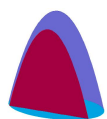
5. Find the volume of the solid which results by revolving $y = \sin x$ about the x axis for $x \in [0, \pi]$.



6. A circle of radius r in the xy plane is the base of a solid which has the property that cross sections perpendicular to the x axis are equilateral triangles. Find the volume of this solid.

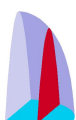
7. The region between $y = x^2$ and $y = x^3$ for $x \in [0, 1]$ is revolved about the y axis. Find the volume of the resulting solid using the method of cross sections. Now find it using the method of shells.

8. A square having each side equal to r in the xy plane is the base of a solid which has the property that cross sections perpendicular to the x axis are equilateral triangles. Find the volume of this solid.



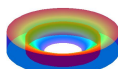
9. The ellipse, $\frac{x^2}{4} + \frac{y^2}{9} \leq 1$ is the base of a solid which has the property that cross sections perpendicular to the x axis are parabolas of height equal to the width of the ellipse. What is the volume of this solid?

10. The area between $y = \sqrt{x}$ and $x = 4$ is the base of a solid which has the property that cross sections perpendicular to the x axis are parabolas of height equal to the width of the base. What is the volume of this solid?
11. The circle $x^2 + y^2 \leq 4$ is the base of a solid which has the property that cross sections perpendicular to the y axis are isosceles right triangles having one leg equal to the width of the base. Find the volume of the solid.
12. The circle $x^2 + y^2 \leq 4$ is the base of a solid which has the property that cross sections perpendicular to the y axis are isosceles right triangles having the hypotenuse equal to the width of the base. Find the volume of the solid.



13. The part of the ellipse $\frac{x^2}{4} + \frac{y^2}{9} \leq 1$ which is in the first quadrant is the base of a solid which has the property that cross sections perpendicular to the x axis are parabolas of height equal to the width of the base. What is the volume of this solid?
14. The ellipse, $x^2 + \frac{y^2}{9} \leq 1$ is the base of a solid which has the property that cross sections perpendicular to the x axis are circles having a diameter perpendicular to the x axis and equal to the width of the ellipse. What is the volume of this solid? This looks like a football.
15. The bounded region between $y = x^2$ and $y = x$ is revolved about the x axis. Find the volume of the solid which results.

16. The region between $y = \ln x$, and the x axis for $x \in [1, 3]$ is revolved about the y axis. What is the volume of the resulting solid?
17. The region between $y = x^2, y = 1$ for $x \in [0, 1]$ is revolved about the line, $x = -4$. Find the volume of the solid which results.
18. The region between $y = \arctan(x)$, and the x axis for $x \in [0, 2]$ is revolved about the y axis. Find the volume of the resulting solid.



19. The region between $y = \arctan(x)$, and the x axis for $x \in [0, 2]$ is revolved about the line $x = -1$. Find the volume of the resulting solid.
20. The region between $y = \sin(x)$, and the x axis for $x \in [0, \pi]$ is revolved about the y axis. Find the volume of the resulting solid.
21. The region between $y = \sin(x)$, and the x axis for $x \in [0, \pi]$ is revolved about the line $x = -1$. Find the volume of the resulting solid.

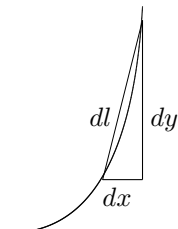


22. The region between $y = \sin(x)$, and the x axis for $x \in [0, \pi]$ is revolved about the line $x = 5$. Find the volume of the resulting solid.
23. The region between $y = \sin(x)$, and the x axis for $x \in [0, \pi]$ is revolved about the line $y = 2$. Find the volume of the resulting solid.
24. The region between $y = 1 + \sin x$ and the x axis for $x \in [0, 2\pi]$ is revolved about the y axis. Find the volume of the solid which results.
25. The region between $y = 2 + \sin 3x$ and the x axis for $x \in [0, \pi/3]$ is revolved about the line $x = -1$. Find the volume of the solid which results.
26. The region between $y = x^3 - x$ and the x axis for $x \geq 0$ is revolved about the line $x = -1$. Find the volume of the solid which results.
27. The region between $y = \sin 2x$ and the x axis for $x \in [0, \pi]$ is revolved about the line $x = -1$. Find the volume of the solid which results.

7.6 Lengths Of Curves And Areas Of Surfaces Of Revolution

7.6.1 Lengths

The same techniques can be used to compute lengths of the graph of a function, $y = f(x)$. Consider the following picture.



which depicts a small right triangle attached as shown to the graph of a function, $y = f(x)$ for $x \in [a, b]$. If the triangle is small enough, this shows the length of the curve joined by the hypotenuse of the right triangle is essentially equal to the length of the hypotenuse. Thus, $(dl)^2 = (dx)^2 + (dy)^2$ and dividing by $(dx)^2$ yields

$$\frac{dl}{dx} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2} = \sqrt{1 + f'(x)^2}, l(a) = 0 \quad (7.6)$$

as an initial value problem for the function, $l(x)$ which gives the length of this curve on $[a, x]$.

This definition gives the right answer for the length of a straight line. To see this, consider a straight line through the points (a, b) and (c, d) where $a < c$. Then the right answer is given by the Pythagorean theorem or distance formula and is $\sqrt{(a - c)^2 + (d - b)^2}$. What is obtained from the above initial value problem? The equation of the line is $f(x) = b + \left(\frac{d-b}{c-a}\right)(x - a)$ and so $f'(x) = \left(\frac{d-b}{c-a}\right)$. Therefore, letting l denote the arc length function,

$$\frac{dl}{dx} = \sqrt{1 + \left(\frac{d-b}{c-a}\right)^2}, l(a) = 0.$$

Thus $l(x) = \sqrt{1 + \left(\frac{d-b}{c-a}\right)^2} (x - a)$ and in particular, $l(c)$, the length of the line is given by

$$\sqrt{1 + \left(\frac{d-b}{c-a}\right)^2} (c - a) = \sqrt{(a - c)^2 + (b - c)^2}$$

as hoped. Thus this differential equation gives the right answer in the familiar cases but it also can be used to find lengths for more general curves than straight lines.

Thus as before, the following procedure can be used.

Procedure 7.6.1 *To find the length of the graph of the function $y = f(x)$ for $x \in [a, b]$, find an antiderivative $F \in \int \sqrt{1 + f'(x)^2} dx$. Then compute $F(b) - F(a)$ to get the length.*

The reason this works is that as above, if $l(x)$ gives the length for $x \in [a, x]$, then it satisfies the initial value problem 7.6. By Corollary 3.6.4 $F(x) = l(x) + C$ for some

C. Therefore,

$$F(b) - F(a) = l(b) + C - (l(a) + C) = l(b)$$

which equals the length.

Here is another familiar example.

Example 7.6.2 Find the length of the part of the circle having radius r which is between the points $(0, r)$ and $(\frac{\sqrt{2}}{2}r, \frac{\sqrt{2}}{2}r)$.

Here the function is $f(x) = \sqrt{r^2 - x^2}$ and so $f'(x) = -x/\sqrt{r^2 - x^2}$. Therefore, the differential equation is

$$\frac{dl}{dx} = \sqrt{1 + \frac{x^2}{(r^2 - x^2)}} = \sqrt{\frac{r^2}{r^2 - x^2}} = \frac{r}{\sqrt{r^2 - x^2}}.$$

Therefore, l is an antiderivative of this last function. Using a trig substitution, $x = r \sin \theta$, it follows $dx = r \cos(\theta) d\theta$ and so

$$\begin{aligned} \int \frac{r}{\sqrt{r^2 - x^2}} dx &= \int \frac{1}{\sqrt{1 - \sin^2 \theta}} r \cos(\theta) d\theta \\ &= r \int d\theta = r\theta + C \end{aligned}$$

Hence changing back to the variable x it follows $l(x) = r \arcsin(\frac{x}{r}) + C$. Then by Procedure 7.6.1, the length is

$$\begin{aligned} &r \arcsin\left(\frac{r}{r}\right) - r \arcsin\left(\frac{1}{r} \frac{\sqrt{2}}{2} r\right) \\ &= r \frac{\pi}{2} - r \frac{\pi}{4} = r \frac{\pi}{4}. \end{aligned}$$

Note this gives the length of one eighth of the circle and so from this the length of the whole circle should be $2r\pi$. Here is another example

Example 7.6.3 Find the length of the graph of $y = x^2$ between $x = 0$ and $x = 1$.

Here $f'(x) = 2x$ and so the initial value problem to be solved is

$$\frac{dl}{dx} = \sqrt{1 + 4x^2}, \quad l(0) = 0.$$

Thus, getting the exact answer depends on finding

$$\int \sqrt{1 + 4x^2} dx.$$

Use the trig. substitution, $2x = \tan u$ so $dx = \frac{1}{2} (\sec^2 u) du$. Therefore, making this substitution and using 6.7 on Page 158,

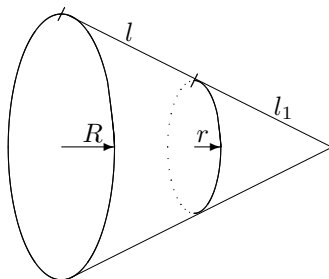
$$\begin{aligned} \int \sqrt{1 + 4x^2} dx &= \frac{1}{2} \int (\sec^3 u) du \\ &= \frac{1}{4} (\tan u) (\sec u) + \frac{1}{4} \ln |\sec u + \tan u| + C \\ &= \frac{1}{4} (2x) (\sqrt{1 + 4x^2}) + \frac{1}{4} \ln |2x + \sqrt{1 + 4x^2}| + C \end{aligned}$$

and since $l(0) = 0$ it must be the case that $C = 0$ and so the desired length is

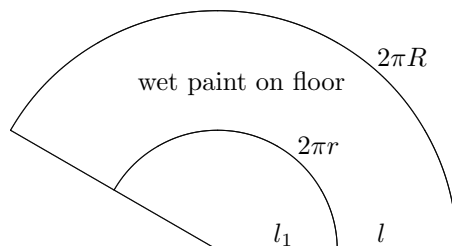
$$l(1) = \frac{1}{2} \sqrt{5} + \frac{1}{4} \ln |2 + \sqrt{5}|$$

7.6.2 Surfaces Of Revolution

The problem of finding the surface area of a solid of revolution is closely related to that of finding the length of a graph. First consider the following picture of the frustum of a cone in which it is desired to find the lateral surface area. In this picture, the frustum of the cone is the left part which has an l next to it and the lateral surface area is this part of the area of the cone.



To do this, imagine painting the sides and rolling the shape on the floor for exactly one revolution. The wet paint would make the following shape.



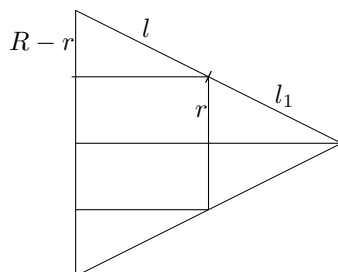
What would be the area of this wet paint? Its area would be the difference between the areas of the two sectors shown, one having radius l_1 and the other having radius $l + l_1$. Both of these have the same central angle equal to

$$\frac{2\pi R}{2\pi(l + l_1)} 2\pi = \frac{2\pi R}{l + l_1}.$$

Therefore, by Example 7.2.6 on Page 184, this area is

$$(l + l_1)^2 \frac{\pi R}{(l + l_1)} - l_1^2 \frac{\pi R}{(l + l_1)} = \pi R l \frac{l + 2l_1}{l + l_1}$$

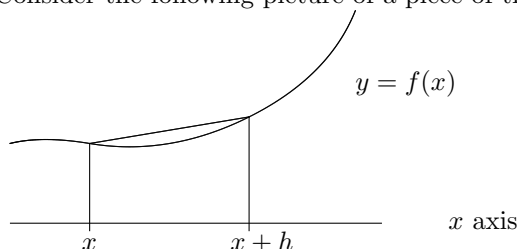
The view from the side is



and so by similar triangles, $l_1 = lr / (R - r)$. Therefore, substituting this into the above, the area of this frustum is

$$\pi R l \frac{l + 2\left(\frac{lr}{R-r}\right)}{l + \left(\frac{lr}{R-r}\right)} = \pi l (R + r) = 2\pi l \left(\frac{R + r}{2}\right).$$

Now consider a function, f , defined on an interval, $[a, b]$ and suppose it is desired to find the area of the surface which results when the graph of this function is revolved about the x axis. Consider the following picture of a piece of this graph.



Let $A(x)$ denote the area which results from revolving the graph of the function restricted to $[a, x]$ about the x axis. Then from the above formula for the area of a frustum,

$$\frac{A(x+h) - A(x)}{h} \approx 2\pi \frac{1}{h} \sqrt{h^2 + (f(x+h) - f(x))^2} \left(\frac{f(x+h) + f(x)}{2} \right)$$

where $\approx$ denotes that these are close to being equal and the approximation gets increasingly good as $h \rightarrow 0$. Therefore, rewriting this a little yields

$$\frac{A(x+h) - A(x)}{h} \approx 2\pi \sqrt{1 + \left(\frac{f(x+h) - f(x)}{h} \right)^2} \left(\frac{f(x+h) + f(x)}{2} \right)$$

Therefore, taking the limit as $h \rightarrow 0$, and using $A(a) = 0$, this yields the following initial value problem for A which can be used to find the area of a surface of revolution.

$$A'(x) = 2\pi f(x) \sqrt{1 + f'(x)^2}, \quad A(a) = 0.$$

What would happen if you revolved about the y axis? I will leave it to you to verify this would lead to the initial value problem

$$A'(x) = 2\pi x \sqrt{1 + f'(x)^2}, \quad A(a) = 0.$$

As before, this results in the following simple procedure for finding the surface area of a surface of revolution.

Procedure 7.6.4 To find the surface area of a surface obtained by revolving the graph of $y = f(x)$ for $x \in [a, b]$ about the x axis, find $F \in \int 2\pi f(x) \sqrt{1 + f'(x)^2} dx$ and compute $F(b) - F(a)$. Similarly, to get the area of the graph rotated about the y axis, find $G \in \int 2\pi x \sqrt{1 + f'(x)^2} dx$ and compute $G(b) - G(a)$.

Example 7.6.5 Find the surface area of the surface obtained by revolving the function, $y = r$ for $x \in [a, b]$ about the x axis. Of course this is just the cylinder of radius r and height $b - a$ so this area should equal $2\pi r(b - a)$. (Imagine painting it and rolling it on the floor and then taking the area of the rectangle which results.)

Using the above initial value problem, solve

$$A'(x) = 2\pi r \sqrt{1 + 0^2}, \quad A(a) = 0.$$

The solution is $A(x) = 2\pi r(x - a)$. Therefore, $A(b) = 2\pi r(b - a)$ as expected.

Example 7.6.6 Find the surface area of a sphere of radius r .

Here the function involved is $f(x) = \sqrt{r^2 - x^2}$ for $x \in [-r, r]$ and it is to be revolved about the x axis. In this case

$$f'(x) = \frac{-x}{\sqrt{r^2 - x^2}}$$

and so the initial value problem is of the form

$$A'(x) = 2\pi \overbrace{\sqrt{r^2 - x^2}}^{=r} \sqrt{1 + \frac{x^2}{r^2 - x^2}}, \quad A(-r) = 0$$

Thus, simplifying the above yields $A'(x) = 2\pi r$ and so $A(x) = 2\pi r x + C$ and since $A(-r) = 0$, it follows that $C = 2\pi r^2$. Therefore, the surface area equals $A(r) = 2\pi r^2 + 2\pi r^2 = 4\pi r^2$. Using the procedure, you could also simply compute $2\pi r(r) - 2\pi r(-r) = 4\pi r^2$.

7.7 Exercises

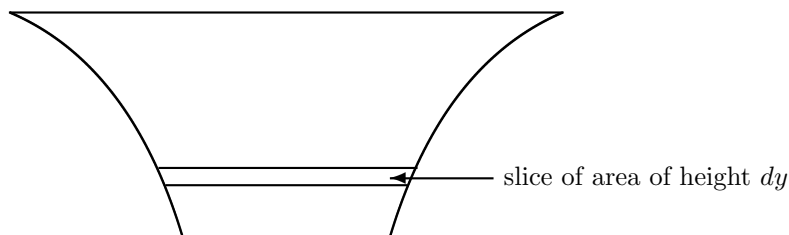
1. Find the length of the graph of $y = \ln(\cos x)$ for $x \in [0, \pi/4]$.
2. The curve defined by $y = \ln(\cos x)$ for $x \in [0, 1]$ is revolved about the y axis. Find an integral for the area of the surface of revolution.
3. Find the length of the graph of $y = x^{1/2} - \frac{x^{3/2}}{3}$ for $x \in [1, 3]$.
4. The graph of the function, $y = x^3$ is revolved about the x axis for $x \in [0, 1]$. Find the area of the resulting surface of revolution.
5. The graph of the function, $y = x^3$ is revolved about the y axis for $x \in [0, 1]$. Find the area of the resulting surface of revolution. **Hint:** Formulate this in terms of x and use a change of variables.
6. The graph of the function, $y = \ln x$ is revolved about the y axis for $x \in [1, 2]$. Find the area of the resulting surface of revolution. **Hint:** Consider x as a function of y .
7. The graph of the function, $y = \ln x$ is revolved about the x axis for $x \in [1, 2]$. Find the area of the resulting surface of revolution. If you can't do the integral, set it up.
8. Find the length of $y = \cosh(x)$ for $x \in [0, 1]$.
9. Find the length of $y = 2x^2 - \frac{1}{16} \ln x$ for $x \in [1, 2]$.
10. The curve defined by $y = 2x^2 - \frac{1}{16} \ln x$ for $x \in [1, 2]$ is revolved about the y axis. Find the area of the resulting surface of revolution.
11. Find the length of $y = x^2 - \frac{1}{8} \ln x$ for $x \in [1, 2]$.
12. The curve defined by $y = x^2 - \frac{1}{8} \ln x$ for $x \in [1, 2]$ is revolved about the y axis. Find the area of the resulting surface of revolution.
13. The curve defined by $y = \cosh(x)$ for $x \in [0, 1]$ is revolved about the x axis. Find the area of the resulting surface of revolution.
14. The curve defined by $y = \cosh(x)$ for $x \in [0, 1]$ is revolved about the line $y = -3$. Find the area of the resulting surface of revolution.

15. For a a positive real number, find the length of $y = \frac{ax^2}{2} - \frac{1}{4a} \ln x$ for $x \in [1, 2]$. Of course your answer should depend on a .
16. The graph of the function, $y = x^2$ for $x \in [0, 1]$ is revolved about the x axis. Find the area of the surface of revolution.
17. The graph of the function, $y = \sqrt{x}$ for $x \in [0, 1]$ is revolved about the y axis. Find the area of the surface of revolution. **Hint:** Switch x and y and then use the previous problem.
18. The graph of the function, $y = x^{1/2} - \frac{x^{3/2}}{3}$ is revolved about the y axis. Find the area of the surface of revolution if $x \in [0, 2]$.
19. The graph of the function, $y = \sinh x$ for $x \in [0, 1]$ is revolved about the x axis. Find the area of the surface of revolution.
20. * The ellipse, $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ is revolved about the x axis. Find the area of the surface of revolution.
21. Find the length of the graph of $y = \frac{2}{3}(x-1)^{3/2}$ for $x \in [2, 3]$.
22. The curve defined by $y = \frac{2}{3}(x-1)^{3/2}$ for $x \in [1, 2]$ is revolved about the y axis. Find the area of the resulting surface of revolution.
23. Suppose $f'(x) = \sqrt{\sec^2 x - 1}$ and $f(0) = 0$. Find the length of the graph of $y = f(x)$ for $x \in [0, 1]$.
24. The curve defined by $y = f(x)$ for $x \in [0, \pi]$ is revolved about the y axis where $f'(x) = \sqrt{(2 + \sin x)^2 - 1}$, $f(0) = 1$. Find the area of the resulting surface of revolution.
25. Revolve $y = 1/x$ for $x \in [1, R]$ about the x axis. Find the area of this surface of revolution. Now show the limit of what you got as $R \rightarrow \infty$ does not exist. Next find the volume of this solid of revolution. Show the limit of this as $R \rightarrow \infty$ is finite. This infinite solid has infinite area but finite volume.

7.8 Force On A Dam And Work

7.8.1 Force On A Dam

Imagine you are a fish swimming in a lake behind a dam and you are interested in the total force acting on the dam. The following picture is what you would see.



The reason you would be interested in that long thin slice of area having essentially the same depth, say at y feet is because the pressure in the water at that depth is

constant and equals $62.5y$ pounds per square foot¹. Therefore, the total force the water exerts on the long thin slice is

$$dF = 62.5yL(y) dy$$

where $L(y)$ denotes the length of the slice. Therefore, the total force on the dam up to depth y is obtained as a solution to the initial value problem,

$$\frac{dF}{dy} = 62.5yL(y), \quad F(0) = 0.$$

Example 7.8.1 Suppose the width of a dam at depth y feet equals $L(y) = 1000 - y$ and its depth is 500 feet. Find the total force in pounds exerted on the dam.

From the above, this is obtained as the solution to the initial value problem

$$\frac{dF}{dy} = 62.5y(1000 - y), \quad F(0) = 0$$

which is $F(y) = -20.83y^3 + 31250y^2$. The total force on the dam would be

$$F(500) = -20.83(500)^3 + 31250(500)^2 = 5,208,750,000.0$$

pounds. In tons this is 2,604,375. That is a lot of force.

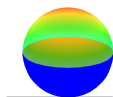
7.8.2 Work

Now suppose you are pumping water from a tank of depth d to a height of H feet above the top of the water in the tank. Suppose also that at depth y below the surface, the area of a cross section having constant depth is $A(y)$. The total weight of a slice of water having thickness dy at this depth is $62.5A(y)dy$ and the pump needs to lift this weight a distance of $y + H$ feet. Therefore, the work done is $dW = (y + H)62.5A(y)dy$. An initial value problem for the work done to pump the water down to a depth of y feet would be

$$\frac{dW}{dy} = (y + H)62.5A(y), \quad W(0) = 0.$$

The reason for the initial condition is that the pump has done no work to pump no water. If the weight of the fluid per cubic foot were different than 62.5 you would do the same things but replace the number.

Example 7.8.2 A spherical storage tank sitting on the ground having radius 30 feet is half filled with a fluid which weighs 50 pounds per cubic foot. How much work is done to pump this fluid to a height of 100 feet?



Letting r denote the radius of a cross section y feet below the level of the fluid, $r^2 + y^2 = 900$. Therefore,

$$r = \sqrt{900 - y^2}.$$

¹Later on a nice result on hydrostatic pressure will be presented which will verify this assertion. Here 62.5 is the weight in pounds of a cubic foot of water. If you like, think of a column of water of height y having base area equal to 1 square foot. Then the total force acting on this base area would be $62.5 \times y$ pounds. It turns out that since the water is incompressible, this is roughly the force acting on a square foot situated in any manner at this depth.

It follows the area of the cross section at depth y is $\pi(900 - y^2)$. Here $H = 70$ and so the initial value problem to solve is

$$\frac{dW}{dy} = (y + 70) 50\pi (900 - y^2), \quad W(0) = 0.$$

Therefore, $W(y) = 50\pi \left(-\frac{1}{4}y^4 - \frac{70}{3}y^3 + 450y^2 + 63\,000y\right)$ and the total work in foot pounds equals

$$W(30) = 50\pi \left(-\frac{1}{4}(30)^4 - \frac{70}{3}(30)^3 + 450(30)^2 + 63\,000(30)\right) = 73,125,000\pi$$

In general, the work done by a constant force in a straight line equals the product of the force times the distance over which it acts. This is an over simplification and it will be made more correct later. If the force is varying with respect to position, then you have to use calculus to compute the work. For now, consider the following examples.

Example 7.8.3 *A 500 pound safe is lifted 10 feet. How much work is done?*

The work is $500 \times 10 = 5000$ foot pounds.

Example 7.8.4 *The force needed to stretch a spring x feet past its equilibrium position is kx . This is known as Hooke's law and is a good approximation as long as the spring is not stretched too far. If $k = 3$, how much work is needed to stretch the spring a distance of 2 feet beyond its equilibrium position? The constant, k is called the spring constant. Different springs would have different spring constants. The units on k are pounds/foot.*

This is a case of a variable force. To stretch the spring from x to $x+dx$ requires $3xdx$ foot pounds of work. Therefore, letting W denote the work up till time x , $dW = 3xdx$ and so the initial value problem is

$$\frac{dW}{dx} = 3x, \quad W(0) = 0.$$

Thus $W(2) = \frac{3}{2}(2^2) = 6$ foot pounds because an antiderivative for $3x$ is $\frac{3}{2}x^2$.

7.9 Exercises

1. The main span of the Portage Lake lift bridge² weighs 4,400,000 pounds. How much work is done in raising this main span to a height of 100 feet?
2. A cylindrical storage tank having radius 20 feet and length 40 feet is filled with a fluid which weighs 50 pounds per cubic foot. This tank is lying on its side on the ground. Find the total force acting on the ends of the tank by the fluid.
3. Suppose the tank in Problem 2 is filled to a depth of 8 feet. Find an integral for the work needed to pump the fluid to a height of 50 feet.
4. A conical hole is filled with water which has weight 62.5 pounds per cubic feet. If the depth of the hole is 20 feet and the radius of the hole is 10 feet, how much work is needed to pump the water to a height of 10 feet above the ground?

²This is the heaviest lift bridge in the world. It joins the towns of Houghton and Hancock in the upper peninsula of Michigan spanning Portage lake. It provides 250 feet of clear channel for ships and can provide as much as 100 feet of vertical clearance. The lifting machinery is at the top of two massive towers 180 feet above the water. Aided by 1,100 ton counter weights on each tower, sixteen foot gears pull on 42 cables to raise the bridge. This usually creates impressive traffic jams on either side of the lake. The motion up and down of this span is quite slow.

5. Suppose the spring constant is 2 pounds per foot. Find the work needed to stretch the spring 3 feet beyond equilibrium.
6. A 20 foot chain lies on the ground. It weighs 5 pounds per foot. How much work is done to lift one end of the chain to a height of 20 feet?
7. A dam 500 feet high has a width at depth y equal to $4000 - 2y$ feet. What is the total force on the dam if it is filled?
8. *When the bucket is filled with water it weighs 30 pounds and when empty it weighs 2 pounds and the person on top of a 100 foot building exerts a constant force of 40 pounds. The bucket is full at the bottom but leaks at the rate of .1 cubic feet per second. How much work does the person on the top of the building do in lifting the bucket to the top? Will the bucket be empty when it reaches the top? You can use Newton's law that force equals mass times acceleration. You can neglect the weight of the rope.
9. In the situation of the above problem, suppose the person on the top maintains a constant velocity of 1 foot per second and the bucket leaks at the rate of .1 pound per second. How much work does he do and is the bucket empty when it reaches the top?
10. A silo is 10 feet in diameter and at a height of 30 feet there is a hemispherical top. The silage weighs 10 pounds per cubic foot. How much work was done in filling it to the very top?
11. A cylindrical storage tank having radius 10 feet is filled with water to a depth of 20 feet. If the storage tank stands upright on its circular base, what is the total force the water exerts on the sides of the tank? **Hint:** The pressure in the water at depth y is $62.5y$ pounds per square foot.
12. A spherical storage tank having radius 10 feet is filled with water. What is the total force the water exerts on the storage tank? **Hint:** The pressure in the water at depth y is $62.5y$ consider the area corresponding to a slice at height y . This is a surface of revolution and you know how to deal with these. The area of this slice times the pressure gives the total force acting on it.
13. A water barrel which is 11 inches in radius and 34 inches high is filled with water. If it is standing on end, what is the total force acting on the circular sides of the barrel?
14. Find the total force acting on the circular sides of the cylinder in Problem 2.

Other Differential Equations*

These are equations in which more is involved than simply taking an antiderivative.

8.1 The Equation $y' + a(t)y = b(t)$

The homogeneous first order constant coefficient linear differential equation is a differential equation of the form

$$y' + ay = 0. \quad (8.1)$$

It is arguably the most important differential equation in existence. Generalizations of it include the entire subject of linear differential equations and even many of the most important partial differential equations occurring in applications.

Here is how to find the solutions to this equation. Multiply both sides of the equation by e^{at} . Then use the product and chain rules to verify that

$$e^{at}(y' + ay) = \frac{d}{dt}(e^{at}y) = 0.$$

Therefore, since the derivative of the function $t \rightarrow e^{at}y(t)$ equals zero, it follows this function must equal some constant, C . Consequently, $ye^{at} = C$ and so $y(t) = Ce^{-at}$. This shows that if there is a solution of the equation, $y' + ay = 0$, then it must be of the form Ce^{-at} for some constant, C . You should verify that every function of the form, $y(t) = Ce^{-at}$ is a solution of the above differential equation, showing this yields all solutions. This proves the following theorem.

Theorem 8.1.1 *The solutions to the equation, $y' + ay = 0$ consist of all functions of the form, Ce^{-at} where C is some constant.*

Example 8.1.2 *Radioactive substances decay in the following way. The rate of decay is proportional to the amount present. In other words, letting $A(t)$ denote the amount of the radioactive substance at time t , $A(t)$ satisfies the following initial value problem.*

$$A'(t) = -k^2 A(t), \quad A(0) = A_0$$

where A_0 is the initial amount of the substance. What is the solution to the initial value problem?

Write the differential equation as $A'(t) + k^2 A(t) = 0$. From Theorem 8.1.1 the solution is

$$A(t) = Ce^{-k^2 t}$$

and it only remains to find C . Letting $t = 0$, it follows $A_0 = A(0) = C$. Thus $A(t) = A_0 \exp(-k^2 t)$.

Now consider a slightly harder equation.

$$y' + a(t)y = b(t).$$

In the easier case, you multiplied both sides by e^{at} . In this case, you multiply both sides by $e^{A(t)}$ where $A'(t) = a(t)$. In other words, you find an antiderivative of $a(t)$ and multiply both sides of the equation by e raised to that function. (It will be shown later in Theorem 9.3.4 on Page 221 that such an A always exists provided a is continuous.) Thus

$$e^{A(t)}(y' + a(t)y) = e^{A(t)}b(t).$$

Now you notice that this becomes

$$\frac{d}{dt}(e^{A(t)}y) = e^{A(t)}b(t). \quad (8.2)$$

This follows from the chain rule.

$$\frac{d}{dt}(e^{A(t)}y) = A'(t)e^{A(t)}y + e^{A(t)}y' = e^{A(t)}(y' + a(t)y).$$

Then from 8.2,

$$e^{A(t)}y \in \int e^{A(t)}b(t) dt.$$

Therefore, to find the solution, you find a function in $\int e^{A(t)}b(t) dt$, say $F(t)$, and

$$e^{A(t)}y = F(t) + C$$

for some constant, C , so the solution is given by $y = e^{-A(t)}F(t) + e^{-A(t)}C$. This proves the following theorem.

Theorem 8.1.3 *The solutions to the equation, $y' + a(t)y = b(t)$ consist of all functions of the form*

$$y = e^{-A(t)}F(t) + e^{-A(t)}C$$

where $F(t) \in \int e^{A(t)}b(t) dt$ and C is a constant.

Example 8.1.4 *Find the solution to the initial value problem*

$$y' + 2ty = \sin(t)e^{-t^2}, \quad y(0) = 3.$$

Multiply both sides by e^{t^2} because $t^2 \in \int t dt$. Then $\frac{d}{dt}(e^{t^2}y) = \sin(t)$ and so $e^{t^2}y = -\cos(t) + C$. Hence the solution is of the form $y(t) = -\cos(t)e^{-t^2} + Ce^{-t^2}$. It only remains to choose C in such a way that the initial condition is satisfied. From the initial condition, $3 = y(0) = -1 + C$ and so $C = 4$. Therefore, the solution is $y = -\cos(t)e^{-t^2} + 4e^{-t^2}$. Now at this point, you should check and see if it works. It needs to solve both the initial condition and the differential equation.

Finally, here is a uniqueness theorem.

Theorem 8.1.5 *If $a(t)$ is a continuous function, there is at most one solution to the initial value problem, $y' + a(t)y = b(t)$, $y(r) = y_0$.*

Proof: If there were two solutions, y_1 , and y_2 , then letting $w = y_1 - y_2$, it follows $w' + a(t)w = 0$ and $w(r) = 0$. Then multiplying both sides of the differential equation by $e^{A(t)}$ where $A'(t) = a(t)$, (It will be shown later in Theorem 9.3.4 on Page 221 that such an A always exists provided a is continuous.) it follows

$$(e^{A(t)}w)' = 0$$

and so $e^{A(t)}w(t) = C$ for some constant, C . However, $w(r) = 0$ and so this constant can only be 0. Hence $w = 0$ and so $y_1 = y_2$.

8.2 Separable Differential Equations

Definition 8.2.1 *Separable differential equations are those which can be written in the form*

$$\frac{dy}{dx} = \frac{f(x)}{g(y)}.$$

The reason these are called separable is that if you formally cross multiply,

$$g(y) dy = f(x) dx$$

and the variables are “separated”. The x variables are on one side and the y variables are on the other.

Proposition 8.2.2 *If $G'(y) = g(y)$ and $F'(x) = f(x)$, then if the equation, $F(x) - G(y) = c$ specifies y as a differentiable function of x , then $x \rightarrow y(x)$ solves the separable differential equation*

$$\frac{dy}{dx} = \frac{f(x)}{g(y)}. \quad (8.3)$$

Proof: Differentiate both sides of $F(x) - G(y) = c$ with respect to x . Using the chain rule,

$$F'(x) - G'(y) \frac{dy}{dx} = 0.$$

Therefore, since $F'(x) = f(x)$ and $G'(y) = g(y)$, $f(x) = g(y) \frac{dy}{dx}$ which is equivalent to 8.3.

Example 8.2.3 *Find the solution to the initial value problem,*

$$y' = \frac{x}{y^2}, \quad y(0) = 1.$$

This is a separable equation and in fact, $y^2 dy = x dx$ so the solution to the differential equation is of the form

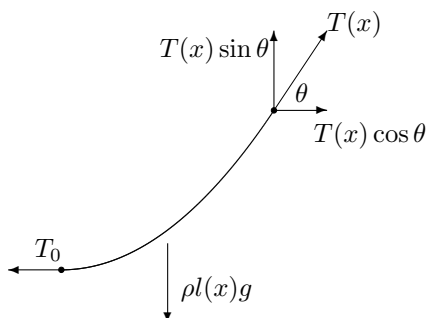
$$\frac{y^3}{3} - \frac{x^2}{2} = C \quad (8.4)$$

and it only remains to find the constant, C . To do this, you use the initial condition. Letting $x = 0$, it follows $\frac{1}{3} = C$ and so

$$\frac{y^3}{3} - \frac{x^2}{2} = \frac{1}{3}$$

Example 8.2.4 *What is the equation of a hanging chain?*

Consider the following picture of a portion of this chain.



In this picture, ρ denotes the density of the chain which is assumed to be constant and g is the acceleration due to gravity. $T(x)$ and T_0 represent the magnitude of the tension in the chain at x and at 0 respectively, as shown. Let the bottom of the chain be at the origin as shown. If this chain does not move, then all these forces acting on it must balance. In particular,

$$T(x) \sin \theta = l(x) \rho g, \quad T(x) \cos \theta = T_0.$$

Therefore, dividing these yields

$$\frac{\sin \theta}{\cos \theta} = l(x) \overbrace{\rho g / T_0}^{\equiv c}.$$

Now letting $y(x)$ denote the y coordinate of the hanging chain corresponding to x ,

$$\frac{\sin \theta}{\cos \theta} = \tan \theta = y'(x).$$

Therefore, this yields

$$y'(x) = cl(x).$$

Now differentiating both sides of the differential equation,

$$y''(x) = cl'(x) = c\sqrt{1 + y'(x)^2}$$

and so

$$\frac{y''(x)}{\sqrt{1 + y'(x)^2}} = c.$$

Let $z(x) = y'(x)$ so the above differential equation becomes

$$\frac{z'(x)}{\sqrt{1 + z^2}} = c.$$

Therefore, $\int \frac{z'(x)}{\sqrt{1 + z^2}} dx = cx + d$. Change the variable in the antiderivative letting $u = z(x)$ and this yields

$$\begin{aligned} \int \frac{z'(x)}{\sqrt{1 + z^2}} dx &= \int \frac{du}{\sqrt{1 + u^2}} = \sinh^{-1}(u) + C \\ &= \sinh^{-1}(z(x)) + C. \end{aligned}$$

by 5.10 on Page 128. Therefore, combining the constants of integration,

$$\sinh^{-1}(y'(x)) = cx + d$$

and so

$$y'(x) = \sinh(cx + d).$$

Therefore,

$$y(x) = \frac{1}{c} \cosh(cx + d) + k$$

where d and k are some constants and $c = \rho g / T_0$. Curves of this sort are called catenaries. Note these curves result from an assumption that the only forces acting on the chain are as shown.

8.3 Exercises

1. For x sufficiently large, let $f(x) = (\ln x)^{\ln x}$. Find $f'(x)$. How big does x need to be in order for this to make sense. **Hint:** You should have $\ln x > 0$.
2. In the hanging chain problem the picture and the derivation involved an assumption that at its lowest point, the chain was horizontal. Imagine lifting the end higher and higher and you will see this might not be the case in general. Can you modify the above derivation for the hanging chain to show that even in this case the chain will be in the form of a catenary?
3. Find the solution to the initial value problem,

$$y' = 1 + y^2, \quad y(0) = 0.$$

4. Verify that for any constant, C , the function, $y(t) = Ce^{-at}$ solves the differential equation, $y' + ay = 0$.
5. In Example 8.1.2 the half life is the time it takes for half of the original amount of the substance to decay. Find a formula for the half life assuming you know k^2 .
6. There are ten grams of a radioactive substance which is allowed to decay for five years. At the end of the five years there are 9.5 grams of the substance left. Find the half life of the substance. **Hint:** Use the given information to find k^2 and then use Problem 5.
7. The giant arch in St. Louis is in the form of an inverted catenary. Why?
8. Sometimes banks compound interest continuously. One way to think of this is to let the amount in the account satisfy the initial value problem,

$$A'(t) = rA(t), \quad A(0) = A_0$$

where here $A(t)$ is the amount at time t measured in years, r is the interest rate per year, and A_0 is the initial amount. Find $A(t)$ explicitly. If \$100 is placed in an account which is compounded continuously at 6% per year, how many years will it be before there is \$200 in the account? **Hint:** In this case, $r = .06$.

9. An object falling through the air experiences a force of gravity and air resistance. Later, it will be shown that this implies the velocity satisfies a differential equation of the form

$$v' = \frac{g}{m} - kv$$

where k is a positive constant and $\frac{g}{m}$ is also a positive constant. Find the solutions to this differential equation and determine what happens to v as t gets large. If you do it right you will find the terminal velocity.

10. Solve the initial value problem

$$A'(t) = rA(t) + 1 + \sin(t), \quad A(0) = 1$$

and describe its behavior as $t \rightarrow \infty$. Assume r is a positive constant. what r is a negative constant?

11. A population is growing at the rate of 11% per year. This means it satisfies the differential equation, $A' = .11A$. Find the time it takes for the population to double.

12. A substance is decaying at the rate of .01 per year. Find the half life of the substance.
13. The half live of a substance is 40 years. Find the rate of decay of the substance.
14. A sample of 4 grams of a radioactive substance has decayed to 3 grams in 5 days. Find the half life of the substance. Give your answer in terms of logarithms.
15. \$1000 is deposited in an account that earns interest at the rate of 5% per year compounded continuously. How much will be in the account after 10 years?
16. A sample of one ounce of water from a water supply is cultured in a Petri dish. After four hours, there are 3000.0 bacteria present and after six hours there are 7000 bacteria present. How many were present in the original sample?
17. Carbon 14 is a radioactive isotope of carbon and it is produced at a more or less constant rate in the earth's atmosphere by radiation from the sun. It also decays at a rate proportional to the amount present. Show this implies that, assuming this has been going on for billions of years, it is reasonable to assume the amount of Carbon 14 in the atmosphere is essentially constant over time. **Hint:** By assumption, if A is the amount of Carbon 14 in the atmosphere, $\frac{dA}{dt} = -k^2 A + r$ where k^2 is the constant of decay described above and r is the constant rate of production. Now show $A'(t) = Ce^{-k^2 t}$. Conclude $A(t) = D - (C/k^2)e^{-k^2 t}$. What happens to this over a long period of time?
18. The method of carbon dating is based on the result of Problem 17. When an animal or plant is alive it absorbs carbon from the atmosphere and so when it is living it has a known percentage of carbon 14. When it dies, it quits absorbing carbon and the carbon 14 begins to decay. After some time, t , the amount of carbon 14 can be measured and on this basis, the time since the death of the animal or plant can be estimated. Given the half life of carbon 14 is 5730 years and the amount of carbon 14 in a mummy is .3 what it was at the time of death, how long has it been since the mummy was alive? (To see how to come up with this figure for the half life, see Problem 6. You do experiments and take measurements over a smaller period of time.)
19. The half life of carbon 14 is known to be 5730 years. A certain tree stump is known to be 4000 years old. What percentage of the original carbon 14 should it contain?
20. One model for population growth is to assume the rate of growth is proportional to both the population and the difference between some maximum sustainable population and the population. The reason for this is that when the population gets large enough, there begin to be insufficient resources. Thus $\frac{dA}{dt} = kA(M - A)$, where k and M are positive constants. Show this is a separable differential equation and its solutions are of the form

$$A(t) = \frac{M}{1 + CM e^{-kMt}}$$

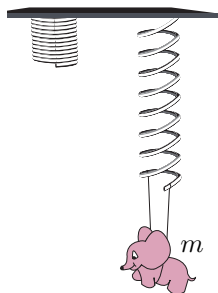
where C is a constant. Given three measurements of population at three equally spaced times, show how to predict the maximum sustainable population¹.

¹This has been done with the earth's population and the maximum sustainable population has been exceeded. Therefore, the model is far too simplistic for human population growth. However, it would work somewhat better for predicting the growth of things like bacteria.

21. Homogeneous differential equations are those which can be written in the form $y' = f\left(\frac{y}{x}\right)$. For example, $y' = \frac{x^2}{y^2+x^2}$. There is a trick to solving such equations. You define a new variable, $v = y/x$ and then write the differential equation in terms of v rather than y . Show that $xv' + v = y'$ and so the differential equation reduces to $xv' = f(v) - v$, a separable differential equation. Use the technique to solve $y' = 1 + 2\frac{y}{x}$, $y(1) = 1$.
22. Solve the following initial value problems involving homogeneous differential equations.
- (a) $y' = 1 + \left(\frac{y}{x}\right) + \left(\frac{y}{x}\right)^2$, $y(1) = 1$.
- (b) $y' = \tan\left(\frac{y}{x}\right) + \frac{y}{x}$, $y(1) = 1$.
- (c) $y' = \frac{x^3+xy^2+y^3}{x^2y+xy^2}$, $y(1) = 1$.
23. * Suppose f is a function defined on $\mathbb{R}$ and it satisfies the functional equation, $f(a+b) = f(a)f(b) + 3ab$. Show $f(0)$ must equal either 1 or 0. Show that if f is to be differentiable, then $f(0) = 1$. Assuming $f(0) = 1$ and $f'(0) = 2$, find $f(x)$ if it exists.

8.4 The Equations Of Undamped And Damped Oscillation

Consider a garage door spring. These springs exert a force which resists extension. Attach such a spring to the ceiling and attach a mass, m , to the bottom end of the spring as shown in the following picture. Any mass will do. It does not have to be a small elephant.



The weight of this mass, mg , is a downward force which extends the spring, moving the bottom end of the spring downward a distance l where the upward force exerted by the spring exactly balances the downward force exerted on the mass by gravity. It has been experimentally observed that as long as the extension, z , of such a spring is not too great, the restoring force exerted by the spring is of the form kz where k is some constant which depends on the spring. (It would be different for a slinky than for a garage door spring.) This is known as Hooke's law which is the simplest model for elastic materials. Therefore, $mg = kl$. Now let y be the displacement from this equilibrium position of the bottom of the spring with the positive direction being up. Thus the acceleration of the spring is y'' . The extension of the spring in terms of y is $(l - y)$. Then Newton's second law² along with Hooke's law imply

$$my'' = k(l - y) - mg$$

²This important law will be discussed more thoroughly later. I assume you have seen it in a physics class by now.

and since $kl - mg = 0$, this yields

$$my'' + ky = 0.$$

Dividing by m and letting $\omega^2 = k/m$ yields the equation for undamped oscillation,

$$y'' + \omega^2 y = 0.$$

Based on physical reasoning just presented, there should be a solution to this equation. It is the displacement of the bottom end of a spring from the equilibrium position. However, it is not enough to base questions of existence in mathematics on physical intuition, although it is sometimes done. The following theorem gives the necessary existence and uniqueness results. The equation is the equation of undamped oscillations. It occurs in modeling a weight on a spring but it also occurs in many other physical settings.

Theorem 8.4.1 *The initial value problem,*

$$y'' + \omega^2 y = 0, \quad y(0) = y_0, y'(0) = y_1 \quad (8.5)$$

has a unique solution and this solution is

$$y(t) = y_0 \cos(\omega t) + \frac{y_1}{\omega} \sin(\omega t). \quad (8.6)$$

Proof: You should verify that 8.6 does indeed provide a solution to the initial value problem. It only remains to verify uniqueness. Suppose then that y_1 and y_2 both solve the initial value problem, 8.5. Let $w = y_1 - y_2$. Then you should verify that $w'' + \omega^2 w = 0$, $w(0) = 0 = w'(0)$. Then multiplying both sides of the differential equation by w' it follows

$$w''w' + \omega^2 ww' = 0.$$

However, $w''w' = \frac{1}{2} \frac{d}{dt} (w')^2$ and $ww' = \frac{1}{2} \frac{d}{dt} (w^2)$ so the above equation reduces to

$$\frac{1}{2} \frac{d}{dt} \left((w')^2 + w^2 \right) = 0.$$

Therefore, $(w')^2 + w^2$ is equal to some constant. However, when $t = 0$, this shows the constant must be zero. Therefore, $y_1 - y_2 = w = 0$. This proves the uniqueness.

Now consider another sort of differential equation,

$$y'' - a^2 y = 0, \quad a > 0 \quad (8.7)$$

To give the complete solution, let $Dy \equiv y'$. Then the differential equation may be written as

$$(D + a)(D - a)y = 0.$$

Let $z = (D - a)y$. Thus $(D + a)z = 0$ and so $z(t) = C_1 e^{-at}$ from Theorem 8.1.1 on Page 203. Therefore,

$$(D - a)y \equiv y' - ay = C_1 e^{-at}.$$

Multiply both sides of this last equation by e^{-at} . By the product and chain rules,

$$\frac{d}{dt} (e^{-at} y) = C_1 e^{-2at}.$$

Therefore,

$$e^{-at} y = \frac{C_1}{-2a} e^{-2at} + C_2$$

and so

$$y = \frac{C_1}{-2a}e^{-at} + C_2e^{at}.$$

Now since C_1 is arbitrary, it follows any solution of 8.7 is of the form $y = C_1e^{-at} + C_2e^{at}$. Now you should verify that any expression of this form actually solves the equation, 8.7. This proves most of the following theorem.

Theorem 8.4.2 *Every solution of the differential equation, $y'' - a^2y = 0$ is of the form $C_1e^{-at} + C_2e^{at}$ for some constants C_1 and C_2 provided $a > 0$. In the case when $a = 0$, every solution of $y'' = 0$ is of the form $C_1t + C_2$ for some constants, C_1 and C_2 .*

All that remains of the proof is to do the part when $a = 0$ which is left as an exercise involving the mean value theorem.

Now consider the differential equation of damped oscillation. In the example of the object bobbing on the end of a spring,

$$my'' = -ky$$

where k was the spring constant and m was the mass of the object. Suppose the object is also attached to a dash pot. This is a device which resists motion like a shock absorber on a car. You know how these work. If the car is just sitting still the shock absorber applies no force to the car. It only gives a force in response to up and down motion of the car and you assume this force is proportional to the velocity and opposite the velocity. Thus in our spring example, you would have

$$my'' = -ky - \delta^2 y'$$

where δ^2 is the constant of proportionality of the resisting force. Dividing by m and adjusting the coefficients, such damped oscillation satisfies an equation of the form,

$$y'' + 2by' + ay = 0. \quad (8.8)$$

Actually this is a general homogeneous second order equation, more general than what results from damped oscillation. Concerning the solutions to this equation, the following theorem is given. In this theorem the first case is referred to as the underdamped case. The second case is called the critically damped case and the third is called the overdamped case.

Theorem 8.4.3 *Suppose $b^2 - a < 0$. Then all solutions of 8.8 are of the form*

$$e^{-bt} (C_1 \cos(\omega t) + C_2 \sin(\omega t)). \quad (8.9)$$

where $\omega = \sqrt{a - b^2}$ and C_1 and C_2 are constants. In the case that $b^2 - a = 0$ the solutions of 8.8 are of the form

$$e^{-bt} (C_1 + C_2 t). \quad (8.10)$$

In the case that $b^2 - a > 0$ the solutions are of the form

$$e^{-bt} (C_1 e^{-rt} + C_2 e^{rt}), \quad (8.11)$$

where $r = \sqrt{b^2 - a}$.

Proof: Let $z = e^{bt}y$ and write 8.8 in terms of z . Thus, z is a solution to the equation

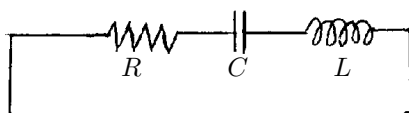
$$z'' + (a - b^2)z = 0. \quad (8.12)$$

If $b^2 - a < 0$, then by Theorem 8.4.1, $z(t) = C_1 \cos(\omega t) + C_2 \sin(\omega t)$ where $\omega = \sqrt{a - b^2}$. Therefore,

$$y = e^{-bt} (C_1 \cos(\omega t) + C_2 \sin(\omega t))$$

as claimed. The other two cases are completely similar. They use Theorem 8.4.2 rather than Theorem 8.4.1.

Example 8.4.4 An important example of these equations occurs in an electrical circuit having a capacitor, a resistor, and an inductor in series as shown in the following picture.



The voltage drop across the inductor is $L \frac{di}{dt}$ where i is the current and L is the inductance. The voltage drop across the resistor is Ri where R is the resistance. This is according to Ohm's law. The voltage drop across the capacitor is $v = \frac{Q}{C}$ where Q is the charge on the capacitor and C is a constant called the capacitance. The current equals the rate of change of the charge on the capacitor. Thus $i = Q' = Cv'$. When these voltages are summed, you must get zero because there is no voltage source in the circuit. Thus $L \frac{di}{dt} + Ri + \frac{Q}{C} = 0$ and written in terms of the voltage drop across the capacitor, this becomes $LCv'' + CRv' + v = 0$, a second order linear differential equation of the sort discussed above.

8.5 Exercises

1. Verify that $y = C_1 e^{-at} + C_2 e^{at}$ solves the differential equation, 8.7.
2. Verify 8.12.
3. Verify that all solutions to the differential equation, $y'' = 0$ are of the form $y = C_1 t + C_2$.
4. Show that for all $x \geq 0$,

$$\sin x \leq x - \frac{x^3}{6} + \frac{x^5}{120}. \quad (8.13)$$

Also verify $\sin x \leq x - \frac{x^3}{6}$. **Hint:** Let $f(x) = x - \frac{x^3}{6} + \frac{x^5}{120} - \sin x$. Then $f(0) = 0$. You need to show $f'(x) > 0$ for $x > 0$. You won't be able to do this directly. Consider $f'(x) = 1 - \frac{x^2}{2} + \frac{x^4}{24} - \cos x$. Then $f'(0) = 0$. You need to show $f''(x) > 0$. You won't be able to do this directly. Consider $f''(x) = -x + \frac{x^3}{6} + \sin x$. Then $f''(0) = 0$. You need to show $f'''(x) > 0$. You won't be able to do this directly. Continue this way. You will eventually have $f^{(k)}(0) = 0$ and it will be obvious that $f^{(k+1)}(x) > 0$.

5. Using Problem 4, along with similar estimates for lower bounds and upper bounds for sine and cosine, estimate $\sin(.1)$ and $\cos(.1)$. Give upper and lower bounds for these numbers.
6. Using Problem 4 and Theorem 2.9.5 on Page 60, establish the limit,

$$\lim_{x \rightarrow 0} \frac{\sin(x)}{x} = 1.$$

7. A mass of ten Kilograms is suspended from a spring attached to the ceiling. This mass causes the end of the spring to be displaced a distance of 39.2 *cm*. The mass end of the spring is then pulled down a distance of one *cm*. and released. Find the displacement from the equilibrium position of the end of the spring as a function of time. Assume the acceleration of gravity is 9.8 *meters/sec*².
8. Keep everything the same in Problem 7 except suppose the suspended end of the spring is also attached to a dash pot which provides a force opposite the direction of the velocity having magnitude $10\sqrt{19}|v|$ Newtons for $|v|$ the speed. Give the displacement as before.
9. In 8.8 consider the equation in which $b = -1$ and $a = 3$. Explain why this equation describes a physical system which has some dubious properties.
10. Solve the initial value problem, $y'' + 5y' - y = 0$, $y(0) = 1$, $y'(0) = 0$.
11. Solve the initial value problem $y'' + 2y' + 2y = 0$, $y(0) = 0$, $y'(0) = 1$.
12. Suppose $f(t) = a \cos \omega t + b \sin \omega t$. Show there exists ϕ such that

$$f(t) = \sqrt{a^2 + b^2} \sin(\omega t + \phi).$$

Hint: $f(t) = \sqrt{a^2 + b^2} \left(\frac{a}{\sqrt{a^2 + b^2}} \cos \omega t + \frac{b}{\sqrt{a^2 + b^2}} \sin \omega t \right)$ and $\left(\frac{b}{\sqrt{a^2 + b^2}}, \frac{a}{\sqrt{a^2 + b^2}} \right)$ is a point on the unit circle so it is of the form $(\cos \phi, \sin \phi)$. Now recall the formulas for the sine and cosine of sums of angles. Can you also write $f(t) = \sqrt{a^2 + b^2} \cos(\omega t + \phi)$? Explain.

13. In the case of undamped oscillation show the solution can be written in the form $A \cos(\omega t - \phi)$ where ϕ is some angle called a phase shift and a constant, A , called the amplitude.
14. Using Problem 4 and Theorem 2.9.5, establish the limit,

$$\lim_{x \rightarrow 0} \frac{1 - \cos x}{x} = 0.$$

15. Is the derivative of a function always continuous? **Hint:** Consider

$$h(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

Show $h'(0) = 0$. What is $h'(x)$ for $x \neq 0$?

The Integral

9.0.1 Outcomes

1. Describe upper and lower sums and give a correct definition of the integral.
2. Understand theorems about the existence of the integral of compositions of continuous functions with Riemann integrable functions.
3. Understand and be able to use intelligently all fundamental properties of the integral.
4. Understand the proof of the fundamental theorem of calculus and be able to use this theorem.
5. Understand a mathematically rigorous treatment of the logarithm.
6. Understand and be able to use various integration techniques in the context of definite integrals.
7. Describe the difference between an improper integral and an integral. Also tell what it means for an improper integral to converge and understand and use theorems which enable you to make this determination.
8. Understand and work problems involving improper integrals.

The integral originated in attempts to find areas of various shapes and the ideas involved in finding integrals are much older than the ideas related to finding derivatives. In fact, Archimedes¹ was finding areas of various curved shapes about 250 B.C. The integral is needed to remove some of the mathematical loose ends and also to enable the study of more general problems involving differential equations. It will also be useful for formulating other physical models. The technique used for finding the area of a circular segment presented early in the book was essentially that employed by Archimedes and contains the essential ideas for the integral. The main difference is that here the triangles will be replaced with rectangles. You may be wondering what the fuss is about. Areas have already been found as solutions of differential equations. However, there is a profound difference between what is about to be presented and what has just been done. It is related to the fundamental mathematical question of existence. As an illustration, consider the problem of finding the area between $y = e^{x^2}$ and the x axis for

¹Archimedes 287-212 B.C. found areas of curved regions by stuffing them with simple shapes which he knew the area of and taking a limit. He also made fundamental contributions to physics. The story is told about how he determined that a gold smith had cheated the king by giving him a crown which was not solid gold as had been claimed. He did this by finding the amount of water displaced by the crown and comparing with the amount of water it should have displaced if it had been solid gold.

$x \in [0, 1]$. As pointed out earlier, the area is obtained as a solution to the initial value problem,

$$A'(x) = e^{x^2}, \quad A(0) = 0.$$

So what is the solution to this initial value problem? By Theorem 6.1.1 there is at most one solution, but what is the solution? Does it even exist? More generally, for which functions, f does there exist a solution to the initial value problem, $y'(x) = f(x), y(0) = y_0$? These questions are typical of mathematics. There are usually two aspects to a mathematical concept. One is the question of existence and the other is how to find that which exists. The two questions are often very different and one can have a good understanding of one without having any idea how to go about considering the other. However, both are absolutely essential. In the preceding chapter the only thing considered was the second question.

9.1 Upper And Lower Sums

The Riemann integral pertains to bounded functions which are defined on a bounded interval. Let $[a, b]$ be a closed interval. A set of points in $[a, b]$, $\{x_0, \dots, x_n\}$ is a partition if

$$a = x_0 < x_1 < \dots < x_n = b.$$

Such partitions are denoted by P or Q . For f a bounded function defined on $[a, b]$, let

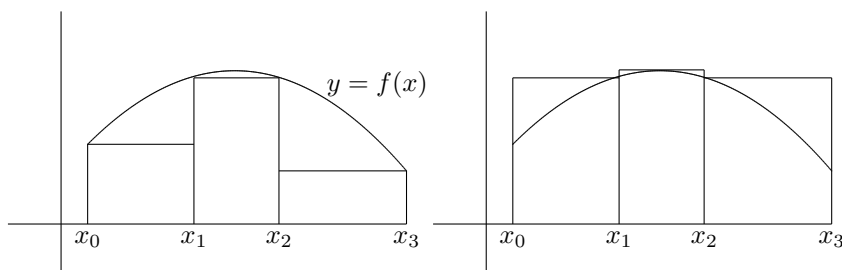
$$M_i(f) \equiv \sup\{f(x) : x \in [x_{i-1}, x_i]\},$$

$$m_i(f) \equiv \inf\{f(x) : x \in [x_{i-1}, x_i]\}.$$

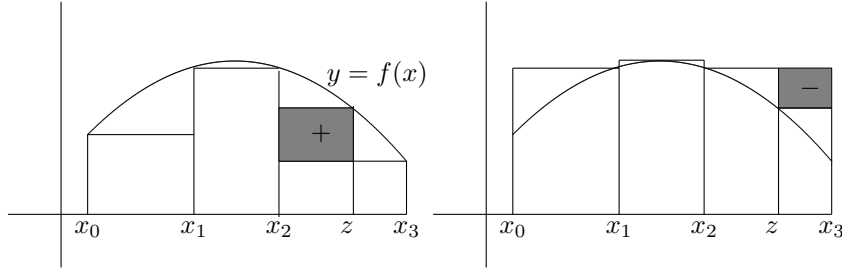
Also let $\Delta x_i \equiv x_i - x_{i-1}$. Then define upper and lower sums as

$$U(f, P) \equiv \sum_{i=1}^n M_i(f) \Delta x_i \text{ and } L(f, P) \equiv \sum_{i=1}^n m_i(f) \Delta x_i$$

respectively. The numbers, $M_i(f)$ and $m_i(f)$, are well defined real numbers because f is assumed to be bounded and $\mathbb{R}$ is complete. Thus the set $S = \{f(x) : x \in [x_{i-1}, x_i]\}$ is bounded above and below. In the following picture, the sum of the areas of the rectangles in the picture on the left is a lower sum for the function in the picture and the sum of the areas of the rectangles in the picture on the right is an upper sum for the same function which uses the same partition.



What happens when you add in more points in a partition? The following pictures illustrate in the context of the above example. In this example a single additional point, labeled z has been added in.



Note how the lower sum got larger by the amount of the area in the shaded rectangle and the upper sum got smaller by the amount in the other shaded rectangle. In general this is the way it works and this is shown in the following lemma.

Lemma 9.1.1 *If $P \subseteq Q$ then*

$$U(f, Q) \leq U(f, P), \text{ and } L(f, P) \leq L(f, Q).$$

Proof: This is verified by adding in one point at a time. Thus let $P = \{x_0, \dots, x_n\}$ and let $Q = \{x_0, \dots, x_k, y, x_{k+1}, \dots, x_n\}$. Thus exactly one point, y , is added between x_k and x_{k+1} . Now the term in the upper sum which corresponds to the interval $[x_k, x_{k+1}]$ in $U(f, P)$ is

$$\sup \{f(x) : x \in [x_k, x_{k+1}]\} (x_{k+1} - x_k) \quad (9.1)$$

and the term which corresponds to the interval $[x_k, x_{k+1}]$ in $U(f, Q)$ is

$$\sup \{f(x) : x \in [x_k, y]\} (y - x_k) + \sup \{f(x) : x \in [y, x_{k+1}]\} (x_{k+1} - y) \quad (9.2)$$

$$\equiv M_1 (y - x_k) + M_2 (x_{k+1} - y) \quad (9.3)$$

All the other terms in the two sums coincide. Now $\sup \{f(x) : x \in [x_k, x_{k+1}]\} \geq \max(M_1, M_2)$ and so the expression in 9.2 is no larger than

$$\begin{aligned} & \sup \{f(x) : x \in [x_k, x_{k+1}]\} (x_{k+1} - y) + \sup \{f(x) : x \in [x_k, x_{k+1}]\} (y - x_k) \\ &= \sup \{f(x) : x \in [x_k, x_{k+1}]\} (x_{k+1} - x_k), \end{aligned}$$

the term corresponding to the interval, $[x_k, x_{k+1}]$ and $U(f, P)$. This proves the first part of the lemma pertaining to upper sums because if $Q \supseteq P$, one can obtain Q from P by adding in one point at a time and each time a point is added, the corresponding upper sum either gets smaller or stays the same. The second part is similar and is left as an exercise.

Lemma 9.1.2 *If P and Q are two partitions, then*

$$L(f, P) \leq U(f, Q).$$

Proof: By Lemma 9.1.1,

$$L(f, P) \leq L(f, P \cup Q) \leq U(f, P \cup Q) \leq U(f, Q).$$

Definition 9.1.3

$$\bar{I} \equiv \inf \{U(f, Q) \text{ where } Q \text{ is a partition}\}$$

$$\underline{I} \equiv \sup \{L(f, P) \text{ where } P \text{ is a partition}\}.$$

Note that $\underline{I}$ and $\bar{I}$ are well defined real numbers.

Theorem 9.1.4 $\underline{I} \leq \bar{I}$.

Proof: From Lemma 9.1.2,

$$\underline{I} = \sup\{L(f, P) \text{ where } P \text{ is a partition}\} \leq U(f, Q)$$

because $U(f, Q)$ is an upper bound to the set of all lower sums and so it is no smaller than the least upper bound. Therefore, since Q is arbitrary,

$$\begin{aligned} \underline{I} &= \sup\{L(f, P) \text{ where } P \text{ is a partition}\} \\ &\leq \inf\{U(f, Q) \text{ where } Q \text{ is a partition}\} \equiv \bar{I} \end{aligned}$$

where the inequality holds because it was just shown that $\underline{I}$ is a lower bound to the set of all upper sums and so it is no larger than the greatest lower bound of this set. This proves the theorem.

Definition 9.1.5 A bounded function f is Riemann integrable, written as

$$f \in R([a, b])$$

if

$$\underline{I} = \bar{I}$$

and in this case,

$$\int_a^b f(x) dx \equiv \underline{I} = \bar{I}.$$

Thus, in words, the Riemann integral is the unique number which lies between all upper sums and all lower sums if there is such a unique number.

Proposition 9.1.6 Let S be a nonempty set and suppose $\sup(S)$ exists. Then for every $\delta > 0$,

$$S \cap (\sup(S) - \delta, \sup(S)] \neq \emptyset.$$

If $\inf(S)$ exists, then for every $\delta > 0$,

$$S \cap [\inf(S), \inf(S) + \delta) \neq \emptyset.$$

Proof: Consider the first claim. If the indicated set equals $\emptyset$, then $\sup(S) - \delta$ is an upper bound for S which is smaller than $\sup(S)$, contrary to the definition of $\sup(S)$ as the least upper bound. In the second claim, if the indicated set equals $\emptyset$, then $\inf(S) + \delta$ would be a lower bound which is larger than $\inf(S)$ contrary to the definition of $\inf(S)$.

This proposition implies the following theorem which is used to determine the question of Riemann integrability.

Theorem 9.1.7 A bounded function f is Riemann integrable if and only if for all $\varepsilon > 0$, there exists a partition P such that

$$U(f, P) - L(f, P) < \varepsilon. \quad (9.4)$$

Proof: First assume f is Riemann integrable. Then let P and Q be two partitions such that

$$U(f, Q) < \bar{I} + \varepsilon/2, \quad L(f, P) > \underline{I} - \varepsilon/2.$$

Then since $\underline{I} = \bar{I}$,

$$U(f, Q \cup P) - L(f, P \cup Q) \leq U(f, Q) - L(f, P) < \bar{I} + \varepsilon/2 - (\underline{I} - \varepsilon/2) = \varepsilon.$$

Now suppose that for all $\varepsilon > 0$ there exists a partition such that 9.4 holds. Then for given ε and partition P corresponding to ε

$$\bar{I} - \underline{I} \leq U(f, P) - L(f, P) \leq \varepsilon.$$

Since ε is arbitrary, this shows $\underline{I} = \bar{I}$ and this proves the theorem.

The condition described in the theorem is called the Riemann criterion.

Not all bounded functions are Riemann integrable. For example, let

$$f(x) \equiv \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases} \quad (9.5)$$

Then if $[a, b] = [0, 1]$ all upper sums for f equal 1 while all lower sums for f equal 0. Therefore the Riemann criterion is violated for $\varepsilon = 1/2$.

9.2 Exercises

1. Prove the second half of Lemma 9.1.1 about lower sums.
2. Verify that for f given in 9.5, the lower sums on the interval $[0, 1]$ are all equal to zero while the upper sums are all equal to one.
3. Let $f(x) = 1 + x^2$ for $x \in [-1, 3]$ and let $P = \{-1, -\frac{1}{3}, 0, \frac{1}{2}, 1, 2\}$. Find $U(f, P)$ and $L(f, P)$.
4. Show that if $f \in R([a, b])$, there exists a partition, $\{x_0, \dots, x_n\}$ such that for any $z_k \in [x_k, x_{k+1}]$,

$$\left| \int_a^b f(x) dx - \sum_{k=1}^n f(z_k)(x_k - x_{k-1}) \right| < \varepsilon$$

This sum, $\sum_{k=1}^n f(z_k)(x_k - x_{k-1})$, is called a Riemann sum and this exercise shows that the integral can always be approximated by a Riemann sum.

5. Let $P = \{1, 1\frac{1}{4}, 1\frac{1}{2}, 1\frac{3}{4}, 2\}$. Find upper and lower sums for the function, $f(x) = \frac{1}{x}$ using this partition. What does this tell you about $\ln(2)$?
6. If $f \in R([a, b])$ and f is changed at finitely many points, show the new function is also in $R([a, b])$ and has the same integral as the unchanged function.
7. Consider the function, $y = x^2$ for $x \in [0, 1]$. Show this function is Riemann integrable and find the integral using the definition and the formula

$$\sum_{k=1}^n k^2 = \frac{1}{3}(n+1)^3 - \frac{1}{2}(n+1)^2 + \frac{1}{6}(n+1)$$

which you should verify by using math induction. This is not a practical way to find integrals in general.

8. Define a “left sum” as

$$\sum_{k=1}^n f(x_{k-1})(x_k - x_{k-1})$$

and a “right sum”,

$$\sum_{k=1}^n f(x_k)(x_k - x_{k-1}).$$

Also suppose that all partitions have the property that $x_k - x_{k-1}$ equals a constant, $(b - a)/n$ so the points in the partition are equally spaced, and define the integral to be the number these right and left sums get close to as n gets larger and larger. Show that for f given in 9.5, $\int_0^x f(t) dt = 1$ if x is rational and $\int_0^x f(t) dt = 0$ if x is irrational. It turns out that the correct answer should always equal zero for that function, regardless of whether x is rational. This is shown in more advanced courses when the Lebesgue integral is studied. This illustrates why the method of defining the integral in terms of left and right sums is nonsense.

9.3 Functions Of Riemann Integrable Functions

It is often necessary to consider functions of Riemann integrable functions and a natural question is whether these are Riemann integrable. The following theorem gives a partial answer to this question. This is not the most general theorem which will relate to this question but it will be enough for the needs of this book.

Theorem 9.3.1 *Let f, g be bounded functions and let $f([a, b]) \subseteq [c_1, d_1]$ and $g([a, b]) \subseteq [c_2, d_2]$. Let $H : [c_1, d_1] \times [c_2, d_2] \rightarrow \mathbb{R}$ satisfy,*

$$|H(a_1, b_1) - H(a_2, b_2)| \leq K[|a_1 - a_2| + |b_1 - b_2|]$$

for some constant K . Then if $f, g \in R([a, b])$ it follows that $H \circ (f, g) \in R([a, b])$.

Proof: In the following claim, $M_i(h)$ and $m_i(h)$ have the meanings assigned above with respect to some partition of $[a, b]$ for the function, h .

Claim: The following inequality holds.

$$|M_i(H \circ (f, g)) - m_i(H \circ (f, g))| \leq K[|M_i(f) - m_i(f)| + |M_i(g) - m_i(g)|].$$

Proof of the claim: By the above proposition, there exist $x_1, x_2 \in [x_{i-1}, x_i]$ be such that

$$H(f(x_1), g(x_1)) + \eta > M_i(H \circ (f, g)),$$

and

$$H(f(x_2), g(x_2)) - \eta < m_i(H \circ (f, g)).$$

Then

$$\begin{aligned} & |M_i(H \circ (f, g)) - m_i(H \circ (f, g))| \\ & < 2\eta + |H(f(x_1), g(x_1)) - H(f(x_2), g(x_2))| \\ & < 2\eta + K[|f(x_1) - f(x_2)| + |g(x_1) - g(x_2)|] \\ & \leq 2\eta + K[|M_i(f) - m_i(f)| + |M_i(g) - m_i(g)|]. \end{aligned}$$

Since $\eta > 0$ is arbitrary, this proves the claim.

Now continuing with the proof of the theorem, let P be such that

$$\sum_{i=1}^n (M_i(f) - m_i(f)) \Delta x_i < \frac{\varepsilon}{2K}, \quad \sum_{i=1}^n (M_i(g) - m_i(g)) \Delta x_i < \frac{\varepsilon}{2K}.$$

Then from the claim,

$$\sum_{i=1}^n (M_i(H \circ (f, g)) - m_i(H \circ (f, g))) \Delta x_i$$

$$< \sum_{i=1}^n K [|M_i(f) - m_i(f)| + |M_i(g) - m_i(g)|] \Delta x_i < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this shows $H \circ (f, g)$ satisfies the Riemann criterion and hence $H \circ (f, g)$ is Riemann integrable as claimed. This proves the theorem.

This theorem implies that if f, g are Riemann integrable, then so is $af + bg, |f|, f^2$, along with infinitely many other such continuous combinations of Riemann integrable functions. For example, to see that $|f|$ is Riemann integrable, let $H(a, b) = |a|$. Clearly this function satisfies the conditions of the above theorem and so $|f| = H(f, f) \in R([a, b])$ as claimed. The following theorem gives an example of many functions which are Riemann integrable.

Theorem 9.3.2 *Let $f : [a, b] \rightarrow \mathbb{R}$ be either increasing or decreasing on $[a, b]$. Then $f \in R([a, b])$.*

Proof: Let $\varepsilon > 0$ be given and let

$$x_i = a + i \left(\frac{b-a}{n} \right), \quad i = 0, \dots, n.$$

Then since f is increasing,

$$\begin{aligned} U(f, P) - L(f, P) &= \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \left(\frac{b-a}{n} \right) \\ &= (f(b) - f(a)) \left(\frac{b-a}{n} \right) < \varepsilon \end{aligned}$$

whenever n is large enough. Thus the Riemann criterion is satisfied and so the function is Riemann integrable. The proof for decreasing f is similar.

Corollary 9.3.3 *Let $[a, b]$ be a bounded closed interval and let $\phi : [a, b] \rightarrow \mathbb{R}$ be Lipschitz continuous. Then $\phi \in R([a, b])$. Recall that a function, ϕ , is Lipschitz continuous if there is a constant, K , such that for all x, y ,*

$$|\phi(x) - \phi(y)| < K|x - y|.$$

Proof: Let $f(x) = x$. Then by Theorem 9.3.2, f is Riemann integrable. Let $H(a, b) \equiv \phi(a)$. Then by Theorem 9.3.1 $H \circ (f, f) = \phi \circ f = \phi$ is also Riemann integrable. This proves the corollary.

In fact, it is enough to assume ϕ is continuous, although this is harder. This is the content of the next theorem which is where the difficult theorems about continuity and uniform continuity are used.

Theorem 9.3.4 *Suppose $f : [a, b] \rightarrow \mathbb{R}$ is continuous. Then $f \in R([a, b])$.*

Proof: By Corollary 2.13.5 on Page 70, f is uniformly continuous on $[a, b]$. Therefore, if $\varepsilon > 0$ is given, there exists a $\delta > 0$ such that if $|x_i - x_{i-1}| < \delta$, then $M_i - m_i < \frac{\varepsilon}{b-a}$. Let

$$P \equiv \{x_0, \dots, x_n\}$$

be a partition with $|x_i - x_{i-1}| < \delta$. Then

$$U(f, P) - L(f, P) < \sum_{i=1}^n (M_i - m_i)(x_i - x_{i-1}) < \frac{\varepsilon}{b-a} (b-a) = \varepsilon.$$

By the Riemann criterion, $f \in R([a, b])$. This proves the theorem.

9.4 Properties Of The Integral

The integral has many important algebraic properties. First here is a simple lemma.

Lemma 9.4.1 *Let S be a nonempty set which is bounded above and below. Then if $-S \equiv \{-x : x \in S\}$,*

$$\sup(-S) = -\inf(S) \quad (9.6)$$

and

$$\inf(-S) = -\sup(S). \quad (9.7)$$

Proof: Consider 9.6. Let $x \in S$. Then $-x \leq \sup(-S)$ and so $x \geq -\sup(-S)$. It follows that $-\sup(-S)$ is a lower bound for S and therefore, $-\sup(-S) \leq \inf(S)$. This implies $\sup(-S) \geq -\inf(S)$. Now let $-x \in -S$. Then $x \in S$ and so $x \geq \inf(S)$ which implies $-x \leq -\inf(S)$. Therefore, $-\inf(S)$ is an upper bound for $-S$ and so $-\inf(S) \geq \sup(-S)$. This shows 9.6. Formula 9.7 is similar and is left as an exercise.

In particular, the above lemma implies that for $M_i(f)$ and $m_i(f)$ defined above $M_i(-f) = -m_i(f)$, and $m_i(-f) = -M_i(f)$.

Lemma 9.4.2 *If $f \in R([a, b])$ then $-f \in R([a, b])$ and*

$$-\int_a^b f(x) dx = \int_a^b -f(x) dx.$$

Proof: The first part of the conclusion of this lemma follows from Theorem 9.3.2 since the function $\phi(y) \equiv -y$ is Lipschitz continuous. Now choose P such that

$$\int_a^b -f(x) dx - L(-f, P) < \varepsilon.$$

Then since $m_i(-f) = -M_i(f)$,

$$\varepsilon > \int_a^b -f(x) dx - \sum_{i=1}^n m_i(-f) \Delta x_i = \int_a^b -f(x) dx + \sum_{i=1}^n M_i(f) \Delta x_i$$

which implies

$$\varepsilon > \int_a^b -f(x) dx + \sum_{i=1}^n M_i(f) \Delta x_i \geq \int_a^b -f(x) dx + \int_a^b f(x) dx.$$

Thus, since ε is arbitrary,

$$\int_a^b -f(x) dx \leq -\int_a^b f(x) dx$$

whenever $f \in R([a, b])$. It follows

$$\int_a^b -f(x) dx \leq -\int_a^b f(x) dx = -\int_a^b -(-f(x)) dx \leq \int_a^b -f(x) dx$$

and this proves the lemma.

Theorem 9.4.3 *The integral is linear,*

$$\int_a^b (\alpha f + \beta g)(x) dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx.$$

whenever $f, g \in R([a, b])$ and $\alpha, \beta \in \mathbb{R}$.

Proof: First note that by Theorem 9.3.1, $\alpha f + \beta g \in R([a, b])$. To begin with, consider the claim that if $f, g \in R([a, b])$ then

$$\int_a^b (f + g)(x) dx = \int_a^b f(x) dx + \int_a^b g(x) dx. \quad (9.8)$$

Let P_1, Q_1 be such that

$$U(f, Q_1) - L(f, Q_1) < \varepsilon/2, \quad U(g, P_1) - L(g, P_1) < \varepsilon/2.$$

Then letting $P \equiv P_1 \cup Q_1$, Lemma 9.1.1 implies

$$U(f, P) - L(f, P) < \varepsilon/2, \quad \text{and} \quad U(g, P) - L(g, P) < \varepsilon/2.$$

Next note that

$$m_i(f + g) \geq m_i(f) + m_i(g), \quad M_i(f + g) \leq M_i(f) + M_i(g).$$

Therefore,

$$L(g + f, P) \geq L(f, P) + L(g, P), \quad U(g + f, P) \leq U(f, P) + U(g, P).$$

For this partition,

$$\begin{aligned} \int_a^b (f + g)(x) dx &\in [L(f + g, P), U(f + g, P)] \\ &\subseteq [L(f, P) + L(g, P), U(f, P) + U(g, P)] \end{aligned}$$

and

$$\int_a^b f(x) dx + \int_a^b g(x) dx \in [L(f, P) + L(g, P), U(f, P) + U(g, P)].$$

Therefore,

$$\begin{aligned} \left| \int_a^b (f + g)(x) dx - \left(\int_a^b f(x) dx + \int_a^b g(x) dx \right) \right| &\leq \\ U(f, P) + U(g, P) - (L(f, P) + L(g, P)) &< \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

This proves 9.8 since ε is arbitrary.

It remains to show that

$$\alpha \int_a^b f(x) dx = \int_a^b \alpha f(x) dx.$$

Suppose first that $\alpha \geq 0$. Then

$$\begin{aligned} \int_a^b \alpha f(x) dx &\equiv \sup\{L(\alpha f, P) : P \text{ is a partition}\} = \\ \alpha \sup\{L(f, P) : P \text{ is a partition}\} &\equiv \alpha \int_a^b f(x) dx. \end{aligned}$$

If $\alpha < 0$, then this and Lemma 9.4.2 imply

$$\begin{aligned} \int_a^b \alpha f(x) dx &= \int_a^b (-\alpha)(-f(x)) dx \\ &= (-\alpha) \int_a^b (-f(x)) dx = \alpha \int_a^b f(x) dx. \end{aligned}$$

This proves the theorem.

Theorem 9.4.4 *If $f \in R([a, b])$ and $f \in R([b, c])$, then $f \in R([a, c])$ and*

$$\int_a^c f(x) dx = \int_a^b f(x) dx + \int_b^c f(x) dx. \quad (9.9)$$

Proof: Let P_1 be a partition of $[a, b]$ and P_2 be a partition of $[b, c]$ such that

$$U(f, P_i) - L(f, P_i) < \varepsilon/2, \quad i = 1, 2.$$

Let $P \equiv P_1 \cup P_2$. Then P is a partition of $[a, c]$ and

$$\begin{aligned} U(f, P) - L(f, P) \\ = U(f, P_1) - L(f, P_1) + U(f, P_2) - L(f, P_2) < \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned} \quad (9.10)$$

Thus, $f \in R([a, c])$ by the Riemann criterion and also for this partition,

$$\begin{aligned} \int_a^b f(x) dx + \int_b^c f(x) dx &\in [L(f, P_1) + L(f, P_2), U(f, P_1) + U(f, P_2)] \\ &= [L(f, P), U(f, P)] \end{aligned}$$

and

$$\int_a^c f(x) dx \in [L(f, P), U(f, P)].$$

Hence by 9.10,

$$\left| \int_a^c f(x) dx - \left(\int_a^b f(x) dx + \int_b^c f(x) dx \right) \right| < U(f, P) - L(f, P) < \varepsilon$$

which shows that since ε is arbitrary, 9.9 holds. This proves the theorem.

Corollary 9.4.5 *Let $[a, b]$ be a closed and bounded interval and suppose that*

$$a = y_1 < y_2 < \cdots < y_l = b$$

and that f is a bounded function defined on $[a, b]$ which has the property that f is either increasing on $[y_j, y_{j+1}]$ or decreasing on $[y_j, y_{j+1}]$ for $j = 1, \dots, l-1$. Then $f \in R([a, b])$.

Proof: This follows from Theorem 9.4.4 and Theorem 9.3.2.

The symbol, $\int_a^b f(x) dx$ when $a > b$ has not yet been defined.

Definition 9.4.6 *Let $[a, b]$ be an interval and let $f \in R([a, b])$. Then*

$$\int_b^a f(x) dx \equiv - \int_a^b f(x) dx.$$

Note that with this definition,

$$\int_a^a f(x) dx = - \int_a^a f(x) dx$$

and so

$$\int_a^a f(x) dx = 0.$$

Theorem 9.4.7 *Assuming all the integrals make sense,*

$$\int_a^b f(x) dx + \int_b^c f(x) dx = \int_a^c f(x) dx.$$

Proof: This follows from Theorem 9.4.4 and Definition 9.4.6. For example, assume

$$c \in (a, b).$$

Then from Theorem 9.4.4,

$$\int_a^c f(x) dx + \int_c^b f(x) dx = \int_a^b f(x) dx$$

and so by Definition 9.4.6,

$$\begin{aligned} \int_a^c f(x) dx &= \int_a^b f(x) dx - \int_c^b f(x) dx \\ &= \int_a^b f(x) dx + \int_b^c f(x) dx. \end{aligned}$$

The other cases are similar.

The following properties of the integral have either been established or they follow quickly from what has been shown so far.

$$\text{If } f \in R([a, b]) \text{ then if } c \in [a, b], f \in R([a, c]), \quad (9.11)$$

$$\int_a^b \alpha dx = \alpha(b - a), \quad (9.12)$$

$$\int_a^b (\alpha f + \beta g)(x) dx = \alpha \int_a^b f(x) dx + \beta \int_a^b g(x) dx, \quad (9.13)$$

$$\int_a^b f(x) dx + \int_b^c f(x) dx = \int_a^c f(x) dx, \quad (9.14)$$

$$\int_a^b f(x) dx \geq 0 \text{ if } f(x) \geq 0 \text{ and } a < b, \quad (9.15)$$

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx. \quad (9.16)$$

The only one of these claims which may not be completely obvious is the last one. To show this one, note that

$$|f(x)| - f(x) \geq 0, \quad |f(x)| + f(x) \geq 0.$$

Therefore, by 9.15 and 9.13, if $a < b$,

$$\int_a^b |f(x)| dx \geq \int_a^b f(x) dx$$

and

$$\int_a^b |f(x)| dx \geq - \int_a^b f(x) dx.$$

Therefore,

$$\int_a^b |f(x)| dx \geq \left| \int_a^b f(x) dx \right|.$$

If $b < a$ then the above inequality holds with a and b switched. This implies 9.16.

9.5 Fundamental Theorem Of Calculus

With these properties, it is easy to prove the fundamental theorem of calculus². Let $f \in R([a, b])$. Then by 9.11 $f \in R([a, x])$ for each $x \in [a, b]$. The first version of the fundamental theorem of calculus is a statement about the derivative of the function

$$x \rightarrow \int_a^x f(t) dt.$$

Theorem 9.5.1 *Let $f \in R([a, b])$ and let*

$$F(x) \equiv \int_a^x f(t) dt.$$

Then if f is continuous at $x \in (a, b)$,

$$F'(x) = f(x).$$

Proof: Let $x \in (a, b)$ be a point of continuity of f and let h be small enough that $x + h \in [a, b]$. Then by using 9.14,

$$h^{-1}(F(x+h) - F(x)) = h^{-1} \int_x^{x+h} f(t) dt.$$

Also, using 9.12,

$$f(x) = h^{-1} \int_x^{x+h} f(x) dt.$$

Therefore, by 9.16,

$$\begin{aligned} |h^{-1}(F(x+h) - F(x)) - f(x)| &= \left| h^{-1} \int_x^{x+h} (f(t) - f(x)) dt \right| \\ &\leq \left| h^{-1} \int_x^{x+h} |f(t) - f(x)| dt \right|. \end{aligned}$$

Let $\varepsilon > 0$ and let $\delta > 0$ be small enough that if $|t - x| < \delta$, then

$$|f(t) - f(x)| < \varepsilon.$$

Therefore, if $|h| < \delta$, the above inequality and 9.12 shows that

$$|h^{-1}(F(x+h) - F(x)) - f(x)| \leq |h|^{-1} \varepsilon |h| = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this shows

$$\lim_{h \rightarrow 0} h^{-1}(F(x+h) - F(x)) = f(x)$$

and this proves the theorem.

Note this gives existence for the initial value problem,

$$F'(x) = f(x), \quad F(a) = 0$$

whenever f is Riemann integrable and continuous.³

The next theorem is also called the fundamental theorem of calculus.

²This theorem is why Newton and Leibnitz are credited with inventing calculus. The integral had been around for thousands of years and the derivative was by their time well known. However the connection between these two ideas had not been fully made although Newton's predecessor, Isaac Barrow had made some progress in this direction.

³Of course it was proved that if f is continuous on a closed interval, $[a, b]$, then $f \in R([a, b])$ but this is a hard theorem using the difficult result about uniform continuity.

Theorem 9.5.2 Let $f \in R([a, b])$ and suppose there exists an antiderivative for f , G , such that

$$G'(x) = f(x)$$

for every point of (a, b) and G is continuous on $[a, b]$. Then

$$\int_a^b f(x) dx = G(b) - G(a). \quad (9.17)$$

Proof: Let $P = \{x_0, \dots, x_n\}$ be a partition satisfying

$$U(f, P) - L(f, P) < \varepsilon.$$

Then

$$\begin{aligned} G(b) - G(a) &= G(x_n) - G(x_0) \\ &= \sum_{i=1}^n G(x_i) - G(x_{i-1}). \end{aligned}$$

By the mean value theorem,

$$\begin{aligned} G(b) - G(a) &= \sum_{i=1}^n G'(z_i)(x_i - x_{i-1}) \\ &= \sum_{i=1}^n f(z_i) \Delta x_i \end{aligned}$$

where z_i is some point in $[x_{i-1}, x_i]$. It follows, since the above sum lies between the upper and lower sums, that

$$G(b) - G(a) \in [L(f, P), U(f, P)],$$

and also

$$\int_a^b f(x) dx \in [L(f, P), U(f, P)].$$

Therefore,

$$\left| G(b) - G(a) - \int_a^b f(x) dx \right| < U(f, P) - L(f, P) < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, 9.17 holds. This proves the theorem.

The following notation is often used in this context. Suppose F is an antiderivative of f as just described with F continuous on $[a, b]$ and $F' = f$ on (a, b) . Then

$$\int_a^b f(x) dx = F(b) - F(a) \equiv F(x) \Big|_a^b.$$

Recall how many interesting problems can be reduced to initial value problems. This was true of work, area, arc length and many other examples. The examples given were cooked up to work out and you could actually solve the initial value problem using known functions. What if you couldn't do this? The next theorem is a significant existence theorem which tells you that solutions of the initial value problem exist.

Theorem 9.5.3 Suppose f is a continuous function defined on an interval, (a, b) , $c \in (a, b)$, and $y_0 \in \mathbb{R}$. Then there exists a unique solution to the initial value problem,

$$F'(x) = f(x), \quad F(c) = y_0.$$

This solution is given by

$$F(x) = y_0 + \int_c^x f(t) dt. \quad (9.18)$$

Proof: From Theorem 9.3.4, it follows the integral in 9.18 is well defined. Now by the fundamental theorem of calculus, $F'(x) = f(x)$. Therefore, F solves the given differential equation. Also, $F(c) = y_0 + \int_c^c f(t) dt = y_0$ so the initial condition is also satisfied. This establishes the existence part of the theorem.

Suppose F and G both solve the initial value problem. Then $F'(x) - G'(x) = f(x) - f(x) = 0$ and so $F(x) - G(x) = C$ for some constant, C . However, $F(c) - G(c) = y_0 - y_0 = 0$ and so the constant C can only equal 0. This proves the uniqueness part of the theorem.

Example 9.5.4 Find the area between $y = x^2$ and $y = x^3$ for $x \in [0, 1]$.

You need to solve the initial value problem, $A'(x) = x^2 - x^3$, $A'(0) = 0$. The answer is then $A(1)$. By Theorem 9.5.3, $A(x) = 0 + \int_0^x (t^2 - t^3) dt$ and so the answer is

$$A(1) = \int_0^1 (t^2 - t^3) dt = \frac{1}{12}$$

Example 9.5.5 A cylinder standing on its end is filled with water which weighs 62.5 pounds per cubic foot is 5 feet in radius and is 10 feet tall. Find the total force on the sides of the cylinder.

Remember the force acting on the sides of a horizontal slice of the cylinder y feet below the top of the fluid is $(2\pi \times 5 \times 62.5 \times y) dy = dF$. Thus you need to find a solution to the initial value problem

$$\frac{dF}{dy} = (2\pi \times 5 \times 62.5 \times y), \quad F(0) = 0$$

and in this case, the answer equals $F(10)$. By Theorem 9.5.3 and Theorem 9.5.2,

$$F(10) = \int_0^{10} (2\pi)(5)(62.5)(y) dy = 98175 \text{ pounds.}$$

An informal way of looking at this is $dF = (2\pi \times 5 \times 62.5 \times y) dy$ for y between 0 and 10. The dF is an “infinitesimal” piece of the total force. To get the total force you just sum the dF 's. Thus $F = \int_0^{10} (2\pi)(5)(62.5)(y) dy$. The reason the integral sign looks like an S is that it is a sort of a sum. You are summing up “infinitesimal” contributions to obtain the total. This last statement is mathematical gobeldygook. I have told you nothing about infinitesimals. I have only used the term in an evocative manner. However, it turns out to be a useful way of thinking about things because it allows you to remember what to do without fussing with the initial value problems.

Example 9.5.6 Find the area of the surface of revolution formed by revolving $y = x^2 - \frac{1}{8} \ln x$ about the x axis for $x \in [1, 2]$.

An infinitesimal contribution is of the form

$$dA = 2\pi \left(x^2 - \frac{1}{8} \ln x \right) \sqrt{1 + \left(\frac{1}{8} \frac{16x^2 - 1}{x} \right)^2} dx = 2\pi \left(x^2 - \frac{1}{8} \ln x \right) \frac{1}{8} \frac{1 + 16x^2}{x} dx$$

Summing these with the integral and using Theorem 9.5.2 yields

$$\int_1^2 2\pi \left(x^2 - \frac{1}{8} \ln x \right) \frac{1}{8} \frac{1 + 16x^2}{x} dx = \frac{63}{4}\pi - \frac{1}{64}\pi \ln^2 2 - \pi \ln 2$$

Isn't it great that things worked out to get a nice answer? What if it hadn't. Sometimes things don't work out.

Example 9.5.7 Find the area of the surface of revolution formed by revolving $y = x^2 - \ln x$ about the x axis for $x \in [1, 2]$.

In this case an infinitesimal contribution to this area is

$$dA = 2\pi (x^2 - \ln x) \frac{1}{x} \sqrt{4x^4 - 3x^2 + 1} dx$$

and so the total area is given by the integral,

$$A = \int_1^2 2\pi (x^2 - \ln x) \frac{1}{x} \sqrt{4x^4 - 3x^2 + 1} dx.$$

What now? Can you find an antiderivative in terms of functions you have names for and use Theorem 9.5.2 to evaluate it? If you want to waste lots of time, give it a try. If you can do this one, it is very easy to find important examples that you can't do by finding antiderivatives in terms of known functions. Examples are integrals which have integrands equal to $\sin x/x$, e^{-x^2} , $\sqrt{1+x^4}$, $\sin(x^2)$, and many others. Sometimes these occur in important applications. What do you do then when Theorem 9.5.2 has failed? The answer is you use a numerical method to compute the integral. The most rudimentary numerical method goes right to the definition of the integral. It is described in the following definition. A much better method is described in the exercises. An extensive study of numerical methods for finding integrals will not be attempted in this book but it is very important you be aware that such methods are available. When you use a calculator or computer algebra system to compute an integral numerically, the machine is using a numerical method, probably one which is much more sophisticated than any presented in beginning calculus.

Definition 9.5.8 Let f be a bounded function defined on a closed interval $[a, b]$ and let $P \equiv \{x_0, \dots, x_n\}$ be a partition of the interval. Suppose $z_i \in [x_{i-1}, x_i]$ is chosen. Then the sum

$$\sum_{i=1}^n f(z_i) (x_i - x_{i-1})$$

is known as a Riemann sum. Also,

$$\|P\| \equiv \max \{|x_i - x_{i-1}| : i = 1, \dots, n\}.$$

Proposition 9.5.9 Suppose $f \in R([a, b])$. Then there exists a partition, $P \equiv \{x_0, \dots, x_n\}$ with the property that for any choice of $z_k \in [x_{k-1}, x_k]$,

$$\left| \int_a^b f(x) dx - \sum_{k=1}^n f(z_k) (x_k - x_{k-1}) \right| < \varepsilon.$$

Proof: Choose P such that $U(f, P) - L(f, P) < \varepsilon$ and then both $\int_a^b f(x) dx$ and $\sum_{k=1}^n f(z_k) (x_k - x_{k-1})$ are contained in $[L(f, P), U(f, P)]$ and so the claimed inequality must hold. This proves the proposition.

It is significant because it gives a way of approximating the integral.

Example 9.5.10 Use a Riemann sum to approximate $\int_0^1 \sin(x^2) dx$.

I will use the partition, $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ and I will pick the midpoint of each subinterval to evaluate the function. This particular Riemann sum is called the **midpoint sum**. Thus the Riemann sum approximating the integral is

$$\sin\left(\frac{1}{64}\right)\left(\frac{1}{4}\right) + \sin\left(\left(\frac{3}{8}\right)^2\right)\frac{1}{4} + \sin\left(\left(\frac{5}{8}\right)^2\right)\frac{1}{4} + \sin\left(\left(\frac{7}{8}\right)^2\right)\frac{1}{4} = .30739$$

Using a computer algebra system to evaluate this integral gives

$$\int_0^1 \sin(x^2) dx = .31027$$

so you see in this case, the primitive Riemann sum approach yielded a pretty good answer. Of course it is important to know how close you are to the true answer and this involves the concept of error estimates. The following is a rudimentary error estimate which tells how well such a mid point sum approximates an integral.

Theorem 9.5.11 *Let f be a continuous function defined on the interval, $[a, b]$ and suppose also that f' exists on (a, b) and there exists $K \geq |f'(x)|$ for all $x \in (a, b)$. Suppose also that $\{x_0, x_1, \dots, x_n\}$ is a partition in which the points are equally spaced. Thus $x_k = a + k\frac{b-a}{n}$. Then*

$$\left| \text{midpoint sum} - \int_a^b f(x) dx \right| < K \frac{(b-a)^2}{n}.$$

Proof: You estimate the error on each interval, $[x_{k-1}, x_k]$ and then sum up these estimates. The contribution of the midpoint sum on this interval equals $f\left(\frac{x_{k-1}+x_k}{2}\right)\left(\frac{b-a}{n}\right)$. Thus the error corresponding to this interval is $\left|f\left(\frac{x_{k-1}+x_k}{2}\right)\left(\frac{b-a}{n}\right) - \int_{x_{k-1}}^{x_k} f(t) dt\right|$. By the mean value theorem for integrals, Problem 11 on Page 233 there exists $z_k \in (x_{k-1}, x_k)$ such that

$$\int_{x_{k-1}}^{x_k} f(t) dt = f(z_k)(x_k - x_{k-1}) = f(z_k)\left(\frac{b-a}{n}\right).$$

Therefore, this error equals

$$\left|f\left(\frac{x_{k-1}+x_k}{2}\right) - f(z_k)\right|\left(\frac{b-a}{n}\right)$$

which by the mean value theorem is no more than

$$|f'(w_k)|\left|\frac{x_{k-1}+x_k}{2} - z_k\right|\left(\frac{b-a}{n}\right) \leq K\left(\frac{b-a}{2n}\right)\left(\frac{b-a}{n}\right).$$

Since there are n of these, it follows the total error is no more than $K\frac{(b-a)^2}{n}$ as claimed. This proves the theorem.

Note that to get an error estimate, you need information on the derivative of the function, something which has nothing to do with the existence of the integral. This is typical of error estimates in numerical methods. You need to know something extra beyond what you need to get existence.

9.6 The Riemann Integral

The definition of Riemann integrability given in this chapter is also called Darboux integrability and the integral defined as the unique number which lies between all upper sums and all lower sums which is given in this chapter is called the Darboux integral. The definition of the Riemann integral in terms of Riemann sums is given next.

Definition 9.6.1 *A bounded function, f defined on $[a, b]$ is said to be Riemann integrable if there exists a number, I with the property that for every $\varepsilon > 0$, there exists $\delta > 0$ such that if*

$$P \equiv \{x_0, x_1, \dots, x_n\}$$

is any partition having $\|P\| < \delta$, and $z_i \in [x_{i-1}, x_i]$,

$$\left| I - \sum_{i=1}^n f(z_i)(x_i - x_{i-1}) \right| < \varepsilon.$$

The number $\int_a^b f(x) dx$ is defined as I .

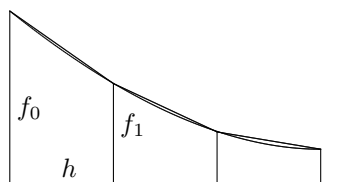
Thus, there are two definitions of the integral, this one in terms of Riemann sums and the earlier one which defined the integral to be the number which is between all the upper sums and lower sums. It turns out they are equivalent which is the following theorem of Darboux.

Theorem 9.6.2 *A bounded function defined on $[a, b]$ is Riemann integrable in the sense of Definition 9.6.1 if and only if it is integrable in the sense of Darboux. Furthermore the two integrals coincide.*

The proof of this theorem is left for the exercises in Problems 16 - 18. It isn't essential that you understand this theorem so if it does not interest you, leave it out. Note that it implies that given a Riemann integrable function f in either sense, it can be approximated by Riemann sums whenever $\|P\|$ is sufficiently small. Both versions of the integral are obsolete but entirely adequate for most applications and as a point of departure for a more up to date and satisfactory integral. The reason for using the Darboux approach to the integral is that all the existence theorems are easier to prove in this context.

9.7 Exercises

1. Let $F(x) = \int_{x^2}^{x^3} \frac{t^5+7}{t^7+87t^6+1} dt$. Find $F'(x)$.
2. Let $F(x) = \int_2^x \frac{1}{1+t^4} dt$. Sketch a graph of F and explain why it looks the way it does.
3. Use a midpoint sum with 4 subintervals to estimate $\int_1^2 \frac{1}{t} dt$ and give an estimate of how close your approximation is to the true answer.
4. Use a midpoint sum with 4 subintervals to estimate $\int_0^1 \frac{1}{t^2+1} dt$ and give an estimate of how close your approximation is to the true answer.
5. There is a general procedure for estimating the integral of a function, f on an interval, $[a, b]$. Form a uniform partition, $P = \{x_0, x_1, \dots, x_n\}$ where for each j , $x_j - x_{j-1} = h$. Let $f_i = f(x_i)$ and assuming $f \geq 0$ on the interval $[x_{i-1}, x_i]$, approximate the area above this interval and under the curve with the area of a trapezoid having vertical sides, f_{i-1} , and f_i as shown in the following picture.



Thus $\frac{1}{2} \left(\frac{f_i + f_{i-1}}{2} \right)$ approximates the area under the curve. Show that adding these up yields

$$\frac{h}{2} [f_0 + 2f_1 + \cdots + 2f_{n-1} + f_n]$$

as an approximation to $\int_a^b f(x) dx$. This is known as the trapezoid rule. Verify that if $f(x) = mx + b$, the trapezoid rule gives the exact answer for the integral. Would this be true of upper and lower sums for such a function? Can you show that in the case of the function, $f(t) = 1/t$ the trapezoid rule will always yield an answer which is too large for $\int_1^2 \frac{1}{t} dt$?

6. Let there be three equally spaced points, $x_{i-1}, x_{i-1} + h \equiv x_i$, and $x_i + 2h \equiv x_{i+1}$. Suppose also a function, f , has the value f_{i-1} at x , f_i at $x + h$, and f_{i+1} at $x + 2h$. Then consider

$$\begin{aligned} g_i(x) \equiv & \frac{f_{i-1}}{2h^2} (x - x_i)(x - x_{i+1}) - \frac{f_i}{h^2} (x - x_{i-1})(x - x_{i+1}) \\ & + \frac{f_{i+1}}{2h^2} (x - x_{i-1})(x - x_i). \end{aligned}$$

Check that this is a second degree polynomial which equals the values f_{i-1}, f_i , and f_{i+1} at the points x_{i-1}, x_i , and x_{i+1} respectively. The function, g_i is an approximation to the function, f on the interval $[x_{i-1}, x_{i+1}]$. Also,

$$\int_{x_{i-1}}^{x_{i+1}} g_i(x) dx$$

is an approximation to $\int_{x_{i-1}}^{x_{i+1}} f(x) dx$. Show $\int_{x_{i-1}}^{x_{i+1}} g_i(x) dx$ equals

$$\frac{hf_{i-1}}{3} + \frac{hf_i}{3} + \frac{hf_{i+1}}{3}.$$

Now suppose n is even and $\{x_0, x_1, \dots, x_n\}$ is a partition of the interval, $[a, b]$ and the values of a function, f defined on this interval are $f_i = f(x_i)$. Adding these approximations for the integral of f on the succession of intervals,

$$[x_0, x_2], [x_2, x_4], \dots, [x_{n-2}, x_n],$$

show that an approximation to $\int_a^b f(x) dx$ is

$$\frac{h}{3} [f_0 + 4f_1 + 2f_2 + 4f_3 + 2f_4 + \cdots + 4f_{n-1} + f_n].$$

This is called Simpson's rule. Use Simpson's rule to compute an approximation to $\int_1^2 \frac{1}{t} dt$ letting $n = 4$. Compare with the answer from a calculator or computer.

7. A mine shaft has the shape $y = \cosh(.01x) - 5000$ where the units are in feet. Thus this mine shaft has a depth of a little less than one mile occurring when $x = 0$. The surface corresponds to $y = 0$ which occurs when $x = 921$ feet (approximately). A

cable which weighs 5 pounds per foot extends from this point on the surface down to the very bottom of this mine shaft. Set up an integral which will give the work required to haul this cable up to the surface and use a computer algebra system or calculator to compute this integral. Next suppose there is a 5 ton ore skip on the bottom of this cable. How much work would be required to haul up the cable and the 5 ton skip of ore⁴?

8. Let a and b be positive numbers and consider the function,

$$F(x) = \int_0^{ax} \frac{1}{a^2 + t^2} dt + \int_b^{a/x} \frac{1}{a^2 + t^2} dt.$$

Show that F is a constant.

9. Solve the following initial value problem from ordinary differential equations which is to find a function y such that

$$y'(x) = \frac{x^7 + 1}{x^6 + 97x^5 + 7}, \quad y(10) = 5.$$

10. If $F, G \in \int f(x) dx$ for all $x \in \mathbb{R}$, show $F(x) = G(x) + C$ for some constant, C . Use this to give a different proof of the fundamental theorem of calculus which has for its conclusion $\int_a^b f(t) dt = G(b) - G(a)$ where $G'(x) = f(x)$.
11. Suppose f is continuous on $[a, b]$. Show there exists $c \in (a, b)$ such that

$$f(c) = \frac{1}{b-a} \int_a^b f(x) dx.$$

Hint: You might consider the function $F(x) \equiv \int_a^x f(t) dt$ and use the mean value theorem for derivatives and the fundamental theorem of calculus.

12. Suppose f and g are continuous functions on $[a, b]$ and that $g(x) \neq 0$ on (a, b) . Show there exists $c \in (a, b)$ such that

$$f(c) \int_a^b g(x) dx = \int_a^b f(x) g(x) dx.$$

Hint: Define $F(x) \equiv \int_a^x f(t) g(t) dt$ and let $G(x) \equiv \int_a^x g(t) dt$. Then use the Cauchy mean value theorem on these two functions.

13. Consider the function

$$f(x) \equiv \begin{cases} \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

Is f Riemann integrable? Explain why or why not.

⁴In the upper peninsula of Michigan, there are copper mines having inclined shafts which achieve a depth of roughly one mile. These shafts start off very steep and become less so toward the bottom to allow for the sagging in the cable. One such is shaft 2 for the Quincy mine in Hancock Michigan. The problem of hauling up so much heavy wire cable was solved by letting out cable and an empty ore skip on one of the dual skipways while taking up the cable and the full ore skip on the other. The Hoist which did the work of hauling the skip weighs 880 tons and is the largest steam hoist ever built. It was operated between 1920 and 1931. This hoist could bring up a 10 ton skip of ore at the rate of 16 meters per second. The huge steam hoist is housed in its own building and the cables from this hoist passed over head on large pulleys to the mine shaft. Once the cable broke. Imagine what happened at the top and at the bottom.

14. Prove the second part of Theorem 9.3.2 about decreasing functions.
15. Find the following limits by identifying them as a Riemann sum for a certain integral.

(a) $\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{n} \sin\left(\frac{k\pi}{n}\right)$

(b) $\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{n+k}$

(c) $\lim_{n \rightarrow \infty} \frac{4}{n^3} \sum_{k=1}^n k^2$

(d) $\lim_{n \rightarrow \infty} \sum_{k=1}^{2n} \frac{n}{2n^2 - 2kn + k^2}$

16. * Suppose f is a bounded function defined on $[a, b]$ and $|f(x)| < M$ for all $x \in [a, b]$. Now let Q be a partition having n points, $\{x_0^*, \dots, x_n^*\}$ and let P be any other partition. Show that

$$|U(f, P) - L(f, P)| \leq 2Mn \|P\| + |U(f, Q) - L(f, Q)|.$$

Hint: Write the sum for $U(f, P) - L(f, P)$ and split this sum into two sums, the sum of terms for which $[x_{i-1}, x_i]$ contains at least one point of Q , and terms for which $[x_{i-1}, x_i]$ does not contain any points of Q . In the latter case, $[x_{i-1}, x_i]$ must be contained in some interval, $[x_{k-1}^*, x_k^*]$. Therefore, the sum of these terms should be no larger than $|U(f, Q) - L(f, Q)|$.

17. * $\uparrow$ If $\varepsilon > 0$ is given and f is a Darboux integrable function defined on $[a, b]$, show there exists $\delta > 0$ such that whenever $\|P\| < \delta$, then

$$|U(f, P) - L(f, P)| < \varepsilon.$$

18. * $\uparrow$ Prove Theorem 9.6.2.

9.8 Return Of The Wild Assumption



The entire treatment of exponential functions and logarithms up till this time has been based on the Wild Assumption on Page 109. This was a totally unjustified assumption that exponential functions existed and were differentiable. Some pictures were drawn by a computer as evidence. Based on this outrageous wild assumption, logarithms were defined. It is now possible to establish Wild Assumption 4.3.1.

Define

$$L_1(x) \equiv \int_1^x \frac{1}{t} dt. \quad (9.19)$$

There is no problem in writing this integral because the function, $f(t) = 1/t$ is decreasing.

Theorem 9.8.1 *The function, $L_1 : (0, \infty) \rightarrow \mathbb{R}$ satisfies the following properties.*

$$L_1(xy) = L_1(x) + L_1(y), \quad L_1(1) = 0, \quad (9.20)$$

The function, L_1 is one to one and onto, strictly increasing, and its graph is concave downward. In addition to this, whenever $\frac{m}{n} \in \mathbb{Q}$

$$L_1\left(\sqrt[n]{x^m}\right) = \frac{m}{n}L_1(x). \quad (9.21)$$

Proof: Fix $y > 0$ and let

$$f(x) = L_1(xy) - (L_1(x) + L_1(y))$$

Then by Theorem 3.2.6 on Page 88 and the Fundamental theorem of calculus, Theorem 9.5.1,

$$f'(x) = y \left(\frac{1}{xy} \right) - \frac{1}{x} = 0.$$

Therefore, by Corollary 3.6.4 on Page 99, $f(x)$ is a constant. However, $f(1) = 0$ and so this proves 9.20.

From the Fundamental theorem of calculus, Theorem 9.5.1, $L_1'(x) = \frac{1}{x} > 0$ and so L_1 is a strictly increasing function and is therefore one to one. Also the second derivative equals

$$L_1''(x) = \frac{-1}{x^2} < 0$$

showing that the graph of L_1 is concave down.

Now consider the assertion that L_1 is onto. First note that from the definition, $L_1(2) > 0$. In fact,

$$L_1(2) \geq 1/2$$

as can be seen by looking at a lower sum for $\int_1^2 (1/t) dt$. Now if $x > 0$

$$L_1(x \times x) = L_1x + L_1x = 2L_1x.$$

Also,

$$\begin{aligned} L_1((x)(x)(x)) &= L_1((x)(x)) + L_1(x) \\ &= L_1(x) + L_1(x) + L_1(x) \\ &= 3L_1(x). \end{aligned}$$

Continuing in this way it follows that for any positive integer, n ,

$$L_1(x^n) = nL_1(x). \quad (9.22)$$

Therefore, $L_1(x)$ achieves arbitrarily large values as x gets increasingly large because you can take $x = 2$ in 9.22 and use the definition of L_1 to verify that $L_1(2) > 0$. Now if $x > 0$

$$0 = L_1\left(\overbrace{\left(\frac{1}{x}\right)(x)}^{=1}\right) = L_1\left(\frac{1}{x}\right) + L_1(x)$$

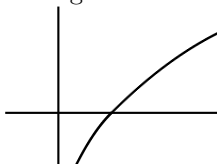
showing that

$$L_1(x^{-1}) = L_1\left(\frac{1}{x}\right) = -L_1x. \quad (9.23)$$

You see that $\left(\frac{1}{2}\right)^n$ can be made as close to zero as desired by taking n sufficiently large. Also, from 9.22,

$$L_1\left(\frac{1}{2^n}\right) = nL_1\left(\frac{1}{2}\right) = -nL_1(2)$$

showing that $L_1(x)$ gets arbitrarily large in the negative direction provided that x is a sufficiently small positive number. Since L_1 is continuous, the intermediate value theorem may be used to fill in all the numbers in between. Thus the picture of the graph of L_1 looks like the following



where the graph approaches the y axis as x gets close to 0.

It only remains to verify the claim about raising x to a rational power. From 9.22 and 9.23, for any integer, n , positive or negative,

$$L_1(x^n) = nL_1(x).$$

Therefore, letting m, n be integers,

$$mL_1(x) = L_1(x^m) = L_1\left(\left(\sqrt[n]{x^m}\right)^n\right) = nL_1\left(\sqrt[n]{x^m}\right)$$

and so

$$\frac{m}{n}L_1(x) = L_1\left(\sqrt[n]{x^m}\right).$$

This proves the theorem.

Now Wild Assumption 4.3.1 on Page 109 can be fully justified.

Definition 9.8.2 For $b > 0$ define

$$\exp_b(x) \equiv L_1^{-1}(xL_1(b)). \quad (9.24)$$

Proposition 9.8.3 The function just defined in 9.24 satisfies all conditions of Wild Assumption 4.3.1 on Page 109 and $L_1(b) = \ln b$ as defined on Page 109.

Proof: First let $x = \frac{m}{n}$ where m and n are integers. To verify that $\exp_b(m/n) = \sqrt[n]{b^m}$,

$$\exp_b\left(\frac{m}{n}\right) \equiv L_1^{-1}\left(\frac{m}{n}L_1(b)\right) = L_1^{-1}\left(L_1\left(\sqrt[n]{b^m}\right)\right) = \sqrt[n]{b^m}$$

by 9.21.

That $\exp_b(x) > 0$ follows immediately from the definition of the inverse function. (L_1 is defined on positive real numbers and so L_1^{-1} has values in the positive real numbers.)

Consider the claim that if $h \neq 0$ and $b \neq 1$, then $\exp_b(h) \neq 1$. Suppose then that $\exp_b(h) = 1$. Then doing L_1 to both sides, $hL_1(b) = L_1(1) = 0$. Hence either $h = 0$ or $b = 1$ which are both excluded.

What of the laws of exponents for arbitrary values of x and y ? As part of Wild Assumption 4.3.1, these were assumed to hold.

$$L_1(\exp_b(x+y)) = (x+y)L_1(b)$$

and

$$\begin{aligned} L_1(\exp_b(x)\exp_b(y)) &= L_1(\exp_b(x)) + L_1(\exp_b(y)) \\ &= xL_1(b) + yL_1(b) = (x+y)L_1(b). \end{aligned}$$

Since L_1 is one to one, this shows the first law of exponents holds,

$$\exp_b(x+y) = \exp_b(x) \exp_b(y).$$

From the definition, $\exp_b(1) = b$ and $\exp_b(0) = 1$. Therefore,

$$1 = \exp_b(1 + (-1)) = \exp_b(-1) \exp_b(1) = \exp_b(-1) b$$

showing that $\exp_b(-1) = b^{-1}$. Now

$$L_1(\exp_{ab}(x)) = xL_1(ab) = xL_1(a) + xL_1(b)$$

while

$$\begin{aligned} L_1(\exp_a(x) \exp_b(x)) &= L_1(\exp_a(x)) + L_1(\exp_b(x)) \\ &= xL_1(a) + xL_1(b). \end{aligned}$$

Again, since L_1 is one to one, $\exp_{ab}(x) = \exp_a(x) \exp_b(x)$. Finally,

$$L_1(\exp_{\exp_b(x)}(y)) = yL_1(\exp_b(x)) = yxL_1(b)$$

while

$$L_1(\exp_b(xy)) = xyL_1(b)$$

and so since L_1 is one to one, $\exp_{\exp_b(x)}(y) = \exp_b(xy)$. This establishes all the laws of exponents for arbitrary real values of the exponent.

$\exp'_b$ exists by Theorem 5.1.6 on Page 119. Therefore, $\exp_b$ defined in 9.24 satisfies all the conditions of Wild Assumption 4.3.1.

It remains to consider the derivative of $\exp_b$ and verify $L_1(b) = \ln b$. First,

$$L_1(L_1^{-1}(x)) = x$$

and so

$$L'_1(L_1^{-1}(x)) (L_1^{-1})'(x) = 1.$$

By 9.19,

$$\frac{(L_1^{-1})'(x)}{L_1^{-1}(x)} = 1$$

showing that $(L_1^{-1})'(x) = (L_1^{-1})(x)$. Now by the chain rule,

$$\begin{aligned} \exp'_b(x) &= (L_1^{-1})'(xL_1(b)) L_1(b) \\ &= L_1^{-1}(xL_1(b)) L_1(b) \\ &= \exp_b(x) L_1(b). \end{aligned}$$

From the definition of $\ln b$ on Page 109, it follows $\ln = L_1$.

9.9 Exercises

1. Show $\ln 2 \in [.5, 1]$.
2. Apply the trapezoid rule, (See Problem 5 on Page 231) to estimate $\ln 2$ in the case where $h = 1/5$. Now use a calculator or table to find the exact value of $\ln 2$.

3. Suppose it is desired to find a function, $L : (0, \infty) \rightarrow \mathbb{R}$ which satisfies

$$L(xy) = Lx + Ly, \quad L(1) = 0. \quad (9.25)$$

Show the only differentiable solutions to this equation are functions of the form $L_k(x) = \int_1^x \frac{k}{t} dt$. **Hint:** Fix $x > 0$ and differentiate both sides of the above equation with respect to y . Then let $y = 1$.

4. Recall that $\ln e = 1$. In fact, this was how e was defined. Show that

$$\lim_{y \rightarrow 0+} (1 + yx)^{1/y} = e^x.$$

Hint: Consider $\ln(1 + yx)^{1/y} = \frac{1}{y} \ln(1 + yx) = \frac{1}{y} \int_1^{1+yx} \frac{1}{t} dt$, use upper and lower sums and then the squeezing theorem to verify $\ln(1 + yx)^{1/y} \rightarrow x$. Recall that $x \rightarrow e^x$ is continuous.

5. Logarithms were invented before calculus, one of the inventors being Napier, a Scottish nobleman. His interest in logarithms was computational. Describe how one could use logarithms to find 7^{th} roots for example. Also describe how one could use logarithms to do computations involving large numbers. Describe how you could construct a table for $\log_{10} x$ for x various numbers. Next, how do you suppose Napier did it? Remember, he did not have calculus. **Hint:** $\log_b(35^{1/7}) = \frac{1}{7} \log_b(35)$. You can find the logarithm of the number you are after. If you had a table of logarithms to some base, you could then see what number this corresponded to. Napier essentially found a table of logarithms to the base $b = .999999$ or some such number like that. The reason for the strange choice is that successive integer powers of this number are quite close to each other and the graph of the function b^x is close to a straight line between two successive integers. This is not true for a number like 10. Shortly after calculus was invented, it was realized that the best way to consider logarithms was in terms of integrals. This occurred in the 1600's.

9.10 Techniques Of Integration

The techniques for finding antiderivatives may be used to find integrals.

9.10.1 The Method Of Substitution

Recall

$$\int f(g(x)) g'(x) dx = F(x) + C, \quad (9.26)$$

where $F'(y) = f(y)$.

How does this relate to finding definite integrals? This is based on the following formula in which all the functions are integrable and $F'(y) = f(y)$.

$$\int_a^b f(g(x)) g'(x) dx = \int_{g(a)}^{g(b)} f(y) dy. \quad (9.27)$$

This formula follows from the observation that, by the fundamental theorem of calculus, both sides equal $F(g(b)) - F(g(a))$.

How can you remember this? The easiest way is to use the Leibniz notation. In 9.27 let $y = g(x)$. Then

$$\frac{dy}{dx} = g'(x)$$

and so formally $dy = g'(x) dx$. Then making the substitution

$$\int_a^b \overbrace{f(g(x))}^{f(y)} \overbrace{g'(x) dx}^{dy} = \int_{\gamma}^? f(y) dy.$$

What should go in as the top and bottom limits of the integral? The important thing to remember is that **if you change the variable, you must change the limits!** When $x = a$, it follows that $y = g(a)$ to the bottom limit must equal $g(a)$. Similarly the top limit should be $g(b)$.

Example 9.10.1 Find $\int_1^2 x \sin(x^2) dx$

Let $u = x^2$ so $du = 2x dx$ and so $\frac{du}{2} = x dx$. Therefore, changing the variables gives

$$\int_1^2 x \sin(x^2) dx = \frac{1}{2} \int_1^4 \sin(u) du = -\frac{1}{2} \cos(4) + \frac{1}{2} \cos(1)$$

Sometimes people prefer not to worry about the limits. This is fine provided you don't write anything which is false. The above problem can be done in the following way.

$$\int x \sin(x^2) dx = -\frac{1}{2} \cos(x^2) + C$$

and so from an application of the fundamental theorem of calculus

$$\begin{aligned} \int_1^2 x \sin(x^2) dx &= -\frac{1}{2} \cos(x^2) \Big|_1^2 \\ &= -\frac{1}{2} \cos(4) + \frac{1}{2} \cos(1). \end{aligned}$$

Example 9.10.2 Find the area of the ellipse,

$$\frac{(y - \beta)^2}{b^2} + \frac{(x - \alpha)^2}{a^2} = 1.$$

If you sketch the ellipse, you see that it suffices to find the area of the top right quarter for $y \geq \beta$ and $x \geq \alpha$ and multiply by 4 since the bottom half is just a reflection of the top half about the line, $y = \beta$ and the left top quarter is just the reflection of the top right quarter reflected about the line, $x = \alpha$. Thus the area of the ellipse is

$$4 \int_{\alpha}^{\alpha+a} b \sqrt{1 - \frac{(x - \alpha)^2}{a^2}} dx$$

Change the variables, letting $u = \frac{x - \alpha}{a}$. Then $du = \frac{1}{a} dx$ and so upon changing the limits to correspond to the new variables, this equals

$$4ba \int_0^1 \overbrace{\sqrt{1 - u^2}}^{\pi/4} du = 4 \times ab \times \frac{1}{4} \pi = \pi ab$$

because the integral in the above is just one quarter of the unit circle and so has area equal to $\pi/4$.

9.10.2 Integration By Parts

Recall the following proposition for finding antiderivative.

Proposition 9.10.3 *Let u and v be differentiable functions for which*

$$\int u(x) v'(x) dx, \int u'(x) v(x) dx$$

are nonempty. Then

$$uv - \int u'(x) v(x) dx = \int u(x) v'(x) dx. \quad (9.28)$$

In terms of integrals, this is stated in the following proposition.

Proposition 9.10.4 *Let u and v be differentiable functions on $[a, b]$ such that $uv', u'v \in R([a, b])$. Then*

$$\int_a^b u(x) v'(x) dx = uv(x) \Big|_a^b - \int_a^b u'(x) v(x) dx \quad (9.29)$$

Proof: Use the product rule and properties of integrals to write

$$\begin{aligned} \int_a^b u(x) v'(x) dx &= \int_a^b (uv)'(x) dx - \int_a^b u'(x) v(x) dx \\ &= uv(x) \Big|_a^b - \int_a^b u'(x) v(x) dx. \end{aligned}$$

This proves the proposition.

Example 9.10.5 *Find $\int_0^\pi x \sin(x) dx$*

Let $u(x) = x$ and $v'(x) = \sin(x)$. Then applying 9.28,

$$\begin{aligned} \int_0^\pi x \sin(x) dx &= (-\cos(x)) x \Big|_0^\pi - \int_0^\pi (-\cos(x)) dx \\ &= -\pi \cos(\pi) = \pi. \end{aligned}$$

Example 9.10.6 *Find $\int_0^1 x e^{2x} dx$*

Let $u(x) = x$ and $v'(x) = e^{2x}$. Then from 9.29

$$\begin{aligned} \int_0^1 x e^{2x} dx &= \frac{e^{2x}}{2} x \Big|_0^1 - \int_0^1 \frac{e^{2x}}{2} dx \\ &= \frac{e^{2x}}{2} x \Big|_0^1 - \frac{e^{2x}}{4} \Big|_0^1 \\ &= \frac{e^2}{2} - \left(\frac{e^2}{4} - \frac{1}{4} \right) \\ &= \frac{e^2}{4} + \frac{1}{4} \end{aligned}$$

9.11 Exercises

1. Find the integrals.

- (a) $\int_0^4 x e^{-3x} dx$
- (b) $\int_2^3 \frac{1}{x(\ln(|x|))^2} dx$
- (c) $\int_0^1 x \sqrt{2-x} dx$
- (d) $\int_2^3 (\ln |x|)^2 dx$ **Hint:** Let $u(x) = (\ln |x|)^2$ and $v'(x) = 1$.
- (e) $\int_0^\pi x^3 \cos(x^2) dx$

2. Find $\int_1^2 x \ln(x^2) dx$

3. Find $\int_0^1 e^x \sin(x) dx$

4. Find $\int_0^1 2^x \cos(x) dx$

5. Find $\int_0^2 x^3 \cos(x) dx$

6. Find the integrals.

- (a) $\int_5^6 \frac{x}{\sqrt{2x-3}} dx$
- (b) $\int_2^4 x(3x^2 + 6)^5 dx$
- (c) $\int_0^\pi x \sin(x^2) dx$
- (d) $\int_0^{\pi/4} \sin^3(2x) \cos(2x) dx$
- (e) $\int_0^7 \frac{1}{\sqrt{1+4x^2}} dx$ **Hint:** Remember the $\sinh^{-1}$ function and its derivative.

7. Find the integrals.

- (a) $\int_0^{\pi/9} \sec(3x) dx$
- (b) $\int_0^{\pi/9} \sec^2(3x) \tan(3x) dx$
- (c) $\int_0^5 \frac{1}{3+5x^2} dx$
- (d) $\int_0^1 \frac{1}{\sqrt{5-4x^2}} dx$
- (e) $\int_2^6 \frac{3}{x\sqrt{4x^2-5}} dx$

8. Find the integrals.

- (a) $\int_0^3 x \cosh(x^2 + 1) dx$
- (b) $\int_0^2 x^3 5^{x^4} dx$
- (c) $\int_{-\pi}^\pi \sin(x) 7^{\cos(x)} dx$
- (d) $\int_0^\pi x \sin(x^2) dx$
- (e) $\int_1^2 x^5 \sqrt{2x^2 + 1} dx$ **Hint:** Let $u = 2x^2 + 1$.

9. Find $\int_0^{\pi/4} \sin^2(x) dx$. **Hint:** Derive and use $\sin^2(x) = \frac{1 - \cos(2x)}{2}$.

10. Find the area between the graphs of $y = \sin(2x)$ and $y = \cos(2x)$ for $x \in [0, 2\pi]$.

11. Find the following integrals.

- (a) $\int_1^\pi x^2 \sin(x^3) dx$
- (b) $\int_1^6 \frac{x}{1+x^2} dx$
- (c) $\int_0^5 \frac{1}{1+4x^2} dx$
- (d) $\int_1^4 x^2 \sqrt{1+x} dx$

12. The most important of all differential equations is the first order linear equation, $y' + p(t)y = f(t)$. Show the solution to the initial value problem consisting of this equation and the initial condition, $y(a) = y_a$ is

$$y(t) = e^{-P(t)} y_a$$

$$+ e^{-P(t)} \int_a^t e^{P(s)} f(s) ds,$$

where $P(t) = \int_a^t p(s) ds$. Give conditions under which everything is correct. **Hint:** You use the integrating factor approach. Multiply both sides by $e^{P(t)}$, verify the left side equals

$$\frac{d}{dt} (e^{P(t)} y(t)),$$

and then take the integral, $\int_a^t$ of both sides.

13. Suppose f is a continuous function which is not equal to zero on $[0, b]$. Show that

$$\int_0^b \frac{f(x)}{f(x) + f(b-x)} dx = \frac{b}{2}.$$

Hint: First change the variables to obtain the integral equals

$$\int_{-b/2}^{b/2} \frac{f(y + b/2)}{f(y + b/2) + f(b/2 - y)} dy$$

Next show by another change of variables that this integral equals

$$\int_{-b/2}^{b/2} \frac{f(b/2 - y)}{f(y + b/2) + f(b/2 - y)} dy.$$

Thus the sum of these equals b .

14. Suppose $x_0 \in (a, b)$ and that f is a function which has $n + 1$ continuous derivatives on this interval. Consider the following.

$$\begin{aligned} f(x) &= f(x_0) + \int_{x_0}^x f'(t) dt \\ &= f(x_0) + (t - x) f'(t) \Big|_{x_0}^x \\ &\quad + \int_{x_0}^x (x - t) f''(t) dt \\ &= f(x_0) + f'(x_0)(x - x_0) \\ &\quad + \int_{x_0}^x (x - t) f''(t) dt. \end{aligned}$$

Explain the above steps and continue the process to eventually obtain Taylor's formula,

$$\begin{aligned} f(x) &= \\ f(x_0) &+ \sum_{k=1}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k \\ &+ \frac{1}{n!} \int_{x_0}^x (x - t)^n f^{(n+1)}(t) dt \end{aligned}$$

where $n! \equiv n(n-1) \cdots 3 \cdot 2 \cdot 1$ if $n \geq 1$ and $0! \equiv 1$.

15. In the above Taylor's formula, use Problem 12 on Page 233 to obtain the existence of some z between x_0 and x such that

$$\begin{aligned} f(x) &= \\ f(x_0) &+ \sum_{k=1}^n \frac{f^{(k)}(x_0)}{k!} (x - x_0)^k \\ &+ \frac{f^{(n+1)}(z)}{(n+1)!} (x - x_0)^{n+1}. \end{aligned}$$

Hint: You might consider two cases, the case when $x > x_0$ and the case when $x < x_0$.

16. * There is a general procedure for constructing these methods of approximate integration like the trapezoid rule and Simpson's rule. Consider $[0, 1]$ and divide this interval into n pieces using a uniform partition, $\{x_0, \dots, x_n\}$ where $x_i - x_{i-1} =$

$1/n$ for each i . The approximate integration scheme for a function, f , will be of the form

$$\left(\frac{1}{n}\right) \sum_{i=0}^n c_i f_i \approx \int_0^1 f(x) dx$$

where $f_i = f(x_i)$ and the constants, c_i are chosen in such a way that the above sum gives the exact answer for $\int_0^1 f(x) dx$ where $f(x) = 1, x, x^2, \dots, x^n$. When this has been done, change variables to write

$$\begin{aligned} &\int_a^b f(y) dy \\ &= (b-a) \int_0^1 f(a + (b-a)x) dx \\ &\approx \frac{b-a}{n} \sum_{i=1}^n c_i f\left(a + (b-a)\left(\frac{i}{n}\right)\right) \\ &= \frac{b-a}{n} \sum_{i=1}^n c_i f_i \end{aligned}$$

where $f_i = f\left(a + (b-a)\left(\frac{i}{n}\right)\right)$. Consider the case where $n = 1$. It is necessary to find constants c_0 and c_1 such that

$$\begin{aligned} c_0 + c_1 &= 1 = \int_0^1 1 dx \\ 0c_0 + c_1 &= 1/2 = \int_0^1 x dx. \end{aligned}$$

Show that $c_0 = c_1 = 1/2$, and that this yields the trapezoid rule. Next take $n = 2$ and show the above procedure yields Simpson's rule. Show also that if this integration scheme is applied to any polynomial of degree 3 the result will be exact. That is,

$$\begin{aligned} &\frac{1}{2} \left(\frac{1}{3} f_0 + \frac{4}{3} f_1 + \frac{1}{3} f_2 \right) \\ &= \int_0^1 f(x) dx \end{aligned}$$

whenever $f(x)$ is a polynomial of degree three. Show that if f_i are the values of f at $a, \frac{a+b}{2}$, and b with $f_1 = f\left(\frac{a+b}{2}\right)$, it follows that the above formula gives $\int_a^b f(x) dx$ exactly whenever f is a polynomial of degree three. Obtain an integration scheme for $n = 3$.

17. * Let f have four continuous derivatives on $[x_{i-1}, x_{i+1}]$ where $x_{i+1} = x_{i-1} + 2h$ and $x_i = x_{i-1} + h$. Show using Problem 15, there exists a polynomial of degree three, $p_3(x)$, such that

$$f(x) = p_3(x) + \frac{1}{4!} f^{(4)}(\xi)(x - x_i)^4$$

Now conclude

$$\left| \int_{x_{i-1}}^{x_{i+1}} f(x) dx - \left(\frac{hf_{i-1}}{3} + \frac{hf_i}{3} + \frac{hf_{i+1}}{3} \right) \right| < \frac{10}{3} \frac{M}{4!} h^5,$$

where M satisfies, $M \geq \max \{ |f^{(4)}(t)| : t \in [x_{i-1}, x_i] \}$. Now let $S(a, b, f, 2m)$ denote the approximation to $\int_a^b f(x) dx$ obtained from Simpson's rule using $2m$ equally spaced points. Show

$$\left| \int_a^b f(x) dx - S(a, b, f, 2m) \right| < C(b-a)^5 \frac{1}{m^4}$$

where C is some constant depending on $M \equiv \max \{ |f^{(4)}(t)| : t \in [a, b] \}$.

18. * A **regular Sturm Liouville problem** involves the differential equation, for an unknown function of x which is denoted here by y ,

$$(p(x)y')' + (\lambda q(x) + r(x))y = 0,$$

$$x \in [a, b]$$

and it is assumed that $p(t), q(t) > 0$ for any t along with boundary conditions,

$$\begin{aligned} C_1 y(a) + C_2 y'(a) &= 0 \\ C_3 y(b) + C_4 y'(b) &= 0 \end{aligned}$$

where

$$C_1^2 + C_2^2 > 0, \text{ and } C_3^2 + C_4^2 > 0.$$

There is an immense theory connected to these important problems. The constant, λ is called an eigenvalue. Show that if y is a solution to the above problem corresponding to $\lambda = \lambda_1$ and if z is a solution corresponding to $\lambda = \lambda_2 \neq \lambda_1$, then

$$\int_a^b q(x)y(x)z(x)dx = 0. \quad (9.30)$$

Hint: Do something like this:

$$\begin{aligned} &(p(x)y')'z + \\ &(\lambda_1 q(x) + r(x))yz = 0, \\ &(p(x)z')'y + \\ &(\lambda_2 q(x) + r(x))zy = 0. \end{aligned}$$

Now subtract and either use integration by parts or show

$$\begin{aligned} &(p(x)y')'z - (p(x)z')'y \\ &= ((p(x)y')z - (p(x)z')y)' \end{aligned}$$

and then integrate. Use the boundary conditions to show that $y'(a)z(a) - z'(a)y(a) = 0$ and $y'(b)z(b) - z'(b)y(b) = 0$. The formula, 9.30 is called an orthogonality relation and it makes possible an expansion in terms of certain functions called eigenfunctions.

19. Letting $[a, b] = [-\pi, \pi]$, consider an example of a regular Sturm Liouville problem which is of the form

$$y'' + \lambda y = 0, y(-\pi) = 0, y(\pi) = 0.$$

Show that if $\lambda = n^2$ and $y_n(x) = \sin(nx)$ for n a positive integer, then y_n is a solution to this regular Sturm Liouville problem. In this case, $q(x) = 1$ and so from Problem 18, it must be the case that

$$\int_{-\pi}^{\pi} \sin(nx) \sin(mx) dx = 0$$

if $n \neq m$. Show directly using integration by parts that the above equation is true.

9.12 Improper Integrals

The integral is only defined for certain bounded functions which are defined on closed and bounded intervals. Nevertheless people do consider things like the following: $\int_0^\infty f(t) dt$. Whenever things like this occur they require a special definition. They are called **improper integrals**. In this section a few types of improper integrals will be discussed.

Definition 9.12.1 The symbol $\int_a^\infty f(t) dt$ is defined to equal

$$\lim_{R \rightarrow \infty} \int_a^R f(t) dt$$

whenever this limit exists. If $\lim_{x \rightarrow a+} f(t) = \pm\infty$ but f is integrable on $[a + \delta, b]$ for all small δ , then

$$\int_a^b f(t) dt \equiv \lim_{\delta \rightarrow 0+} \int_{a+\delta}^b f(t) dt$$

whenever this limit exists. Similarly, if $\lim_{x \rightarrow b-} f(t) = \pm\infty$ but f is integrable on $[a, b - \delta]$ for all small δ , then

$$\int_a^b f(t) dt \equiv \lim_{\delta \rightarrow 0+} \int_a^{b-\delta} f(t) dt$$

whenever the limit exists. Finally, if $\lim_{x \rightarrow a+} f(t) = \pm\infty$, then

$$\int_a^\infty f(t) dt \equiv \lim_{R \rightarrow \infty} \int_a^R f(t) dt$$

where the improper integral, $\int_a^R f(t) dt$ is defined above as $\lim_{\delta \rightarrow 0+} \int_{a+\delta}^R f(t) dt$.

You can probably construct other examples of improper integrals such as integrals of the form $\int_{-\infty}^a f(t) dt$. The definitions are analogous to the above.

Example 9.12.2 Find $\int_0^\infty e^{-t} dt$.

From the definition, this equals $\lim_{R \rightarrow \infty} \int_0^R e^{-t} dt = \lim_{R \rightarrow \infty} (1 - e^{-R}) = 1$.

Example 9.12.3 Find $\int_0^1 \frac{1}{\sqrt{x}} dx$.

From the definition this equals $\lim_{\delta \rightarrow 0+} \int_\delta^1 x^{-1/2} dx = \lim_{\delta \rightarrow 0+} (2 - 2\sqrt{\delta}) = 2$.

Sometimes you can argue the improper integral exists even though you can't find it. The following theorem is about this question of existence.

Theorem 9.12.4 Suppose $f(t) \geq 0$ for all $t \in [a, \infty)$ and that $f \in R([a + \delta, R])$ whenever $\delta > 0$ is small enough, for every $R > a$. Suppose also there exists a number, M such that for all $R > a$, the integral $\int_a^R f(t) dt$ exists and

$$\int_a^R f(t) dt \leq M. \quad (9.31)$$

Then $\int_a^\infty f(t) dt$ exists.

If $f(t) \geq 0$ for all $t \in (a, b]$ and there exists M such that

$$\int_{a+\delta}^b f(t) dt \leq M \quad (9.32)$$

for all $\delta > 0$, then $\int_a^b f(t) dt$ exists.

If $f(t) \geq 0$ for all $t \in [a, b]$ and there exists M such that

$$\int_a^{b-\delta} f(t) dt \leq M \quad (9.33)$$

for all $\delta > 0$, then $\int_a^b f(t) dt$ exists.

Proof: Suppose 9.31. Then $I \equiv \text{lub} \left\{ \int_a^R f(t) dt : R > a \right\} \leq M$. It follows that if $\varepsilon > 0$ is given, there exists R_0 such that $\int_a^{R_0} f(t) dt \in (I - \varepsilon, I]$. Then, since $f(t) \geq 0$, it follows that for $R \geq R_0$, and small $\delta > 0$,

$$\int_{a+\delta}^R f(t) dt = \int_{a+\delta}^{R_0} f(t) dt + \int_{R_0}^R f(t) dt.$$

Letting $\delta \rightarrow 0+$,

$$\int_a^R f(t) dt = \int_a^{R_0} f(t) dt + \int_{R_0}^R f(t) dt \geq \int_a^{R_0} f(t) dt.$$

Therefore, whenever $R > R_0$, $\left| \int_a^R f(t) dt - I \right| < \varepsilon$. Since ε is arbitrary, the conditions for

$$I = \lim_{R \rightarrow \infty} \int_a^R f(t) dt$$

are satisfied and so $I = \int_a^\infty f(t) dt$.

Now suppose 9.32. Then $I \equiv \sup \left\{ \int_{a+\delta}^b f(t) dt : \delta > 0 \right\} \leq M$. It follows that if $\varepsilon > 0$ is given, there exists $\delta_0 > 0$ such that

$$\int_{a+\delta_0}^b f(t) dt \in (I - \varepsilon, I].$$

Therefore, if $\delta < \delta_0$,

$$I - \varepsilon < \int_{a+\delta_0}^b f(t) dt \leq \int_{a+\delta}^b f(t) dt \leq I$$

showing that for such δ , $\left| \int_{a+\delta}^b f(t) dt - I \right| < \varepsilon$. This is what is meant by the expression

$$\lim_{\delta \rightarrow 0+} \int_{a+\delta}^b f(t) dt = I$$

and so $I = \int_a^b f(t) dt$.

The last case is entirely similar to this one. This proves the theorem.

There is an easy way to use known convergence of some integrals to verify convergence of others. It is called the limit comparison test.

Theorem 9.12.5 Suppose $f(x), g(x) \geq 0$ for all $x \geq a$ and f, g are Riemann integrable on every interval of the form $[a, b]$. Suppose also that

$$\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = L$$

where L is a positive number. Then $\int_a^\infty f(x) dx$ and $\int_a^\infty g(x) dx$ converge or diverge together. In case the functions may be unbounded at a , then if

$$\lim_{x \rightarrow a+} \frac{f(x)}{g(x)} = L$$

where L is a positive number, then $\int_a^b f(x) dx$ and $\int_a^b g(x) dx$ converge or diverge together. The same can be said if the functions may be unbounded at b except in this case, the condition is

$$\lim_{x \rightarrow b-} \frac{f(x)}{g(x)} = L.$$

Proof: Suppose the limit condition and that $\int_a^\infty f(x) dx$ converges. Then from the limit condition, it follows that there exists R_1 such that if $x > R_1$, then

$$\frac{f(x)}{g(x)} > \frac{L}{2}$$

Therefore, if $R > R_1$,

$$\int_a^R g(x) dx \leq \int_a^{R_1} g(x) dx + \frac{2}{L} \int_{R_1}^R f(x) dx \leq \int_a^{R_1} g(x) dx + \frac{2}{L} \int_a^\infty f(x) dx \equiv M$$

Therefore, from Theorem 9.12.4 $\int_a^\infty g(x) dx$ exists. Next suppose that $\int_a^\infty f(x) dx$ diverges. Since $f(x) \geq 0$, this can only happen if $\lim_{R \rightarrow \infty} \int_a^R f(x) dx = \infty$. There exists R_1 such that if $x > R_1$, then

$$\frac{f(x)}{g(x)} < 2L$$

This must occur because of the limit condition. Therefore, for all $R > R_1$,

$$\int_{R_1}^R f(x) dx \leq 2L \int_{R_1}^R g(x) dx$$

and so, since $\lim_{R \rightarrow \infty} \int_{R_1}^R f(x) dx = \infty$, the same must be true of $\int_{R_1}^R g(x) dx$. The proof of the last assertion is similar. For $\delta > 0$ small enough, then for $\varepsilon < \delta$

$$\frac{L}{2} \int_{a+\varepsilon}^{a+\delta} g(x) dx \leq \int_{a+\varepsilon}^{a+\delta} f(x) dx < 2L \int_{a+\varepsilon}^{a+\delta} g(x) dx$$

and so $\int_a^{a+\delta} f(x) dx$ and $\int_a^{a+\delta} g(x) dx$ have the same convergence properties. The case where the functions are unbounded at b is similar. This proves the theorem.

Example 9.12.6 Does $\int_1^\infty \sin\left(\frac{1}{x^2}\right) dx$ exist?

Yes, it does. To see this note $\lim_{x \rightarrow \infty} \sin\left(\frac{1}{x^2}\right)/1/x^2 = 1$. Therefore, from the limit comparison test, this integral exists because $\int_1^\infty \frac{1}{x^2} dx$ exists. Note I had absolutely no way of finding the antiderivative of $\sin\left(\frac{1}{x^2}\right)$.

Example 9.12.7 Does $\int_0^1 \frac{1}{\sqrt{\sin x}} dx$ exist?

I don't know how to find an antiderivative for this function but the question of existence can still be resolved. Since $\lim_{x \rightarrow 0+} \frac{x}{\sin x} = 1$, it follows that for x small enough, $\frac{x}{\sin x} < \frac{3}{2}$, say for $x < \delta_1$. Then for such x , it follows

$$\frac{2}{3}x < \sin x$$

and so if $\delta < \delta_1$,

$$\int_{\delta}^1 \frac{1}{\sqrt{\sin x}} dx \leq \int_{\delta}^{\delta_1} \sqrt{\frac{3}{2}} \frac{1}{\sqrt{x}} dx + \int_{\delta_1}^1 \frac{1}{\sqrt{\sin \delta_1}} dx.$$

Now using the argument of Example 9.12.3, the first integral in the above is bounded above by $(\sqrt{\delta_1} - \sqrt{\delta})\sqrt{6}$. The second integral equals $\frac{1-\delta_1}{\sqrt{\sin \delta_1}}$. Therefore, the improper integral exists because the conditions of Theorem 9.12.4 with

$$M = \frac{1-\delta_1}{\sqrt{\sin \delta_1}} + (\sqrt{\delta_1} - \sqrt{\delta})\sqrt{6}.$$

Example 9.12.8 The gamma function is defined by $\Gamma(\alpha) \equiv \int_0^{\infty} e^{-t} t^{\alpha-1} dt$ whenever $\alpha > 0$. Does the improper integral exist?

You should supply the details to the following estimate in which δ is a small positive number less than 1 and R is a large positive number.

$$\begin{aligned} \int_{\delta}^R e^{-t} t^{\alpha-1} dt &\leq \int_{\delta}^k e^{-t} t^{\alpha-1} dt + \int_k^R e^{-t} t^{\alpha-1} dt \\ &\leq \int_0^k t^{\alpha-1} dt + \int_k^{\infty} e^{-t/2} dt. \end{aligned}$$

Here k is chosen such that if $t \geq k$,

$$e^{-t} t^{\alpha-1} < e^{-t/2}.$$

Such a k exists because

$$\lim_{t \rightarrow \infty} \frac{e^{-t} t^{\alpha-1}}{e^{-t/2}} = 0.$$

Therefore, let $M \equiv \int_0^k t^{\alpha-1} dt + \int_k^{\infty} e^{-t/2} dt$ and this shows from the above theorem that

$$\int_0^R e^{-t} t^{\alpha-1} dt \leq M$$

for all large R and so $\int_0^{\infty} e^{-t} t^{\alpha-1} dt$ exists.

Sometimes the existence of the improper integral is a little more subtle. This is the case when functions are not all the same sign for example.

Example 9.12.9 Does $\int_0^{\infty} \frac{\sin x}{x} dx$ exist?

You should verify $\int_0^1 \frac{\sin x}{x} dx$ exists and that

$$\begin{aligned} \int_0^R \frac{\sin x}{x} dx &= \int_0^1 \frac{\sin x}{x} dx + \int_1^R \frac{\sin x}{x} dx \\ &= \int_0^1 \frac{\sin x}{x} dx + \cos 1 - \frac{\cos R}{R} - \int_1^R \frac{\cos x}{x^2} dx. \end{aligned}$$

Thus the improper integral exists if it can be shown that $\int_1^\infty \frac{\cos x}{x^2} dx$ exists. However,

$$\int_1^R \frac{\cos x}{x^2} dx = \int_1^R \frac{\cos x + |\cos x|}{x^2} dx - \int_1^R \frac{(|\cos x| - \cos x)}{x^2} dx \quad (9.34)$$

and

$$\begin{aligned} \int_1^R \frac{\cos x + |\cos x|}{x^2} dx &\leq \int_1^R \frac{2}{x^2} dx \leq \int_0^\infty \frac{2}{x^2} dx < \infty \\ \int_1^R \frac{(|\cos x| - \cos x)}{x^2} dx &\leq \int_1^R \frac{2}{x^2} dx \leq \int_0^\infty \frac{2}{x^2} dx < \infty. \end{aligned}$$

Since both integrands are positive, Theorem 9.12.4 applies and the limits

$$\lim_{R \rightarrow \infty} \int_1^R \frac{(|\cos x| - \cos x)}{x^2} dx, \quad \lim_{R \rightarrow \infty} \int_1^R \frac{\cos x + |\cos x|}{x^2} dx$$

both exist and so from 9.34 $\lim_{R \rightarrow \infty} \int_1^R \frac{\cos x}{x^2} dx$ also exists and so $\int_0^\infty \frac{\sin x}{x} dx$ exists.

This is an important example. There are at least two ways to show that $\int_0^\infty \frac{\sin x}{x} dx = \frac{1}{2}\pi$. However, they involve techniques which will not be discussed in this book. It is a standard problem in the subject of complex analysis. The above argument is a special case of the following corollary to Theorem 9.12.4.

Definition 9.12.10 Let f be a real valued function. Then $f^+(t) \equiv \frac{|f(t)| + f(t)}{2}$ and $f^-(t) \equiv \frac{|f(t)| - f(t)}{2}$. Thus $|f(t)| = f^+(t) + f^-(t)$ and $f(t) = f^+(t) - f^-(t)$ while both f^+ and f^- are nonnegative functions.

Corollary 9.12.11 Suppose f is a real valued function, Riemann integrable on every finite interval, and the conditions of Theorem 9.12.4 hold for both f^+ and f^- . Then $\int_0^\infty f(t) dt$ exists.

Corollary 9.12.12 Suppose f is a real valued function, Riemann integrable on every finite interval, and $\int_0^\infty |f(x)| dx$ exists. Then $\int_0^\infty f(x) dx$ exists.

Proof: $0 \leq f^+(x) \leq |f(x)|$, and $0 \leq f^-(x) \leq |f(x)|$ and so for every $R > 0$,

$$\int_0^R f^+(x) dx \leq \int_0^\infty |f(x)| dx$$

and

$$\int_0^R f^-(x) dx \leq \int_0^\infty |f(x)| dx$$

and so it follows from Theorem 9.12.4 that $\int_0^\infty f^+(x) dx$ and $\int_0^\infty f^-(x) dx$ both exist. Therefore,

$$\int_0^\infty f(x) dx \equiv \lim_{R \rightarrow \infty} \left(\int_0^R f^+(x) dx - \int_0^R f^-(x) dx \right)$$

also exists.

Example 9.12.13 Does $\int_0^\infty \cos(x^2) dx$ exist?

This is called a **Fresnel integral** and it has also been evaluated exactly using techniques from complex analysis. In fact $\int_0^\infty \cos(x^2) dx = \frac{1}{4}\sqrt{2}\sqrt{\pi}$. The verification that this integral exists is left to you. First change the variable letting $x^2 = u$ and then integrate by parts. You will eventually get an integral of the form $\int_0^\infty \frac{\sin u}{u^{3/2}} du$. Now

consider $\int_{\delta}^1 \frac{\sin u}{u^{3/2}} du$ where δ is a small positive number. On $[\delta, 1]$, $\sin u$ is nonnegative. You also know that $\sin u \leq u$. Therefore, $\frac{\sin u}{u^{3/2}} \leq \frac{1}{u^{1/2}}$. Thus

$$\int_{\delta}^1 \frac{\sin u}{u^{3/2}} du \leq \int_{\delta}^1 \frac{1}{u^{1/2}} du = 2 - 2\delta^{1/2} \leq 2.$$

Therefore, $\int_0^1 \frac{\sin u}{u^{3/2}} du$ exists. Similarly, $\int_0^R \frac{\sin u}{u^{3/2}} du$ exists. Now

$$\int_0^R \frac{|\sin u|}{u^{3/2}} du \leq \int_0^1 \frac{|\sin u|}{u^{3/2}} du + \int_1^R \frac{1}{u^{3/2}} du \leq 2 + 2$$

and so $\int_0^{\infty} \frac{|\sin u|}{u^{3/2}} du$ exists which shows by Corollary 9.12.12 that $\int_0^{\infty} \frac{\sin u}{u^{3/2}} du$ exists also. I have been a little sketchy on the details. You finish them. This is an especially exciting example because $\lim_{x \rightarrow \infty} \cos(x^2)$ does not exist. This is a contrast to the situation for infinite sums.

9.13 Exercises

- Verify all the details in Example 9.12.8.
- Verify all the details of Example 9.12.9.
- Verify all the details of Example 9.12.13.
- Find the values of p for which $\int_1^{\infty} \frac{1}{t^p} dt$ exists and compute the integral when it does exist.
- Find the values of p for which $\int_2^{\infty} \frac{1}{t(\ln t)^p} dt$ exists and compute the integral when it does exist.
- Determine whether $\int_1^{\infty} \frac{\sin t}{\sqrt{t}} dt$ exists.
- Determine whether $\int_3^{\infty} \frac{\sin t}{\ln t} dt$ exists.
Hint: You might try integrating by parts.
- Determine whether $\int_0^1 \frac{1}{\sqrt{x} + \sin x} dx$ exists.
- Determine whether $\int_1^{\infty} \frac{1}{\sqrt{x} + x^6} dx$ exists.
- Determine whether $\int_0^1 \frac{\sin t}{t} dt$ exists.
- Determine whether $\int_0^1 \frac{\cot t}{t} dt$ exists.
- Determine whether $\int_0^1 \frac{1}{1-t^3} dt$ exists.
- Determine whether $\int_0^1 \frac{1}{1-\sqrt{x}} dx$ exists.
- Determine whether $\int_0^1 \frac{1}{\sqrt{1-x}} dx$ exists.
- Find $\int_0^1 \frac{1}{\sqrt[3]{1-x}} dx$ if it exists.
- Find $\int_0^{\pi/2} \frac{\cos x}{\sqrt{1-\sin x}} dx$ if it exists.
- Find $\int_0^{\infty} \frac{1}{4x^2+9} dx$ if it exists.
- Define and find $\int_{-\infty}^0 e^x dx$. Note the lower limit of integration is $-\infty$.
- Find $\int_0^{\infty} e^{-x} dx$ and then define and find the volume obtained by revolving the graph of $y = e^{-x}$ about the x axis. Define the surface area of the shape obtained by revolving about the x axis and determine whether it is finite.
- When $\int_0^{\infty} f(x) dx$ and $\int_{-\infty}^0 f(x) dx$ both exist, it follows $\int_{-\infty}^{\infty} f(x) dx$ also exists and equals the sum of the two first integrals,

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \int_{-\infty}^0 f(x) dx \\ &+ \int_0^{\infty} f(x) dx. \end{aligned}$$

The normal distribution function is $\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. In this formula, μ is the mean and σ is a positive number called the standard deviation. In statistics, there are things called random variables. These are really just a kind of function and one

of these random variables is said to be normally distributed if the probability that it has a value between a and b is given by the integral $\int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$. You may be wondering why $\sqrt{2\pi}$ occurs in this. It is because it is what is needed to have $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$. Later in the book you will learn how to show this. It is in Problem 14 on Page 574 and depends on changing variables in multiple integrals. For now, show $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$ exists. The importance of this distribution function cannot be over stated.

21. Show $\Gamma(1) = 1 = \Gamma(2)$. Next show $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ and prove that for n a nonnegative integer, $\Gamma(n + 1) = n!$.
22. It can be shown that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Using this and Problem 21, find $\Gamma(\frac{5}{2})$. What is the advantage of the gamma function over the notion of factorials? **Hint:** For what values of x is $\Gamma(x)$ defined?
23. Prove $\int_0^{\infty} \sin(x^2) dx$ exists. This is also called a Fresnel integral.
24. For $\alpha > 0$ find $\int_0^1 t^{\alpha} \ln t dt$ if it exists and if it does not exist, explain why.
25. Prove that for every $\alpha > 0$, $\int_0^1 t^{\alpha-1} dt$ exists and find the answer.
26. Prove that for every $\alpha > 0$,

$$\int_0^1 (\tan t)^{\alpha-1} dt$$

exists.

27. Prove that for every $\alpha > 0$,

$$\int_0^1 (\sin t)^{\alpha-1} dt$$

exists.

28. Recall the area of the surface obtained by revolving the graph of $y = f(x)$ about the x axis for $x \in [a, b]$ for f a positive continuous function

having continuous derivative is given by

$$2\pi \int_a^b f(x) \sqrt{1 + (f'(x))^2} dx.$$

Also the volume of the solid obtained in the same way is given by the integral,

$$\pi \int_a^b (f(x))^2 dx.$$

It seems reasonable to define the surface area and volume for $x \in [a, \infty)$ in terms of an improper integral. Try it on the function, $f(x) = 1/x$ for $x \geq 1$. Show the resulting solid has finite volume but infinite surface area. Thus you could fill it but you couldn't paint it. Sometimes people call this Gabriel's horn.

29. Show $\int_0^{\infty} \frac{2x}{1+x^2} dx$ does not exist but that $\lim_{R \rightarrow \infty} \int_{-R}^R \frac{2x}{1+x^2} dx = 0$. This last limit is called the Cauchy principle value integral. It is not a very respectable thing. Later you might study a subject called complex analysis in which techniques for finding hard integrals are developed. These methods often give the Cauchy principle value. Try to show the following: For every number, A , there exist sequences $a_n, b_n \rightarrow \infty$ such that

$$\lim_{n \rightarrow \infty} \int_{-a_n}^{b_n} \frac{2x}{1+x^2} dx = A.$$

This is true and shows why such principle value integrals are somewhat disreputable.

30. Suppose f is a continuous function which is bounded and defined on $\mathbb{R}$. Show

$$\int_{-\infty}^{\infty} \frac{\varepsilon}{\pi(\varepsilon^2 + (x - x_1)^2)} dx = 1.$$

Next show that

$$\lim_{\varepsilon \rightarrow 0^+} \int_{-\infty}^{\infty} \frac{\varepsilon f(x)}{\pi(\varepsilon^2 + (x - x_1)^2)} dx$$

$$= f(x_1).$$

31. The ideas in these next two problems come from Wooley, who is an engineer I met. He was interested in nonlinear springs and he devised this model for describing collisions. It is a very interesting application of improper integrals. Recall the equation of undamped oscillation,

$$y'' + \beta y = 0$$

where $\beta > 0$ is related to a spring constant. This describes a linear spring. Of course there is no such thing. It is a mathematical idealization which is convenient to use because the solutions involve known functions, sin and cos. Wooley proposed the following equation for a nonlinear spring,

$$y'' + \beta y^\alpha = 0$$

where $\alpha \in (0, 1]$ and $\beta > 0$. As part of the analysis, it was interesting to find β such that the solution to

$$\begin{aligned} y'' + \beta y^\alpha &= 0 \\ y(0) &= 0, y'(0) = 1 \end{aligned} \quad (9.35)$$

also has the property that $y'(1) = 0$ and $y'(t) > 0$ on $(0, 1)$. Multiply the differential equation by y' and then show that

$$\left(\frac{(y')^2}{2} + \frac{\beta y^{\alpha+1}}{\alpha+1} \right)' = 0.$$

Then use the conditions given to conclude

$$(y')^2 + \frac{2\beta y^{\alpha+1}}{\alpha+1} = 1 \quad (9.37)$$

32. $\uparrow$ Now argue that from this it follows

$$\frac{dy}{dt} = \sqrt{1 - \frac{2\beta y^{\alpha+1}}{\alpha+1}}$$

and this is a separable differential equation. Now explain why

$$\int_0^y \frac{dz}{\sqrt{1 - \frac{2\beta z^{\alpha+1}}{\alpha+1}}} = t \quad (9.38)$$

for $t \in (0, 1)$. Now change the variables, letting

$$w = \frac{2\beta z^{\alpha+1}}{\alpha+1}$$

Show 9.38 reduces to

$$\begin{aligned} & \int_0^{\frac{2\beta y^{\alpha+1}}{\alpha+1}} \frac{dw}{\left(\left(\frac{\gamma w}{2\beta} \right)^{1/\gamma} \right)^\alpha \sqrt{1-w}} \\ &= t \end{aligned}$$

where $\gamma = \alpha + 1$, which simplifies to

$$\begin{aligned} & \frac{1}{2^{\frac{1}{\alpha+1}} \beta^{\frac{1}{\alpha+1}} (\alpha+1)^{\frac{\alpha}{\alpha+1}}} \\ & \int_0^{\frac{2\beta y^{\alpha+1}}{\alpha+1}} \frac{dw}{w^{\frac{\alpha}{\alpha+1}} \sqrt{1-w}} = t \end{aligned}$$

Now explain how 9.37 and the condition that $y'(1) = 0$ implies

$$\begin{aligned} & \frac{1}{2^{\frac{1}{\alpha+1}} \beta^{\frac{1}{\alpha+1}} (\alpha+1)^{\frac{\alpha}{\alpha+1}}} \\ & \int_0^1 \frac{dw}{w^{\frac{\alpha}{\alpha+1}} \sqrt{1-w}} = 1 \end{aligned}$$

Explain why this improper integral converges and solve for β to find

$$\begin{aligned} & \left(\frac{1}{(\alpha+1)^{\frac{\alpha}{\alpha+1}} 2^{\frac{1}{\alpha+1}}} \right. \\ & \left. \int_0^1 \frac{dw}{w^{\frac{\alpha}{\alpha+1}} \sqrt{1-w}} \right)^{\alpha+1} = \beta. \end{aligned}$$

What happens in the case that $\alpha = 1$? What should be the correct value of β in this case? If you know β , then there are well known numerical methods which will allow you to find the solution to the initial value problem. Note the correct value of β involves the computation of an improper integral.

Infinite Series

10.0.1 Outcomes

1. Understand the use of Taylor polynomials for approximating a given function and understand a proof of some form for the remainder.
2. Recall and understand the meaning of a convergent series.
3. Recall, use and understand the various tests for determining convergence of series.
4. Recall and explain the difference between conditional and absolute convergence.
5. Be able to justify the interchange in order of summation in a double sum.
6. Understand the use of a Taylor series as the definition of a function.
7. Understand and use correctly various methods for determining the interval of convergence and radius of convergence of a Taylor series.
8. Recall and understand the various operations which can be used on Taylor series and be able to justify their use.

10.1 Approximation By Taylor Polynomials

By now, you have noticed there are two sorts of functions, those which come from a formula like $f(x) = x^2 + 2$ which are easy to evaluate by following a simple procedure, and those which come as short words; things like $\ln(x)$ or $\sin(x)$. This latter type of function is not so easy to evaluate. For example, what is $\sin 2$? Can you get it by doing a simple sequence of operations like you can with $f(x) = x^2 + 2$? How can you find $\sin 2$? It turns out there are many ways to do so. In this section, the method of Taylor polynomials is discussed. The following theorem is called Taylor's theorem. Before presenting it, recall the meaning of $n!$ for n a positive integer. Define $0! \equiv 1 = 1!$ and $(n+1)! \equiv (n+1)n!$ so that $n! = n(n-1) \cdots 1$. In particular, $2! = 2$, $3! = 3 \times 2! = 6$, $4! = 4 \times 3! = 24$, etc. The following is due to Lagrange, about 1790.

Theorem 10.1.1 *Suppose f has $n+1$ derivatives on an interval, (a, b) and let $c \in (a, b)$. Then if $x \in (a, b)$, there exists ξ between c and x such that*

$$f(x) = f(c) + \sum_{k=1}^n \frac{f^{(k)}(c)}{k!} (x-c)^k + \frac{f^{(n+1)}(\xi)}{(n+1)!} (x-c)^{n+1}.$$

(In this formula, the symbol $\sum_{k=1}^0 a_k$ will denote the number 0.)

Proof: If $n = 0$ then the theorem is true because it is just the mean value theorem. Suppose the theorem is true for $n - 1, n \geq 1$. It can be assumed $x \neq c$ because if $x = c$ there is nothing to show. Then there exists K such that

$$f(x) - \left(f(c) + \sum_{k=1}^n \frac{f^{(k)}(c)}{k!} (x-c)^k + K(x-c)^{n+1} \right) = 0 \quad (10.1)$$

In fact,

$$K = \frac{-f(x) + \left(f(c) + \sum_{k=1}^n \frac{f^{(k)}(c)}{k!} (x-c)^k \right)}{(x-c)^{n+1}}.$$

Now define $F(t)$ for t in the closed interval determined by x and c by

$$F(t) \equiv f(x) - \left(f(t) + \sum_{k=1}^n \frac{f^{(k)}(c)}{k!} (x-t)^k + K(x-t)^{n+1} \right).$$

The c in 10.1 got replaced by t .

Therefore, $F(c) = 0$ by the way K was chosen and also $F(x) = 0$. By the mean value theorem or Rolle's theorem, there exists t_1 between x and c such that $F'(t_1) = 0$. Therefore,

$$\begin{aligned} 0 &= f'(t_1) - \sum_{k=1}^n \frac{f^{(k)}(c)}{k!} k(x-t_1)^{k-1} - K(n+1)(x-t_1)^n \\ &= f'(t_1) - \left(f'(c) + \sum_{k=1}^{n-1} \frac{f^{(k+1)}(c)}{k!} (x-t_1)^k \right) - K(n+1)(x-t_1)^n \\ &= f'(t_1) - \left(f'(c) + \sum_{k=1}^{n-1} \frac{f^{(k+1)}(c)}{k!} (x-t_1)^k \right) - K(n+1)(x-t_1)^n \end{aligned}$$

By induction applied to f' , there exists ξ between x and t_1 such that the above simplifies to

$$\begin{aligned} 0 &= \frac{f^{(n)}(\xi)(x-t_1)^n}{n!} - K(n+1)(x-t_1)^n \\ &= \frac{f^{(n+1)}(\xi)(x-t_1)^n}{n!} - K(n+1)(x-t_1)^n \end{aligned}$$

therefore,

$$K = \frac{f^{(n+1)}(\xi)}{(n+1)n!} = \frac{f^{(n+1)}(\xi)}{(n+1)!}$$

and the formula is true for n . This proves the theorem.

The term $\frac{f^{(n+1)}(\xi)}{(n+1)!} (x-c)^{n+1}$, is called the remainder and this particular form of the remainder is called the Lagrange form of the remainder.

Example 10.1.2 Approximate $\sin x$ for x in some open interval containing 0.

Use Taylor's formula just presented and let $c = 0$. Then for $f(x) = \sin x$,

$$f'(x) = \cos x, \quad f''(x) = -\sin x, \quad f'''(x) = -\cos x,$$

etc. Therefore, $f(0) = 0, f'(0) = 1, f''(0) = 0, f'''(0) = -1$, etc. Thus the Taylor polynomial for $\sin x$ is of the form

$$x - \frac{x^3}{3!} + \cdots \pm \frac{x^{2n+1}}{(2n+1)!} = \sum_{k=1}^{n+1} (-1)^{k-1} \frac{x^{2k-1}}{(2k-1)!}$$

while the remainder is of the form

$$\frac{f^{(2n+2)}(\xi) x^{2n+2}}{(2n+2)!}$$

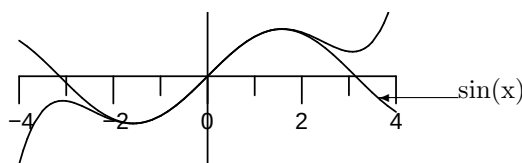
for some ξ between 0 and x . For $n = 2$ in the above, the resulting polynomial is

$$x - \frac{x^3}{3!} + \frac{x^5}{5!}$$

and the error between this polynomial and $\sin x$ must be measured by the remainder term. Therefore,

$$\left| \sin x - \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} \right) \right| \leq \left| \frac{f^{(6)}(\xi) x^6}{6!} \right| \leq \frac{x^6}{6!}.$$

For small x , this error is very small but if x is large, no such conclusion can be drawn. This is illustrated in the following picture drawn by a computer algebra system.



You see from the picture that the polynomial is a very good approximation for the function, $\sin x$ as long as $|x|$ is small but that if $|x|$ gets very large, the approximation is lousy. The above estimate indicates the good approximation holds as long as $|x|$ is small and it quantifies how good the approximation is. Suppose for example, you wanted to find $\sin(.5)$. Then from the above error estimate,

$$\left| \sin(.5) - \left((.5) - \frac{(.5)^3}{3!} + \frac{(.5)^5}{5!} \right) \right| \leq \frac{(.5)^6}{6!} = \frac{1}{46080}$$

so difference between the approximation and $\sin(.5)$ is less than 10^{-4} . If this is used to find $\sin(.1)$ the polynomial approximation would be even closer.

10.2 Exercises

1. Let $p_n(x) = a_0 + \sum_{k=1}^n a_k (x-c)^k$. Show that if you require that $p_n(c) = f(c)$, $p'_n(c) = f'(c)$, $\dots$, $p_n^{(n)}(c) = f^{(n)}(c)$, then this requirement is achieved if and only if $a_0 = f(c)$, $a_1 = f'(c)$, $\dots$, $a_n = \frac{f^{(n)}(c)}{n!}$. Thus the Taylor polynomial of degree n and its first n derivatives agree with the function and its first n derivatives when $x = c$.
2. Find the Taylor polynomials for $\cos x$ for x near 0 along with a formula for the remainder. Use your approximate polynomial to compute $\cos(.5)$ to 3 decimal places and prove your approximation is this good.
3. Find the Taylor polynomials for $x^4 + 2x^3 + x - 7$ for x near 1. Prove that the Taylor polynomial of degree 4 equals the function.
4. Find the Taylor polynomials for $3x^4 + 2x^3 + x^2 - 7$ for x near -1 . Prove that the Taylor polynomial of degree 4 equals the function.

5. Find the Taylor polynomials for $\cosh x$ for x near 0.
6. Find the Taylor polynomials for $\sinh x$ for x near 0.
7. Find the Taylor polynomials for $\ln(1+x)$ for x near 0.
8. Find the Taylor polynomials for $\ln(1-x)$ for x near 0.
9. Find a Taylor polynomial for $\ln\left(\frac{1+x}{1-x}\right)$ and use it to compute $\ln 5$ to three decimal places.
10. Verify that $\lim_{n \rightarrow \infty} \frac{M^n}{n!} = 0$ whenever M is a positive real number. **Hint:** Prove by induction that $M^n/n! \leq (2M)^n/(\sqrt{n})^n$. Now consider what happens when $\sqrt{n}$ is much larger than $2M$.
11. Show that for every $x \in \mathbb{R}$, $\sin(x) = \lim_{n \rightarrow \infty} \sum_{k=1}^n (-1)^{k-1} \frac{x^{2k-1}}{(2k-1)!}$. **Hint:** Use the formula for the error to conclude that

$$\left| \sin x - \sum_{k=1}^n (-1)^{k-1} \frac{x^{2k-1}}{(2k-1)!} \right| \leq \frac{|x|^{2n}}{(2n)!}$$

and then use the result of Problem 10.

12. Show $\cos(x) = \lim_{n \rightarrow \infty} \sum_{k=1}^n (-1)^{k-1} \frac{x^{2k-2}}{(2k-2)!}$ by finding a suitable formula for the remainder and then using an argument similar to that done in Problem 11.
13. Using Problem 10, show $e^x = \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{x^k}{k!}$ for all $x \in \mathbb{R}$.
14. Find a_n such that $\arctan(x) = \lim_{n \rightarrow \infty} \sum_{k=0}^n a_k x^k$ for some values of x . Find the values of x for which the limit is true and prove your result. **Hint:** It is a good idea to use

$$\arctan(x) = \int_0^x \frac{1}{1+t^2} dt,$$

show

$$\frac{1}{1+t^2} = \sum_{k=0}^n (-1)^k t^{2k} \pm \frac{t^{2n+2}}{1+t^2},$$

and then integrate this finite sum from 0 to x . Thus the error would be no larger than

$$\left| \int_0^x \frac{t^{2n+2}}{1+t^2} dt \right| \leq \left| \int_0^x t^{2n+2} dt \right|.$$

15. If you did Problem 14 correctly, you found

$$\arctan x = \lim_{n \rightarrow \infty} \sum_{k=1}^n (-1)^{k-1} \frac{x^{2k-1}}{2k-1}$$

and that this limit will hold for $x \in [-1, 1]$. Use this to verify that

$$\frac{\pi}{4} = \lim_{n \rightarrow \infty} \sum_{k=1}^n (-1)^{k-1} \frac{1}{2k-1}.$$

16. Do for $\ln(1+x)$ what was done for $\arctan(x)$ and find a formula of this sort for $\ln 2$. Use

$$\ln(1+x) = \int_0^x \frac{1}{1+t} dt.$$

17. Repeat 14 for the function $\ln(1 - x)$.
18. Suppose a function $y(x)$ satisfies the initial value problem, $y' = y$, $y(0) = 1$. Find Taylor polynomials for this function. Do you know this function which satisfies the given initial value problem?
19. Suppose a function, $y(x)$ satisfies the initial value problem, $y'' + y = 0$, $y(0) = 0$, $y'(0) = 1$. Find Taylor polynomials for this function. Do you know this function which satisfies the given initial value problem?

10.3 Infinite Series Of Numbers

10.3.1 Basic Considerations

Earlier in Definition 2.11.1 on Page 63 the notion of limit of a sequence was discussed. There is a very closely related concept called an infinite series which is dealt with in this section.

Definition 10.3.1 *Define*

$$\sum_{k=m}^{\infty} a_k \equiv \lim_{n \rightarrow \infty} \sum_{k=m}^n a_k$$

whenever the limit exists and is finite. In this case the series is said to converge. If it does not converge, it is said to diverge. The sequence $\{\sum_{k=m}^n a_k\}_{n=m}^{\infty}$ in the above is called the sequence of partial sums.

From this definition, it should be clear that infinite sums do not always make sense. Sometimes they do and sometimes they don't, depending on the behavior of the partial sums. As an example, consider $\sum_{k=1}^{\infty} (-1)^k$. The partial sums corresponding to this symbol alternate between -1 and 0 . Therefore, there is no limit for the sequence of partial sums. It follows the symbol just written is meaningless and the infinite sum diverges.

Example 10.3.2 *Find the infinite sum, $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$.*

Note $\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1}$ and so $\sum_{n=1}^N \frac{1}{n(n+1)} = \sum_{n=1}^N \left(\frac{1}{n} - \frac{1}{n+1} \right) = -\frac{1}{N+1} + 1$. Therefore,

$$\lim_{N \rightarrow \infty} \sum_{n=1}^N \frac{1}{n(n+1)} = \lim_{N \rightarrow \infty} \left(-\frac{1}{N+1} + 1 \right) = 1.$$

Proposition 10.3.3 *Let $a_k \geq 0$. Then $\{\sum_{k=m}^n a_k\}_{n=m}^{\infty}$ is an increasing sequence. If this sequence is bounded above, then $\sum_{k=m}^{\infty} a_k$ converges and its value equals*

$$\sup \left\{ \sum_{k=m}^n a_k : n = m, m+1, \dots \right\}.$$

When the sequence is not bounded above, $\sum_{k=m}^{\infty} a_k$ diverges.

Proof: It follows $\{\sum_{k=m}^n a_k\}_{n=m}^{\infty}$ is an increasing sequence because

$$\sum_{k=m}^{n+1} a_k - \sum_{k=m}^n a_k = a_{n+1} \geq 0.$$

If it is bounded above, let

$$l \equiv \sup \left\{ \sum_{k=m}^n a_k : n = m, m+1, \dots \right\}$$

be the least upper bound. Then for every $\varepsilon > 0$ there exists p such that

$$l - \varepsilon < \sum_{k=m}^p a_k \leq l$$

since otherwise l would not be the least upper bound of the partial sums. Then since each $a_k \geq 0$, it follows for all $n \geq p$,

$$l - \varepsilon < \sum_{k=m}^p a_k \leq \sum_{k=m}^n a_k \leq l$$

so by definition, the partial sums converge to l . It follows the sequence of partial sums converges to

$$\sup \left\{ \sum_{k=m}^n a_k : n = m, m+1, \dots \right\}.$$

If the sequence of partial sums is not bounded, then it is not a Cauchy sequence and so it does not converge. See Theorem 2.11.13 on Page 66. This proves the proposition.

In the case where $a_k \geq 0$, the above proposition shows there are only two alternatives available. Either the sequence of partial sums is bounded above or it is not bounded above. In the first case convergence occurs and in the second case, the infinite series diverges. For this reason, people will sometimes write $\sum_{k=m}^{\infty} a_k < \infty$ to denote the case where convergence occurs and $\sum_{k=m}^{\infty} a_k = \infty$ for the case where divergence occurs. Be very careful you never think this way in the case where it is not true that all $a_k \geq 0$. For example, the partial sums of $\sum_{k=1}^{\infty} (-1)^k$ are bounded because they are all either -1 or 0 but the series does not converge.

One of the most important examples of a convergent series is the geometric series. This series is $\sum_{n=0}^{\infty} r^n$. The study of this series depends on simple High school algebra and Theorem 2.11.9 on Page 65. Let $S_n \equiv \sum_{k=0}^n r^k$. Then

$$S_n = \sum_{k=0}^n r^k, \quad rS_n = \sum_{k=0}^n r^{k+1} = \sum_{k=1}^{n+1} r^k.$$

Therefore, subtracting the second equation from the first yields

$$(1 - r) S_n = 1 - r^{n+1}$$

and so a formula for S_n is available. In fact, if $r \neq 1$,

$$S_n = \frac{1 - r^{n+1}}{1 - r}.$$

By Theorem 2.11.9, $\lim_{n \rightarrow \infty} S_n = \frac{1}{1-r}$ in the case when $|r| < 1$. Now if $|r| \geq 1$, the limit clearly does not exist because S_n fails to be a Cauchy sequence (Why?). This shows the following.

Theorem 10.3.4 *The geometric series, $\sum_{n=0}^{\infty} r^n$ converges and equals $\frac{1}{1-r}$ if $|r| < 1$ and diverges if $|r| \geq 1$.*

If the series do converge, the following holds.

Theorem 10.3.5 *If $\sum_{k=m}^{\infty} a_k$ and $\sum_{k=m}^{\infty} b_k$ both converge and x, y are numbers, then*

$$\sum_{k=m}^{\infty} a_k = \sum_{k=m+j}^{\infty} a_{k-j} \quad (10.2)$$

$$\sum_{k=m}^{\infty} xa_k + yb_k = x \sum_{k=m}^{\infty} a_k + y \sum_{k=m}^{\infty} b_k \quad (10.3)$$

$$\left| \sum_{k=m}^{\infty} a_k \right| \leq \sum_{k=m}^{\infty} |a_k| \quad (10.4)$$

where in the last inequality, the last sum equals $+\infty$ if the partial sums are not bounded above.

Proof: The above theorem is really only a restatement of Theorem 2.11.6 on Page 64 and the above definitions of infinite series. Thus

$$\sum_{k=m}^{\infty} a_k = \lim_{n \rightarrow \infty} \sum_{k=m}^n a_k = \lim_{n \rightarrow \infty} \sum_{k=m+j}^{n+j} a_{k-j} = \sum_{k=m+j}^{\infty} a_{k-j}.$$

To establish 10.3, use Theorem 2.11.6 on Page 64 to write

$$\begin{aligned} \sum_{k=m}^{\infty} xa_k + yb_k &= \lim_{n \rightarrow \infty} \sum_{k=m}^n xa_k + yb_k \\ &= \lim_{n \rightarrow \infty} \left(x \sum_{k=m}^n a_k + y \sum_{k=m}^n b_k \right) \\ &= x \sum_{k=m}^{\infty} a_k + y \sum_{k=m}^{\infty} b_k. \end{aligned}$$

Formula 10.4 follows from the observation that, from the triangle inequality,

$$\left| \sum_{k=m}^n a_k \right| \leq \sum_{k=m}^n |a_k|$$

and so

$$\left| \sum_{k=m}^{\infty} a_k \right| = \lim_{n \rightarrow \infty} \left| \sum_{k=m}^n a_k \right| \leq \sum_{k=m}^{\infty} |a_k|.$$

Example 10.3.6 Find $\sum_{n=0}^{\infty} \left(\frac{5}{2^n} + \frac{6}{3^n} \right)$.

From the above theorem and Theorem 10.3.4,

$$\begin{aligned} \sum_{n=0}^{\infty} \left(\frac{5}{2^n} + \frac{6}{3^n} \right) &= 5 \sum_{n=0}^{\infty} \frac{1}{2^n} + 6 \sum_{n=0}^{\infty} \frac{1}{3^n} \\ &= 5 \frac{1}{1 - (1/2)} + 6 \frac{1}{1 - (1/3)} = 19. \end{aligned}$$

The following criterion is useful in checking convergence.

Theorem 10.3.7 *The sum $\sum_{k=m}^{\infty} a_k$ converges if and only if for all $\varepsilon > 0$, there exists n_ε such that if $q \geq p \geq n_\varepsilon$, then*

$$\left| \sum_{k=p}^q a_k \right| < \varepsilon. \quad (10.5)$$

Proof: Suppose first that the series converges. Then $\{\sum_{k=m}^n a_k\}_{n=m}^{\infty}$ is a Cauchy sequence by Theorem 2.11.13 on Page 66. Therefore, there exists $n_\varepsilon > m$ such that if $q \geq p - 1 \geq n_\varepsilon > m$,

$$\left| \sum_{k=m}^q a_k - \sum_{k=m}^{p-1} a_k \right| = \left| \sum_{k=p}^q a_k \right| < \varepsilon. \quad (10.6)$$

Next suppose 10.5 holds. Then from 10.6 it follows upon letting p be replaced with $p + 1$ that $\{\sum_{k=m}^n a_k\}_{n=m}^{\infty}$ is a Cauchy sequence and so, by the completeness axiom, it converges. By the definition of infinite series, this shows the infinite sum converges as claimed.

Definition 10.3.8 *A series*

$$\sum_{k=m}^{\infty} a_k$$

is said to converge absolutely if

$$\sum_{k=m}^{\infty} |a_k|$$

converges. If the series does converge but does not converge absolutely, then it is said to converge conditionally.

Theorem 10.3.9 *If $\sum_{k=m}^{\infty} a_k$ converges absolutely, then it converges. Also if each $a_k \geq 0$ and the partial sums are bounded above, then $\sum_{k=m}^{\infty} a_k$ converges.*

Proof: Let $\varepsilon > 0$ be given. Then by assumption and Theorem 10.3.7, there exists n_ε such that whenever $q \geq p \geq n_\varepsilon$,

$$\sum_{k=p}^q |a_k| < \varepsilon.$$

Therefore, from the triangle inequality,

$$\varepsilon > \sum_{k=p}^q |a_k| \geq \left| \sum_{k=p}^q a_k \right|.$$

By Theorem 10.3.7, $\sum_{k=m}^{\infty} a_k$ converges and this proves the first part. The second claim follows from Proposition 10.3.3. This proves the theorem.

In fact, the above theorem is really another version of the completeness axiom. Thus its validity implies completeness. You might try to show this.

Theorem 10.3.10 *(comparison test) Suppose $\{a_n\}$ and $\{b_n\}$ are sequences of non negative real numbers and suppose for all n sufficiently large, $a_n \leq b_n$. Then*

1. *If $\sum_{n=k}^{\infty} b_n$ converges, then $\sum_{n=m}^{\infty} a_n$ converges.*

2. If $\sum_{n=k}^{\infty} a_n$ diverges, then $\sum_{n=m}^{\infty} b_n$ diverges.

Proof: Consider the first claim. From the assumption there exists n^* such that $n^* > \max(k, m)$ and for all $n \geq n^*$ $b_n \geq a_n$. Then if $p \geq n^*$,

$$\begin{aligned} \sum_{n=m}^p a_n &\leq \sum_{n=m}^{n^*} a_n + \sum_{n=n^*+1}^k b_n \\ &\leq \sum_{n=m}^{n^*} a_n + \sum_{n=k}^{\infty} b_n. \end{aligned}$$

Thus the sequence, $\{\sum_{n=m}^p a_n\}_{p=m}^{\infty}$ is bounded above and increasing. Therefore, it converges by Theorem 10.3.9. The second claim is left as an exercise.

Example 10.3.11 Determine the convergence of $\sum_{n=1}^{\infty} \frac{1}{n^2}$.

For $n > 1$,

$$\frac{1}{n^2} \leq \frac{1}{n(n-1)}.$$

Now

$$\begin{aligned} \sum_{n=2}^p \frac{1}{n(n-1)} &= \sum_{n=2}^p \left[\frac{1}{n-1} - \frac{1}{n} \right] \\ &= 1 - \frac{1}{p} \rightarrow 1 \end{aligned}$$

Therefore, letting $a_n = \frac{1}{n^2}$ and $b_n = \frac{1}{n(n-1)}$

A convenient way to implement the comparison test is to use the limit comparison test. This is considered next.

Theorem 10.3.12 Let $a_n, b_n > 0$ and suppose for all n large enough,

$$0 < a < \frac{a_n}{b_n} < b < \infty.$$

Then $\sum a_n$ and $\sum b_n$ converge or diverge together.

Proof: Let n^* be such that $n \geq n^*$, then

$$\frac{a_n}{b_n} > a \text{ and } \frac{a_n}{b_n} < b$$

and so for all such n ,

$$ab_n < a_n < bb_n$$

and so the conclusion follows from the comparison test.

Corollary 10.3.13 If $a_n, b_n > 0$ and $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = r \in (0, \infty)$, then $\sum a_n, \sum b_n$ converge or diverge together.

Proof: The assumption implies that for all n large enough,

$$\frac{r}{2} < \frac{a_n}{b_n} < 2r$$

The conclusion now follows from Theorem 10.3.12.

Example 10.3.14 Determine the convergence of $\sum_{k=1}^{\infty} \frac{1}{\sqrt{n^4+2n+7}}$.

This series converges by the limit comparison test. Compare with the series of Example 10.3.11.

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\left(\frac{1}{n^2}\right)}{\left(\frac{1}{\sqrt{n^4+2n+7}}\right)} &= \lim_{n \rightarrow \infty} \frac{\sqrt{n^4+2n+7}}{n^2} \\ &= \lim_{n \rightarrow \infty} \sqrt{1 + \frac{2}{n^3} + \frac{7}{n^4}} = 1. \end{aligned}$$

Therefore, the series converges with the series of Example 10.3.11. How did I know what to compare with? I noticed that $\sqrt{n^4+2n+7}$ is essentially like $\sqrt{n^4} = n^2$ for large enough n . You see, the higher order term, n^4 dominates the other terms in n^4+2n+7 . Therefore, reasoning that $1/\sqrt{n^4+2n+7}$ is a lot like $1/n^2$ for large n , it was easy to see what to compare with. Of course this is not always easy and there is room for acquiring skill through practice.

To really exploit this limit comparison test, it is desirable to get lots of examples of series, some which converge and some which do not. The tool for obtaining these examples here will be the following wonderful theorem known as the Cauchy condensation test.

Theorem 10.3.15 Let $a_n \geq 0$ and suppose the terms of the sequence $\{a_n\}$ are decreasing. Thus $a_n \geq a_{n+1}$ for all n . Then

$$\sum_{n=1}^{\infty} a_n \text{ and } \sum_{n=0}^{\infty} 2^n a_{2^n}$$

converge or diverge together.

Proof: This follows from the inequality of the following claim.

Claim:

$$\sum_{k=1}^n 2^k a_{2^{k-1}} \geq \sum_{k=1}^{2^n} a_k \geq \sum_{k=0}^n 2^{k-1} a_{2^k}.$$

Proof of the Claim: Note the claim is true for $n = 1$. Suppose the claim is true for n . Then, since $2^{n+1} - 2^n = 2^n$, and the terms, a_n , are decreasing,

$$\begin{aligned} \sum_{k=1}^{n+1} 2^k a_{2^{k-1}} &= 2^{n+1} a_{2^n} + \sum_{k=1}^n 2^k a_{2^{k-1}} \geq 2^{n+1} a_{2^n} + \sum_{k=1}^{2^n} a_k \\ &\geq \sum_{k=1}^{2^{n+1}} a_k \geq 2^n a_{2^{n+1}} + \sum_{k=1}^{2^n} a_k \geq 2^n a_{2^{n+1}} + \sum_{k=0}^n 2^{k-1} a_{2^k} = \sum_{k=0}^{n+1} 2^{k-1} a_{2^k}. \end{aligned}$$

Example 10.3.16 Determine the convergence of $\sum_{k=1}^{\infty} \frac{1}{k^p}$ where p is a positive number. These are called the p series.

Let $a_n = \frac{1}{n^p}$. Then $a_{2^n} = \left(\frac{1}{2^p}\right)^n$. From the Cauchy condensation test the two series

$$\sum_{n=1}^{\infty} \frac{1}{n^p} \text{ and } \sum_{n=0}^{\infty} 2^n \left(\frac{1}{2^p}\right)^n = \sum_{n=0}^{\infty} \left(2^{(1-p)}\right)^n$$

converge or diverge together. If $p > 1$, the last series above is a geometric series having common ratio less than 1 and so it converges. If $p \leq 1$, it is still a geometric series but in this case the common ratio is either 1 or greater than 1 so the series diverges. It follows that the p series converges if $p > 1$ and diverges if $p \leq 1$. In particular, $\sum_{n=1}^{\infty} n^{-1}$ diverges while $\sum_{n=1}^{\infty} n^{-2}$ converges. The following table summarizes the above results.

The p series table

	$p \leq 1$	$p > 1$
$\sum \frac{1}{n^p}$	diverges	converges

Example 10.3.17 Determine the convergence of $\sum_{k=1}^{\infty} \frac{1}{\sqrt{n^2+100n}}$.

Use the limit comparison test.

$$\lim_{n \rightarrow \infty} \frac{\left(\frac{1}{n}\right)}{\left(\frac{1}{\sqrt{n^2+100n}}\right)} = 1$$

and so this series diverges with $\sum_{k=1}^{\infty} \frac{1}{k}$.

Example 10.3.18 Determine the convergence of $\sum_{k=2}^{\infty} \frac{1}{k \ln k}$.

Use the Cauchy condensation test. The above series does the same thing in terms of convergence as the series

$$\sum_{n=1}^{\infty} 2^n \frac{1}{2^n \ln(2^n)} = \sum_{n=1}^{\infty} \frac{1}{n \ln 2}$$

and this series diverges by limit comparison with the series $\sum \frac{1}{n}$.

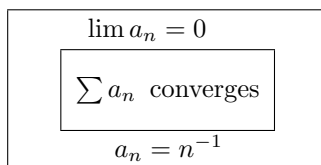
Sometimes it is good to be able to say a series does not converge. The n^{th} term test gives such a condition which is sufficient for this. It is really a corollary of Theorem 10.3.7.

Theorem 10.3.19 If $\sum_{n=m}^{\infty} a_n$ converges, then $\lim_{n \rightarrow \infty} a_n = 0$.

Proof: Apply Theorem 10.3.7 to conclude that

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \sum_{k=n}^n a_k = 0.$$

It is very important to observe that this theorem goes only in one direction. That is, you cannot conclude the series converges if $\lim_{n \rightarrow \infty} a_n = 0$. If this happens, you don't know anything from this information. Recall $\lim_{n \rightarrow \infty} n^{-1} = 0$ but $\sum_{n=1}^{\infty} n^{-1}$ diverges. The following picture is descriptive of the situation.



10.4 Exercises

1. Determine whether the following series converge and give reasons for your answers.
 - (a) $\sum_{n=1}^{\infty} \frac{1}{\sqrt{n^2+n+1}}$
 - (b) $\sum_{n=1}^{\infty} (\sqrt{n+1} - \sqrt{n})$
 - (c) $\sum_{n=1}^{\infty} \frac{(2n)!}{(n!)^2}$
 - (d) $\sum_{n=1}^{\infty} \frac{1}{2n+2}$

(e) $\sum_{n=1}^{\infty} \left(\frac{n}{n+1}\right)^n$

2. Determine whether the following series converge and give reasons for your answers.

(a) $\sum_{n=1}^{\infty} \frac{\ln(k^5)}{k}$

(b) $\sum_{n=1}^{\infty} \frac{\ln(k^5)}{k^{1.01}}$

(c) $\sum_{n=1}^{\infty} \sin\left(\frac{1}{n}\right)$

(d) $\sum_{n=1}^{\infty} \tan\left(\frac{1}{n^2}\right)$

(e) $\sum_{n=1}^{\infty} \cos\left(\frac{1}{n^2}\right)$

(f) $\sum_{n=1}^{\infty} \sin\left(\frac{\sqrt{n}}{n^2+1}\right)$

3. Determine whether the following series converge and give reasons for your answers.

(a) $\sum_{n=1}^{\infty} \frac{2^n+n}{n^{2^n}}$

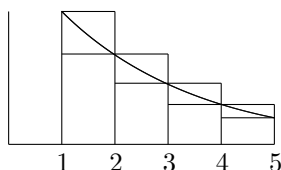
(b) $\sum_{n=1}^{\infty} \frac{2^n+n}{n^2 2^n}$

(c) $\sum_{n=1}^{\infty} \frac{n}{2n+1}$

(d) $\sum_{n=1}^{\infty} \frac{\ln n}{n^2}$

4. Suppose a_n and b_n are just numbers, either positive or negative and that for all n large enough, $0 < a < a_n/b_n < b < \infty$. Show that either one of $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ converges absolutely if and only if the other converges absolutely. This little fact, generalizing Theorem 10.3.12 was pointed out to me by Rodney Forcade.

5. Suppose f is a nonnegative continuous decreasing function defined on $[1, \infty)$. Show the improper integral, $\int_1^{\infty} f(t) dt$ and the sum $\sum_{k=1}^{\infty} f(k)$ converge or diverge together. This is called the integral test. Use this test to verify convergence of $\sum_{k=1}^{\infty} \frac{1}{k^\alpha}$ whenever $\alpha > 1$ and divergence whenever $\alpha \leq 1$. In showing this integral test, it might be helpful to consider the following picture.



In this picture, the graph of the continuous decreasing function is represented. There are rectangles below this curve and rectangles above it. $f(1)$ is the area of the first rectangle on the left which is above the curve while $f(2)$ is the area of the first rectangle on the left which lies below the curve. From the picture, $\int_1^5 f(t) dt$ lies between $f(2) + f(3) + f(4) + f(5)$ and $f(1) + f(2) + f(3) + f(4)$. Generalize to conclude that for any $n \in \mathbb{N}$, it follows that $\sum_{k=2}^n f(k) \leq \int_1^n f(t) dt \leq \sum_{k=1}^{n-1} f(k)$. Now explain why this inequality implies the integral and sum have the same convergence properties.

6. Show using either the Cauchy condensation test or the integral test that

$$\sum_{k=4}^{\infty} \frac{1}{\ln(\ln(k)) \ln(k) k}$$

diverges. Estimate how big N must be in order that $\sum_{k=4}^N \frac{1}{\ln(\ln(k)) \ln(k) k} > 10$. What does this tell you about the wisdom of attempting to determine questions of convergence through experimentation?

7. For p a positive number, determine the convergence of $\sum_{n=2}^{\infty} \frac{1}{n(\ln(n))^p}$ for various values of p .
8. For p a positive number, determine the convergence of

$$\sum_{n=2}^{\infty} \frac{\ln n}{n^p}$$

for various values of p .

9. Determine the convergence of the series $\sum_{n=1}^{\infty} \left(\sum_{k=1}^n \frac{1}{k}\right)^{-n/2}$.
10. Is it possible there could exist a decreasing sequence of positive numbers, $\{a_n\}$ such that $\lim_{n \rightarrow \infty} a_n = 0$ but $\sum_{n=1}^{\infty} \left(1 - \frac{a_{n+1}}{a_n}\right)$ converges?

Hint: You might do something like this. Show

$$\lim_{x \rightarrow 1} \frac{1-x}{-\ln(x)} = \frac{1-x}{\ln(1/x)} = 1$$

Next use a limit comparison test with

$$\sum_{n=1}^{\infty} \ln \left(\frac{a_n}{a_{n+1}} \right)$$

10.5 More Tests For Convergence

10.5.1 Convergence Because Of Cancellation

So far, the tests for convergence have been applied to non negative terms only. Sometimes, a series converges, not because the terms of the series get small fast enough, but because of cancellation taking place between positive and negative terms. A discussion of this involves some simple algebra.

Let $\{a_n\}$ and $\{b_n\}$ be sequences and let

$$A_n \equiv \sum_{k=1}^n a_k, \quad A_{-1} \equiv A_0 \equiv 0.$$

Then if $p < q$

$$\begin{aligned} \sum_{n=p}^q a_n b_n &= \sum_{n=p}^q b_n (A_n - A_{n-1}) = \sum_{n=p}^q b_n A_n - \sum_{n=p}^q b_n A_{n-1} \\ &= \sum_{n=p}^q b_n A_n - \sum_{n=p-1}^{q-1} b_{n+1} A_n = b_q A_q - b_p A_{p-1} + \sum_{n=p}^{q-1} A_n (b_n - b_{n+1}) \end{aligned}$$

This formula is called the partial summation formula. It is just like integration by parts.

Theorem 10.5.1 (*Dirichlet's test*) Suppose A_n is bounded and $\lim_{n \rightarrow \infty} b_n = 0$, with $b_n \geq b_{n+1}$. Then

$$\sum_{n=1}^{\infty} a_n b_n$$

converges.

Proof: This follows quickly from Theorem 10.3.7. Indeed, letting $|A_n| \leq C$, and using the partial summation formula above along with the assumption that the b_n are decreasing,

$$\begin{aligned} \left| \sum_{n=p}^q a_n b_n \right| &= \left| b_q A_q - b_p A_{p-1} + \sum_{n=p}^{q-1} A_n (b_n - b_{n+1}) \right| \\ &\leq C(|b_q| + |b_p|) + C \sum_{n=p}^{q-1} (b_n - b_{n+1}) \\ &= C(|b_q| + |b_p|) + C(b_p - b_q) \end{aligned}$$

and by assumption, this last expression is small whenever p and q are sufficiently large. This proves the theorem.

Definition 10.5.2 If $b_n > 0$ for all n , a series of the form $\sum_k (-1)^k b_k$ or $\sum_k (-1)^{k-1} b_k$ is known as an **alternating series**.

The following corollary is known as the alternating series test.

Corollary 10.5.3 (*alternating series test*) If $\lim_{n \rightarrow \infty} b_n = 0$, with $b_n \geq b_{n+1}$, then $\sum_{n=1}^{\infty} (-1)^n b_n$ converges.

Proof: Let $a_n = (-1)^n$. Then the partial sums of $\sum_n a_n$ are bounded and so Theorem 10.5.1 applies.

In the situation of Corollary 10.5.3 there is a convenient error estimate available.

Theorem 10.5.4 Let $b_n > 0$ for all n such that $b_n \geq b_{n+1}$ for all n and $\lim_{n \rightarrow \infty} b_n = 0$ and consider either $\sum_{n=1}^{\infty} (-1)^n b_n$ or $\sum_{n=1}^{\infty} (-1)^{n-1} b_n$. Then

$$\left| \sum_{n=1}^{\infty} (-1)^n b_n - \sum_{n=1}^N (-1)^n b_n \right| \leq |b_{N+1}|, \quad \left| \sum_{n=1}^{\infty} (-1)^{n-1} b_n - \sum_{n=1}^N (-1)^{n-1} b_n \right| \leq |b_{N+1}|$$

Example 10.5.5 How many terms must I take in the sum, $\sum_{n=1}^{\infty} (-1)^n \frac{1}{n^2+1}$ to be closer than $\frac{1}{10}$ to $\sum_{n=1}^{\infty} (-1)^n \frac{1}{n^2+1}$?

From Theorem 10.5.4, I need to find n such that $\frac{1}{n^2+1} \leq \frac{1}{10}$ and then $n-1$ is the desired value. Thus $n = 3$ and so

$$\left| \sum_{n=1}^{\infty} (-1)^n \frac{1}{n^2+1} - \sum_{n=1}^2 (-1)^n \frac{1}{n^2+1} \right| \leq \frac{1}{10}$$

10.5.2 Ratio And Root Tests

A favorite test for convergence is the ratio test. This is discussed next. It is the at the other extreme. This test is completely oblivious to any sort of cancellation. It only gives absolute convergence or spectacular divergence.

Theorem 10.5.6 Suppose $|a_n| > 0$ for all n and suppose

$$\lim_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|} = r.$$

Then

$$\sum_{n=1}^{\infty} a_n \begin{cases} \text{diverges if } r > 1 \\ \text{converges absolutely if } r < 1 \\ \text{test fails if } r = 1 \end{cases}.$$

Proof: Suppose $r < 1$. Then there exists n_1 such that if $n \geq n_1$, then

$$0 < \left| \frac{a_{n+1}}{a_n} \right| < R$$

where $r < R < 1$. Then

$$|a_{n+1}| < R |a_n|$$

for all such n . Therefore,

$$|a_{n_1+p}| < R |a_{n_1+p-1}| < R^2 |a_{n_1+p-2}| < \cdots < R^p |a_{n_1}| \quad (10.7)$$

and so if $m > n$, then $|a_m| < R^{m-n_1} |a_{n_1}|$. By the comparison test and the theorem on geometric series, $\sum |a_n|$ converges. This proves the convergence part of the theorem.

To verify the divergence part, note that if $r > 1$, then 10.7 can be turned around for some $R > 1$. Showing $\lim_{n \rightarrow \infty} |a_n| = \infty$. Since the n^{th} term fails to converge to 0, it follows the series diverges.

To see the test fails if $r = 1$, consider $\sum n^{-1}$ and $\sum n^{-2}$. The first series diverges while the second one converges but in both cases, $r = 1$. (Be sure to check this last claim.)

The ratio test is very useful for many different examples but it is somewhat unsatisfactory mathematically. One reason for this is the assumption that $a_n > 0$, necessitated by the need to divide by a_n , and the other reason is the possibility that the limit might not exist. The next test, called the root test removes both of these objections.

Theorem 10.5.7 Suppose $|a_n|^{1/n} < R < 1$ for all n sufficiently large. Then

$$\sum_{n=1}^{\infty} a_n \text{ converges absolutely.}$$

If there are infinitely many values of n such that $|a_n|^{1/n} \geq 1$, then

$$\sum_{n=1}^{\infty} a_n \text{ diverges.}$$

Proof: Suppose first that $|a_n|^{1/n} < R < 1$ for all n sufficiently large. Say this holds for all $n \geq n_R$. Then for such n ,

$$\sqrt[n]{|a_n|} < R.$$

Therefore, for such n ,

$$|a_n| \leq R^n$$

and so the comparison test with a geometric series applies and gives absolute convergence as claimed.

Next suppose $|a_n|^{1/n} \geq 1$ for infinitely many values of n . Then for those values of n , $|a_n| \geq 1$ and so the series fails to converge by the n^{th} term test.

Corollary 10.5.8 Suppose $\lim_{n \rightarrow \infty} |a_n|^{1/n}$ exists and equals r . Then

$$\sum_{k=m}^{\infty} a_k \begin{cases} \text{converges absolutely if } r < 1 \\ \text{test fails if } r = 1 \\ \text{diverges if } r > 1 \end{cases}$$

Proof: The first and last alternatives follow from Theorem 10.5.7. To see the test fails if $r = 1$, consider the two series $\sum_{n=1}^{\infty} \frac{1}{n}$ and $\sum_{n=1}^{\infty} \frac{1}{n^2}$ both of which have $r = 1$ but having different convergence properties.

Example 10.5.9 Show that for all $x \in \mathbb{R}$,

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}. \quad (10.8)$$

By Taylor's theorem

$$\begin{aligned} e^x &= 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + e^{\xi_n} \frac{x^{n+1}}{(n+1)!} \\ &= \sum_{k=0}^n \frac{x^k}{k!} + e^{\xi_n} \frac{x^{n+1}}{(n+1)!} \end{aligned} \quad (10.9)$$

where $|\xi_n| \leq |x|$. Now for any $x \in \mathbb{R}$

$$\left| e^{\xi_n} \frac{x^{n+1}}{(n+1)!} \right| \leq e^{|x|} \frac{|x|^{n+1}}{(n+1)!}$$

and an application of the ratio test shows

$$\sum_{n=0}^{\infty} e^{|x|} \frac{|x|^{n+1}}{(n+1)!} < \infty.$$

Therefore, the n^{th} term converges to zero by the n^{th} term test and so for each $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} e^{\xi_n} \frac{x^{n+1}}{(n+1)!} = 0.$$

Therefore, taking the limit in 10.9 it follows 10.8 holds.

10.6 Double Series*

Sometimes it is required to consider double series which are of the form

$$\sum_{k=m}^{\infty} \sum_{j=m}^{\infty} a_{jk} \equiv \sum_{k=m}^{\infty} \left(\sum_{j=m}^{\infty} a_{jk} \right).$$

In other words, first sum on j yielding something which depends on k and then sum these. The major consideration for these double series is the question of when

$$\sum_{k=m}^{\infty} \sum_{j=m}^{\infty} a_{jk} = \sum_{j=m}^{\infty} \sum_{k=m}^{\infty} a_{jk}.$$

In other words, when does it make no difference which subscript is summed over first? In the case of finite sums there is no issue here. You can always write

$$\sum_{k=m}^M \sum_{j=m}^N a_{jk} = \sum_{j=m}^N \sum_{k=m}^M a_{jk}$$

because addition is commutative. However, there are limits involved with infinite sums and the interchange in order of summation involves taking limits in a different order. Therefore, it is not always true that it is permissible to interchange the two sums. A general rule of thumb is this: If something involves changing the order in which two limits are taken, you may not do it without agonizing over the question. In general, limits foul up algebra and also introduce things which are counter intuitive. Here is an example. This example is a little technical. It is placed here just to prove conclusively there is a question which needs to be considered.

Example 10.6.1 Consider the following picture which depicts some of the ordered pairs (m, n) where m, n are positive integers.

$$\begin{array}{cccccccc}
0. & 0. & 0. & 0. & 0. & c. & 0. & -c. \\
0. & 0. & 0. & 0. & c. & 0. & -c. & 0. \\
0. & 0. & 0. & c. & 0. & -c. & 0. & 0. \\
0. & 0. & c. & 0. & -c. & 0. & 0. & 0. \\
0. & c. & 0. & -c. & 0. & 0. & 0. & 0. \\
b. & 0. & -c. & 0. & 0. & 0. & 0. & 0. \\
0. & a. & 0. & 0. & 0. & 0. & 0. & 0.
\end{array}$$

The numbers next to the point are the values of a_{mn} . You see $a_{nn} = 0$ for all n , $a_{21} = a$, $a_{12} = b$, $a_{mn} = c$ for (m, n) on the line $y = 1 + x$ whenever $m > 1$, and $a_{mn} = -c$ for all (m, n) on the line $y = x - 1$ whenever $m > 2$.

Then $\sum_{m=1}^{\infty} a_{mn} = a$ if $n = 1$, $\sum_{m=1}^{\infty} a_{mn} = b - c$ if $n = 2$ and if $n > 2$, $\sum_{m=1}^{\infty} a_{mn} = 0$. Therefore,

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{mn} = a + b - c.$$

Next observe that $\sum_{n=1}^{\infty} a_{mn} = b$ if $m = 1$, $\sum_{n=1}^{\infty} a_{mn} = a + c$ if $m = 2$, and $\sum_{n=1}^{\infty} a_{mn} = 0$ if $m > 2$. Therefore,

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn} = b + a + c$$

and so the two sums are different. Moreover, you can see that by assigning different values of a, b , and c , you can get an example for any two different numbers desired.

Don't become upset by this. It happens because, as indicated above, limits are taken in two different orders. An infinite sum always involves a limit and this illustrates why you must always remember this. This example in no way violates the commutative law of addition which has nothing to do with limits. However, it turns out that if $a_{ij} \geq 0$ for all i, j , then you can always interchange the order of summation. This is shown next and is based on the following lemma. First, some notation should be discussed.

Definition 10.6.2 Let $f(a, b) \in [-\infty, \infty]$ for $a \in A$ and $b \in B$ where A, B are sets which means that $f(a, b)$ is either a number, ∞ , or $-\infty$. The symbol, $+\infty$ is interpreted as a point out at the end of the number line which is larger than every real number. Of course there is no such number. That is why it is called ∞ . The symbol, $-\infty$ is interpreted similarly. Then $\sup_{a \in A} f(a, b)$ means $\sup(S_b)$ where $S_b \equiv \{f(a, b) : a \in A\}$.

Unlike limits, you can take the sup in different orders.

Lemma 10.6.3 Let $f(a, b) \in [-\infty, \infty]$ for $a \in A$ and $b \in B$ where A, B are sets. Then

$$\sup_{a \in A} \sup_{b \in B} f(a, b) = \sup_{b \in B} \sup_{a \in A} f(a, b).$$

Proof: Note that for all a, b , $f(a, b) \leq \sup_{b \in B} \sup_{a \in A} f(a, b)$ and therefore, for all a , $\sup_{b \in B} f(a, b) \leq \sup_{b \in B} \sup_{a \in A} f(a, b)$. Therefore,

$$\sup_{a \in A} \sup_{b \in B} f(a, b) \leq \sup_{b \in B} \sup_{a \in A} f(a, b).$$

Repeat the same argument interchanging a and b , to get the conclusion of the lemma.

Lemma 10.6.4 *If $\{A_n\}$ is an increasing sequence in $[-\infty, \infty]$, then*

$$\sup \{A_n\} = \lim_{n \rightarrow \infty} A_n.$$

Proof: Let $\sup \{A_n : n \in \mathbb{N}\} = r$. In the first case, suppose $r < \infty$. Then letting $\varepsilon > 0$ be given, there exists n such that $A_n \in (r - \varepsilon, r]$. Since $\{A_n\}$ is increasing, it follows if $m > n$, then $r - \varepsilon < A_n \leq A_m \leq r$ and so $\lim_{n \rightarrow \infty} A_n = r$ as claimed. In the case where $r = \infty$, then if a is a real number, there exists n such that $A_n > a$. Since $\{A_k\}$ is increasing, it follows that if $m > n$, $A_m > a$. But this is what is meant by $\lim_{n \rightarrow \infty} A_n = \infty$. The other case is that $r = -\infty$. But in this case, $A_n = -\infty$ for all n and so $\lim_{n \rightarrow \infty} A_n = -\infty$.

Theorem 10.6.5 *Let $a_{ij} \geq 0$. Then $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}$.*

Proof: First note there is no trouble in defining these sums because the a_{ij} are all nonnegative. If a sum diverges, it only diverges to ∞ and so ∞ is the value of the sum. Next note that

$$\sum_{j=r}^{\infty} \sum_{i=r}^{\infty} a_{ij} \geq \sup_n \sum_{j=r}^{\infty} \sum_{i=r}^n a_{ij}$$

because for all j ,

$$\sum_{i=r}^{\infty} a_{ij} \geq \sum_{i=r}^n a_{ij}.$$

Therefore,

$$\begin{aligned} \sum_{j=r}^{\infty} \sum_{i=r}^{\infty} a_{ij} &\geq \sup_n \sum_{j=r}^{\infty} \sum_{i=r}^n a_{ij} = \sup_n \lim_{m \rightarrow \infty} \sum_{j=r}^m \sum_{i=r}^n a_{ij} \\ &= \sup_n \lim_{m \rightarrow \infty} \sum_{i=r}^n \sum_{j=r}^m a_{ij} = \sup_n \sum_{i=r}^n \lim_{m \rightarrow \infty} \sum_{j=r}^m a_{ij} \\ &= \sup_n \sum_{i=r}^n \sum_{j=r}^{\infty} a_{ij} = \lim_{n \rightarrow \infty} \sum_{i=r}^n \sum_{j=r}^{\infty} a_{ij} = \sum_{i=r}^{\infty} \sum_{j=r}^{\infty} a_{ij} \end{aligned}$$

Interchanging the i and j in the above argument proves the theorem.

The following is the fundamental result on double sums.

Theorem 10.6.6 *Let a_{ij} be a number and suppose*

$$\sum_{i=r}^{\infty} \sum_{j=r}^{\infty} |a_{ij}| < \infty.$$

Then

$$\sum_{i=r}^{\infty} \sum_{j=r}^{\infty} a_{ij} = \sum_{j=r}^{\infty} \sum_{i=r}^{\infty} a_{ij}$$

and every infinite sum encountered in the above equation converges.

Proof: By Theorem 10.6.5

$$\sum_{j=r}^{\infty} \sum_{i=r}^{\infty} |a_{ij}| = \sum_{i=r}^{\infty} \sum_{j=r}^{\infty} |a_{ij}| < \infty$$

Therefore, for each j , $\sum_{i=r}^{\infty} |a_{ij}| < \infty$ and for each i , $\sum_{j=r}^{\infty} |a_{ij}| < \infty$. By Theorem 10.3.9 on Page 260, $\sum_{i=r}^{\infty} a_{ij}$, $\sum_{j=r}^{\infty} a_{ij}$ both converge, the first one for every j and the second for every i . Also,

$$\sum_{j=r}^{\infty} \left| \sum_{i=r}^{\infty} a_{ij} \right| \leq \sum_{j=r}^{\infty} \sum_{i=r}^{\infty} |a_{ij}| < \infty$$

and

$$\sum_{i=r}^{\infty} \left| \sum_{j=r}^{\infty} a_{ij} \right| \leq \sum_{i=r}^{\infty} \sum_{j=r}^{\infty} |a_{ij}| < \infty$$

so by Theorem 10.3.9 again,

$$\sum_{j=r}^{\infty} \sum_{i=r}^{\infty} a_{ij}, \quad \sum_{i=r}^{\infty} \sum_{j=r}^{\infty} a_{ij}$$

both exist. It only remains to verify they are equal. Note $0 \leq (|a_{ij}| + a_{ij}) \leq |a_{ij}|$. Therefore, by Theorem 10.6.5 and Theorem 10.3.5 on Page 259

$$\begin{aligned} \sum_{j=r}^{\infty} \sum_{i=r}^{\infty} |a_{ij}| + \sum_{j=r}^{\infty} \sum_{i=r}^{\infty} a_{ij} &= \sum_{j=r}^{\infty} \sum_{i=r}^{\infty} (|a_{ij}| + a_{ij}) \\ &= \sum_{i=r}^{\infty} \sum_{j=r}^{\infty} (|a_{ij}| + a_{ij}) \\ &= \sum_{i=r}^{\infty} \sum_{j=r}^{\infty} |a_{ij}| + \sum_{i=r}^{\infty} \sum_{j=r}^{\infty} a_{ij} \\ &= \sum_{j=r}^{\infty} \sum_{i=r}^{\infty} |a_{ij}| + \sum_{i=r}^{\infty} \sum_{j=r}^{\infty} a_{ij} \end{aligned}$$

and so $\sum_{j=r}^{\infty} \sum_{i=r}^{\infty} a_{ij} = \sum_{i=r}^{\infty} \sum_{j=r}^{\infty} a_{ij}$ as claimed. This proves the theorem.

One of the most important applications of this theorem is to the problem of multiplication of series.

Definition 10.6.7 Let $\sum_{i=r}^{\infty} a_i$ and $\sum_{i=r}^{\infty} b_i$ be two series. For $n \geq r$, define

$$c_n \equiv \sum_{k=r}^n a_k b_{n-k+r}.$$

The series $\sum_{n=r}^{\infty} c_n$ is called the *Cauchy product* of the two series.

It isn't hard to see where this comes from. Formally write the following in the case $r = 0$:

$$(a_0 + a_1 + a_2 + a_3 \cdots)(b_0 + b_1 + b_2 + b_3 \cdots)$$

and start multiplying in the usual way. This yields

$$a_0 b_0 + (a_0 b_1 + b_0 a_1) + (a_0 b_2 + a_1 b_1 + a_2 b_0) + \cdots$$

and you see the expressions in parentheses above are just the c_n for $n = 0, 1, 2, \dots$. Therefore, it is reasonable to conjecture that

$$\sum_{i=r}^{\infty} a_i \sum_{j=r}^{\infty} b_j = \sum_{n=r}^{\infty} c_n$$

and of course there would be no problem with this in the case of finite sums but in the case of infinite sums, it is necessary to prove a theorem. The following is a special case of Merten's theorem.

Theorem 10.6.8 Suppose $\sum_{i=r}^{\infty} a_i$ and $\sum_{j=r}^{\infty} b_j$ both converge absolutely¹. Then

$$\left(\sum_{i=r}^{\infty} a_i \right) \left(\sum_{j=r}^{\infty} b_j \right) = \sum_{n=r}^{\infty} c_n$$

where

$$c_n = \sum_{k=r}^n a_k b_{n-k+r}.$$

Proof: Let $p_{nk} = 1$ if $r \leq k \leq n$ and $p_{nk} = 0$ if $k > n$. Then

$$c_n = \sum_{k=r}^{\infty} p_{nk} a_k b_{n-k+r}.$$

Also,

$$\begin{aligned} \sum_{k=r}^{\infty} \sum_{n=r}^{\infty} p_{nk} |a_k| |b_{n-k+r}| &= \sum_{k=r}^{\infty} |a_k| \sum_{n=r}^{\infty} p_{nk} |b_{n-k+r}| \\ &= \sum_{k=r}^{\infty} |a_k| \sum_{n=k}^{\infty} |b_{n-k+r}| \\ &= \sum_{k=r}^{\infty} |a_k| \sum_{n=k}^{\infty} |b_{n-(k-r)}| \\ &= \sum_{k=r}^{\infty} |a_k| \sum_{m=r}^{\infty} |b_m| < \infty. \end{aligned}$$

Therefore, by Theorem 10.6.6

$$\begin{aligned} \sum_{n=r}^{\infty} c_n &= \sum_{n=r}^{\infty} \sum_{k=r}^n a_k b_{n-k+r} = \sum_{n=r}^{\infty} \sum_{k=r}^{\infty} p_{nk} a_k b_{n-k+r} \\ &= \sum_{k=r}^{\infty} a_k \sum_{n=r}^{\infty} p_{nk} b_{n-k+r} = \sum_{k=r}^{\infty} a_k \sum_{n=k}^{\infty} b_{n-k+r} \\ &= \sum_{k=r}^{\infty} a_k \sum_{m=r}^{\infty} b_m \end{aligned}$$

and this proves the theorem.

¹Actually, it is only necessary to assume one of the series converges and the other converges absolutely. This is known as Merten's theorem and may be read in the 1974 book by Apostol listed in the bibliography.

10.7 Exercises

1. Determine whether the following series converge absolutely, conditionally, or not at all and give reasons for your answers.

$$\begin{aligned} \text{(a)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{1}{\sqrt{n^2+n+1}} \\ \text{(b)} \quad & \sum_{n=1}^{\infty} (-1)^n (\sqrt{n+1} - \sqrt{n}) \\ \text{(c)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{(n!)^2}{(2n)!} \\ \text{(d)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{(2n)!}{(n!)^2} \\ \text{(e)} \quad & \sum_{n=1}^{\infty} \frac{(-1)^n}{2n+2} \\ \text{(f)} \quad & \sum_{n=1}^{\infty} (-1)^n \left(\frac{n}{n+1} \right)^n \\ \text{(g)} \quad & \sum_{n=1}^{\infty} (-1)^n \left(\frac{n}{n+1} \right)^{n^2} \end{aligned}$$

2. Determine whether the following series converge absolutely, conditionally, or not at all and give reasons for your answers.

$$\begin{aligned} \text{(a)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{\ln(k^5)}{k} \\ \text{(b)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{\ln(k^5)}{k^{1.01}} \\ \text{(c)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{10^n}{(1.01)^n} \\ \text{(d)} \quad & \sum_{n=1}^{\infty} (-1)^n \sin\left(\frac{1}{n}\right) \\ \text{(e)} \quad & \sum_{n=1}^{\infty} (-1)^n \tan\left(\frac{1}{n^2}\right) \\ \text{(f)} \quad & \sum_{n=1}^{\infty} (-1)^n \cos\left(\frac{1}{n^2}\right) \\ \text{(g)} \quad & \sum_{n=1}^{\infty} (-1)^n \sin\left(\frac{\sqrt{n}}{n^2+1}\right) \end{aligned}$$

3. Determine whether the following series converge absolutely, conditionally, or not at all and give reasons for your answers.

$$\begin{aligned} \text{(a)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{2^n+n}{n2^n} \\ \text{(b)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{2^n+n}{n^2 2^n} \\ \text{(c)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{n}{2n+1} \\ \text{(d)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{10^n}{n!} \\ \text{(e)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{n^{100}}{1.01^n} \\ \text{(f)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{\ln n}{n^2} \\ \text{(g)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{3^n}{n^3} \\ \text{(h)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{n^3}{3^n} \\ \text{(i)} \quad & \sum_{n=1}^{\infty} (-1)^n \frac{n^3}{n!} \end{aligned}$$

$$\text{(j)} \quad \sum_{n=1}^{\infty} (-1)^n \frac{n!}{n^{100}}$$

4. Find the exact values of the following infinite series if they converge.

$$\begin{aligned} \text{(a)} \quad & \sum_{k=3}^{\infty} \frac{1}{k(k-2)} \\ \text{(b)} \quad & \sum_{k=1}^{\infty} \frac{1}{k(k+1)} \\ \text{(c)} \quad & \sum_{k=3}^{\infty} \frac{1}{(k+1)(k-2)} \\ \text{(d)} \quad & \sum_{k=1}^{\infty} \left(\frac{1}{\sqrt{k}} - \frac{1}{\sqrt{k+1}} \right) \\ \text{(e)} \quad & \sum_{n=1}^{\infty} \ln \left(\frac{(n+1)^2}{n(n+2)} \right) \end{aligned}$$

5. Suppose $\sum_{n=1}^{\infty} a_n$ converges absolutely. Can the same thing be said about $\sum_{n=1}^{\infty} a_n^2$? Explain.

6. A person says a series converges conditionally by the ratio test. Explain why his statement is total nonsense.

7. A person says a series diverges by the alternating series test. Explain why his statement is total nonsense.

8. Find a series which diverges using one test but converges using another if possible. If this is not possible, tell why.

9. If $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ both converge, can you conclude the sum, $\sum_{n=1}^{\infty} a_n b_n$ converges?

10. If $\sum_{n=1}^{\infty} a_n$ converges absolutely, and b_n is bounded, can you conclude $\sum_{n=1}^{\infty} a_n b_n$ converges? What if it is only the case that $\sum_{n=1}^{\infty} a_n$ converges?

11. The logarithm test states the following. Suppose $a_k \neq 0$ for large k and that

$$p = \lim_{k \rightarrow \infty} \frac{\ln \left(\frac{1}{|a_k|} \right)}{\ln k}$$

exists. If $p > 1$, then $\sum_{k=1}^{\infty} a_k$ converges absolutely. If $p < 1$, then the series, $\sum_{k=1}^{\infty} a_k$ does not converge absolutely. Prove this theorem.

12. Prove Theorem 10.5.4. **Hint:** For $\sum_{k=1}^{\infty} (-1)^k b_k$, show the odd partial sums are all at least as small as $\sum_{k=1}^{\infty} (-1)^k b_k$ and are increasing

while the even partial sums are at least as large as $\sum_{k=1}^{\infty} (-1)^k b_k$ and are decreasing.

13. Use Theorem 10.5.4 in the following alternating series to tell how large n must be so that

$$\left| \sum_{k=1}^{\infty} (-1)^k a_k - \sum_{k=1}^n (-1)^k a_k \right|$$

is no larger than the given number.

- $\sum_{k=1}^{\infty} (-1)^k \frac{1}{k}, .001$
 - $\sum_{k=1}^{\infty} (-1)^k \frac{1}{k^2}, .001$
 - $\sum_{k=1}^{\infty} (-1)^k \sin\left(\frac{1}{k}\right), .001$
 - $\sum_{k=1}^{\infty} (-1)^{k-1} \frac{1}{\sqrt{k}}, .001$
 - $\sum_{k=1}^{\infty} (-1)^k \frac{\ln k}{k}, .001$ Note that this one satisfies b_k is decreasing if $k > 3$ but not for all k . Does this matter?
14. For $1 \geq x \geq 0$, and $p \geq 1$, show that $(1-x)^p \geq 1 - px$. **Hint:** This can be done using the mean value theorem from calculus. Define $f(x) \equiv (1-x)^p - 1 + px$ and show that $f(0) = 0$ while $f'(x) \geq 0$ for all $x \in (0, 1)$.
15. Using the result of Problem 14 establish Raabe's Test, an interesting variation on the ratio test. This test says the following. Suppose there exists a constant, C and a number p such that

$$\left| \frac{a_{k+1}}{a_k} \right| \leq 1 - \frac{p}{k+C}$$

for all k large enough. Then if $p > 1$, it follows that $\sum_{k=1}^{\infty} a_k$ converges absolutely. **Hint:** Let $b_k \equiv k - 1 + C$ and note that for all k large enough, $b_k > 1$. Now conclude that there exists an integer, k_0 such that $b_{k_0} > 1$ and for all $k \geq k_0$ the given inequality above holds. Use Problem 14 to conclude that

$$\left| \frac{a_{k+1}}{a_k} \right| \leq 1 - \frac{p}{k+C}$$

$$\leq \left(1 - \frac{1}{k+C}\right)^p = \left(\frac{b_k}{b_{k+1}}\right)^p$$

showing that $|a_k| b_k^p$ is decreasing for $k \geq k_0$. Thus for some constant K

$$\begin{aligned} |a_k| &\leq K/b_k^p \\ &= K/(k-1+C)^p. \end{aligned}$$

Now use comparison theorems and the p series to obtain the conclusion of the theorem.

16. Consider the series $\sum_{k=0}^{\infty} (-1)^k \frac{1}{\sqrt{k+1}}$. Show this series converges and so it makes sense to write $\left(\sum_{k=0}^{\infty} (-1)^k \frac{1}{\sqrt{k+1}}\right)^2$. What about the Cauchy product of this series? Does it even converge? What does this mean about using algebra on infinite sums as though they were finite sums?
17. Verify Theorem 10.6.8 on the two series $\sum_{k=0}^{\infty} 2^{-k}$ and $\sum_{k=0}^{\infty} 3^{-k}$.
18. You can define infinite series of complex numbers in exactly the same way as infinite series of real numbers. That is $w = \sum_{k=1}^{\infty} z_k$ means: For every $\varepsilon > 0$ there exists N such that if $n \geq N$, then $\left|w - \sum_{k=1}^n z_k\right| < \varepsilon$. Here the absolute value is the one which applies to complex numbers. That is, $|a + ib| = \sqrt{a^2 + b^2}$. Show that if $\{b_n\}$ is a decreasing sequence of nonnegative numbers with the property that $\lim_{n \rightarrow \infty} b_n = 0$ and if ω is any complex number which is not equal to 1 but which satisfies $|\omega| = 1$, then $\sum_{k=1}^{\infty} \omega^n b_n$ must converge. Note a sequence of complex numbers, $\{a_n + ib_n\}$ converges to $a + ib$ if and only if $a_n \rightarrow a$ and $b_n \rightarrow b$. See Problem 9 on Page 69. There are quite a few things in this problem you should think about.

10.8 Power Series

10.8.1 Functions Defined In Terms Of Series

Earlier Taylor polynomials were used to approximate known functions such as $\sin x$ and $\ln(1+x)$. A much more exciting idea is to use infinite series as definitions of possibly new functions.

Definition 10.8.1 Let $\{a_k\}_{k=0}^{\infty}$ be a sequence of numbers. The expression,

$$\sum_{k=0}^{\infty} a_k (x-a)^k \quad (10.10)$$

is called a Taylor series centered at a . This is also called a power series centered at a .

In the above definition, x is a variable. Thus you can put in various values of x and ask whether the resulting series of numbers converges. Defining, D to be the set of all values of x such that the resulting series does converge, define a new function, f defined on D as

$$f(x) \equiv \sum_{k=0}^{\infty} a_k (x-a)^k.$$

This might be a totally new function, one which has no name. Nevertheless, much can be said about such functions. The following lemma is fundamental in considering the form of D which always turns out to be an interval centered at a which may or may not contain either end point.

Lemma 10.8.2 Suppose $z \in D$. Then if $|x-a| < |z-a|$, then $x \in D$ also and furthermore, the series $\sum_{k=0}^{\infty} |a_k| |x-a|^k$ converges.

Proof: Let $1 > r = |x-a|/|z-a|$. The n^{th} term test implies

$$\lim_{n \rightarrow \infty} |a_n| |z-a|^n = 0$$

and so for all n large enough,

$$|a_n| |z-a|^n < 1$$

so for such n ,

$$|a_n| |x-a|^n = |a_n| |z-a|^n \frac{|x-a|^n}{|z-a|^n} \leq \frac{|x-a|^n}{|z-a|^n} < r^n$$

Therefore, $\sum_{k=0}^{\infty} |a_k| |x-a|^k$ converges by comparison with the geometric series, $\sum r^n$.

With this lemma, the following fundamental theorem is obtained.

Theorem 10.8.3 Let $\sum_{k=0}^{\infty} a_k (x-a)^k$ be a Taylor series. Then there exists $r \leq \infty$ such that the Taylor series converges absolutely if $|x-a| < r$. Furthermore, if $|x-a| > r$, the Taylor series diverges.

Proof: Let

$$r \equiv \sup \{|y-a| : y \in D\}.$$

Then if $|x-a| < r$, it follows there exists $z \in D$ such that $|z-a| > |x-a|$ since otherwise, r wouldn't be as defined. In fact $|x-a|$ would then be an upper bound to $\{|y-a| : y \in D\}$. Therefore, by the above lemma $\sum_{k=0}^{\infty} |a_k| |x-a|^k$ converges and this proves the first part of this theorem.

Now suppose $|x - a| > r$. If $\sum_{k=0}^{\infty} a_k (x - a)^k$ converges then by the above lemma, r fails to be an upper bound to $\{|y - a| : y \in D\}$ and so the Taylor series must diverge as claimed. This proves the theorem.

From now on D will be referred to as the interval of convergence and r of the above theorem as the radius of convergence. Determining which points of $\{x : |x - a| = r\}$ are in D requires the use of specific convergence tests and can be quite hard. However, the determination of r tends to be pretty easy.

Example 10.8.4 Find the interval of convergence of the Taylor series $\sum_{n=1}^{\infty} \frac{x^n}{n}$.

Use Corollary 10.5.8.

$$\lim_{n \rightarrow \infty} \left(\frac{|x|^n}{n} \right)^{1/n} = \lim_{n \rightarrow \infty} \frac{|x|}{\sqrt[n]{n}} = |x|$$

because $\lim_{n \rightarrow \infty} \sqrt[n]{n} = 1$ and so if $|x| < 1$ the series converges. The endpoints require special attention. When $x = 1$ the series diverges because it reduces to $\sum_{n=1}^{\infty} \frac{1}{n}$. At the other endpoint, however, the series converges because it reduces to $\sum_{n=1}^{\infty} \frac{(-1)^n}{n}$ and the alternating series test applies and gives convergence.

Example 10.8.5 Find the radius of convergence of $\sum_{n=1}^{\infty} \frac{n^n}{n!} x^n$.

Apply the ratio test. Taking the ratio of the absolute values of the $(n+1)^{th}$ and the n^{th} terms

$$\frac{\frac{(n+1)^{(n+1)}}{(n+1)n!} |x|^{n+1}}{\frac{n^n}{n!} |x|^n} = (n+1)^n |x| n^{-n} = |x| \left(1 + \frac{1}{n}\right)^n \rightarrow |x| e$$

Therefore the series converges absolutely if $|x|e < 1$ and diverges if $|x|e > 1$. Consequently, $r = 1/e$.

10.8.2 Operations On Power Series

It is desirable to be able to differentiate, integrate, and multiply power series. The following theorem says one can differentiate power series in the most natural way on the interval of convergence, just as you would differentiate a polynomial. This theorem may seem obvious, but it is a serious mistake to think this. You usually cannot differentiate an infinite series whose terms are functions even if the functions are themselves polynomials. The following is special and pertains to power series. It is another example of the interchange of two limits, in this case, the limit involved in taking the derivative and the limit of the sequence of finite sums.

Theorem 10.8.6 Let $\sum_{n=0}^{\infty} a_n (x - a)^n$ be a Taylor series having radius of convergence $r > 0$ and let

$$f(x) \equiv \sum_{n=0}^{\infty} a_n (x - a)^n \quad (10.11)$$

for $|x - a| < r$. Then

$$f'(x) = \sum_{n=0}^{\infty} a_n n (x - a)^{n-1} = \sum_{n=1}^{\infty} a_n n (x - a)^{n-1} \quad (10.12)$$

and this new differentiated power series, the derived series, has radius of convergence equal to r .

Proof: First it will be shown that the series on the right in 10.12 has the same radius of convergence as the original series. Thus let $|x - a| < r$ and pick y such that

$$|x - a| < |y - a| < r.$$

Then

$$\lim_{n \rightarrow \infty} |a_n| |y - a|^{n-1} = \lim_{n \rightarrow \infty} |a_n| |y - a|^n = 0$$

because

$$\sum_{n=0}^{\infty} |a_n| |y - a|^n < \infty$$

and so, for n large enough,

$$|a_n| |y - a|^{n-1} < 1.$$

Therefore, for large enough n ,

$$\begin{aligned} |a_n| n |x - a|^{n-1} &= |a_n| |y - a|^{n-1} n \left| \frac{x - a}{y - a} \right|^{n-1} \\ &\leq n \left| \frac{x - a}{y - a} \right|^{n-1} \end{aligned}$$

and

$$\sum_{n=1}^{\infty} n \left| \frac{x - a}{y - a} \right|^{n-1}$$

converges by the ratio test. By the comparison test, it follows $\sum_{n=1}^{\infty} a_n n (x - a)^{n-1}$ converges absolutely for any x satisfying $|x - a| < r$. Therefore, the radius of convergence of the derived series is at least as large as that of the original series. On the other hand, if $\sum_{n=1}^{\infty} |a_n| n |x - a|^{n-1}$ converges then by the comparison test, $\sum_{n=1}^{\infty} |a_n| |x - a|^{n-1}$ and therefore $\sum_{n=1}^{\infty} |a_n| |x - a|^n$ also converges which shows the radius of convergence of the derived series is no larger than that of the original series. It remains to verify the assertion about the derivative.

Let $|x - a| < r$ and let $r_1 < r$ be close enough to r that

$$x \in (a - r_1, a + r_1) \subseteq [a - r_1, a + r_1] \subseteq (a - r, a + r).$$

Thus, letting $r_2 \in (r_1, r)$,

$$\sum_{n=0}^{\infty} |a_n| r_1^n, \sum_{n=0}^{\infty} |a_n| r_2^n < \infty \quad (10.13)$$

Letting y be close enough to x , it follows both x and y are in $[a - r_1, a + r_1]$. Then considering the difference quotient,

$$\begin{aligned} \frac{f(y) - f(x)}{y - x} &= \sum_{n=0}^{\infty} a_n (y - x)^{-1} [(y - a)^n - (x - a)^n] \\ &= \sum_{n=1}^{\infty} a_n n z_n^{n-1} \end{aligned} \quad (10.14)$$

where the last equation follows from the mean value theorem and z_n is some point between $x - a$ and $y - a$. Therefore,

$$\frac{f(y) - f(x)}{y - x} = \sum_{n=1}^{\infty} a_n n z_n^{n-1} =$$

$$\begin{aligned}
&= \sum_{n=1}^{\infty} a_n n \left(z_n^{n-1} - (x-a)^{n-1} \right) + \sum_{n=1}^{\infty} a_n n (x-a)^{n-1} \\
&= \sum_{n=2}^{\infty} a_n n (n-1) w_n^{n-2} (z_n - (x-a)) + \sum_{n=1}^{\infty} a_n n (x-a)^{n-1} \quad (10.15)
\end{aligned}$$

where w_n is between z_n and $x-a$. Thus w_n is between $x-a$ and $y-a$ and so

$$w_n + a \in [a - r_1, a + r_1]$$

which implies $|w_n| \leq r_1$. The first sum on the right in 10.15 therefore satisfies

$$\begin{aligned}
\left| \sum_{n=2}^{\infty} a_n n (n-1) w_n^{n-2} (z_n - (x-a)) \right| &\leq |y-x| \sum_{n=2}^{\infty} |a_n| n (n-1) |w_n|^{n-2} \\
&\leq |y-x| \sum_{n=2}^{\infty} |a_n| n (n-1) r_1^{n-2} \\
&= |y-x| \sum_{n=2}^{\infty} |a_n| r_2^{n-2} n (n-1) \left(\frac{r_1}{r_2} \right)^{n-2}
\end{aligned}$$

Now from 10.13, $|a_n| r_2^{n-2} < 1$ for all n large enough. Therefore, for such n ,

$$|a_n| r_2^{n-2} n (n-1) \left(\frac{r_1}{r_2} \right)^{n-2} \leq n (n-1) \left(\frac{r_1}{r_2} \right)^{n-2}$$

and the series $\sum n (n-1) \left(\frac{r_1}{r_2} \right)^{n-2}$ converges by the ratio test. Therefore, there exists a constant, C independent of y such that

$$\sum_{n=2}^{\infty} |a_n| n (n-1) r_1^{n-2} = C < \infty$$

Consequently, from 10.15

$$\left| \frac{f(y) - f(x)}{y - x} - \sum_{n=1}^{\infty} a_n n (x-a)^{n-1} \right| \leq C |y-x|.$$

Taking the limit as $y \rightarrow x$ 10.12 follows. This proves the theorem.

As an immediate corollary, it is possible to characterize the coefficients of a Taylor series.

Corollary 10.8.7 *Let $\sum_{n=0}^{\infty} a_n (x-a)^n$ be a Taylor series with radius of convergence $r > 0$ and let*

$$f(x) \equiv \sum_{n=0}^{\infty} a_n (x-a)^n. \quad (10.16)$$

Then

$$a_n = \frac{f^{(n)}(a)}{n!}. \quad (10.17)$$

Proof: From 10.16, $f(a) = a_0 \equiv f^{(0)}(a)/0!$. From Theorem 10.8.6,

$$f'(x) = \sum_{n=1}^{\infty} a_n n (x-a)^{n-1} = a_1 + \sum_{n=2}^{\infty} a_n n (x-a)^{n-1}.$$

Now let $x = a$ and obtain that $f'(a) = a_1 = f'(a)/1!$. Next use Theorem 10.8.6 again to take the second derivative and obtain

$$f''(x) = 2a_2 + \sum_{n=3}^{\infty} a_n n(n-1)(x-a)^{n-2}$$

let $x = a$ in this equation and obtain $a_2 = f''(a)/2 = f''(a)/2!$. Continuing this way proves the corollary.

This also shows the coefficients of a Taylor series are unique. That is, if

$$\sum_{k=0}^{\infty} a_k (x-a)^k = \sum_{k=0}^{\infty} b_k (x-a)^k$$

for all x in some interval, then $a_k = b_k$ for all k .

Example 10.8.8 Find the power series for $\sin(x)$, and $\cos(x)$ centered at 0 and give the interval of convergence.

First consider $f(x) = \sin(x)$. Then $f'(x) = \cos(x)$, $f''(x) = -\sin(x)$, $f'''(x) = -\cos(x)$ etc. Therefore, from Taylor's formula, Theorem 10.1.1 on Page 253,

$$f(x) = 0 + x + 0 - \frac{x^3}{3!} + 0 + \frac{x^5}{5!} + \cdots + \frac{x^{2n+1}}{(2n+1)!} + \frac{f^{(2n+2)}(\xi_n)}{(2n+2)!}$$

where ξ_n is some number between 0 and x . Furthermore, this equals either $\pm \sin(\xi_n)$ or $\pm \cos(\xi_n)$ and so its absolute value is no larger than 1. Thus

$$\left| \frac{f^{(2n+2)}(\xi_n)}{(2n+2)!} \right| \leq \frac{1}{(2n+2)!}.$$

By the ratio test, it follows that

$$\sum_{n=0}^{\infty} \frac{1}{(2n+2)!} < \infty$$

and so by the comparison test,

$$\sum_{n=0}^{\infty} \left| \frac{f^{(2n+2)}(\xi_n)}{(2n+2)!} \right| < \infty$$

also. Therefore, by the n^{th} term test $\lim_{n \rightarrow \infty} \frac{f^{(2n+2)}(\xi_n)}{(2n+2)!} = 0$. This implies

$$\sin(x) = \sum_{k=0}^n (-1)^k \frac{x^{2k+1}}{(2k+1)!} + \frac{f^{(2n+2)}(\xi_n)}{(2n+2)!}$$

and the last term converges to zero as $n \rightarrow \infty$ for any value of x and therefore,

$$\sin(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k+1}}{(2k+1)!}$$

for all $x \in \mathbb{R}$. By Theorem 10.8.6, you can differentiate both sides, doing the series term by term and obtain

$$\cos(x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{(2k)!}$$

for all $x \in \mathbb{R}$.

Example 10.8.9 Find the sum $\sum_{k=1}^{\infty} k2^{-k}$.

It may not be obvious what this sum equals but with the above theorem it is easy to find. From the formula for the sum of a geometric series, $\frac{1}{1-t} = \sum_{k=0}^{\infty} t^k$ if $|t| < 1$. Differentiate both sides to obtain

$$(1-t)^{-2} = \sum_{k=1}^{\infty} kt^{k-1}$$

whenever $|t| < 1$. Let $t = 1/2$. Then

$$4 = \frac{1}{(1-(1/2))^2} = \sum_{k=1}^{\infty} k2^{-(k-1)}$$

and so if you multiply both sides by 2^{-1} ,

$$2 = \sum_{k=1}^{\infty} k2^{-k}.$$

The following is a very important example known as the binomial series.

Example 10.8.10 Find a Taylor series for the function $(1+x)^\alpha$ centered at 0 valid for $|x| < 1$.

Use Theorem 10.8.6 to do this. First note that if $y(x) \equiv (1+x)^\alpha$, then y is a solution of the following initial value problem.

$$y' - \frac{\alpha}{(1+x)}y = 0, \quad y(0) = 1. \quad (10.18)$$

Next it is necessary to observe there is only one solution to this initial value problem. To see this, multiply both sides of the differential equation in 10.18 by $(1+x)^{-\alpha}$. When this is done one obtains

$$\frac{d}{dx} \left((1+x)^{-\alpha} y \right) = (1+x)^{-\alpha} \left(y' - \frac{\alpha}{(1+x)}y \right) = 0. \quad (10.19)$$

Therefore, from 10.19, there must exist a constant, C , such that

$$(1+x)^{-\alpha} y = C.$$

However, $y(0) = 1$ and so it must be that $C = 1$. Therefore, there is exactly one solution to the initial value problem in 10.18 and it is $y(x) = (1+x)^\alpha$. The strategy for finding the Taylor series of this function consists of finding a series which solves the initial value problem above. Let

$$y(x) \equiv \sum_{n=0}^{\infty} a_n x^n \quad (10.20)$$

be a solution to 10.18. Of course it is not known at this time whether such a series exists. However, the process of finding it will demonstrate its existence. From Theorem 10.8.6 and the initial value problem,

$$(1+x) \sum_{n=0}^{\infty} a_n n x^{n-1} - \sum_{n=0}^{\infty} \alpha a_n x^n = 0$$

and so

$$\sum_{n=1}^{\infty} a_n n x^{n-1} + \sum_{n=0}^{\infty} a_n (n - \alpha) x^n = 0$$

Changing the order variable of summation in the first sum,

$$\sum_{n=0}^{\infty} a_{n+1} (n+1) x^n + \sum_{n=0}^{\infty} a_n (n - \alpha) x^n = 0$$

and from Corollary 10.8.7 and the initial condition for 10.18 this requires

$$a_{n+1} = \frac{a_n (\alpha - n)}{n+1}, a_0 = 1. \quad (10.21)$$

Therefore, from 10.21 and letting $n = 0$, $a_1 = \alpha$. Then using 10.21 again along with this information, $a_2 = \frac{\alpha(\alpha-1)}{2}$. Using the same process, $a_3 = \frac{(\frac{\alpha(\alpha-1)}{2})(\alpha-2)}{3} = \frac{\alpha(\alpha-1)(\alpha-2)}{3!}$. By now you can spot the pattern. In general,

$$a_n = \frac{\overbrace{\alpha(\alpha-1) \cdots (\alpha-n+1)}^{n \text{ of these factors}}}{n!}.$$

Therefore, the candidate for the Taylor series is

$$y(x) = \sum_{n=0}^{\infty} \frac{\alpha(\alpha-1) \cdots (\alpha-n+1)}{n!} x^n.$$

Furthermore, the above discussion shows this series solves the initial value problem on its interval of convergence. It only remains to show the radius of convergence of this series equals 1. It will then follow that this series equals $(1+x)^\alpha$ because of uniqueness of the initial value problem. To find the radius of convergence, use the ratio test. Thus the ratio of the absolute values of $(n+1)^{st}$ term to the absolute value of the n^{th} term is

$$\frac{\left| \frac{\alpha(\alpha-1) \cdots (\alpha-n+1)(\alpha-n)}{(n+1)n!} \right| |x|^{n+1}}{\left| \frac{\alpha(\alpha-1) \cdots (\alpha-n+1)}{n!} \right| |x|^n} = |x| \frac{|\alpha-n|}{n+1} \rightarrow |x|$$

showing that the radius of convergence is 1 since the series converges if $|x| < 1$ and diverges if $|x| > 1$.

The expression, $\frac{\alpha(\alpha-1) \cdots (\alpha-n+1)}{n!}$ is often denoted as $\binom{\alpha}{n}$. With this notation, the following theorem has been established.

Theorem 10.8.11 *Let α be a real number and let $|x| < 1$. Then*

$$(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n.$$

There is a very interesting issue related to the above theorem which illustrates the limitation of power series. The function $f(x) = (1+x)^\alpha$ makes sense for all $x > -1$ but one is only able to describe it with a power series on the interval $(-1, 1)$. Think about this. The above technique is a standard one for obtaining solutions of differential equations and this example illustrates a deficiency in the method. To completely understand power series, it is necessary to take a course in complex analysis. You may have noticed the prominent role played by geometric series. This is no accident. It turns out that the right way to consider Taylor series is through the use of geometric series and something called the Cauchy integral formula of complex analysis. However, these are topics for another course.

You can also integrate power series on their interval of convergence.

Theorem 10.8.12 Let $f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n$ and suppose the radius of convergence is $r > 0$. Then if $|y-a| < r$,

$$\int_a^y f(x) dx = \sum_{n=0}^{\infty} \int_a^y a_n (x-a)^n dx = \sum_{n=0}^{\infty} \frac{a_n (y-a)^{n+1}}{n+1}.$$

Proof: Define $F(y) \equiv \int_a^y f(x) dx$ and $G(y) \equiv \sum_{n=0}^{\infty} \frac{a_n (y-a)^{n+1}}{n+1}$. By Theorem 10.8.6 and the Fundamental theorem of calculus,

$$G'(y) = \sum_{n=0}^{\infty} a_n (y-a)^n = f(y) = F'(y).$$

Therefore, $G(y) - F(y) = C$ for some constant. But $C = 0$ because $F(a) - G(a) = 0$. This proves the theorem.

Next consider the problem of multiplying two power series.

Theorem 10.8.13 Let $\sum_{n=0}^{\infty} a_n (x-a)^n$ and $\sum_{n=0}^{\infty} b_n (x-a)^n$ be two power series having radii of convergence r_1 and r_2 , both positive. Then

$$\left(\sum_{n=0}^{\infty} a_n (x-a)^n \right) \left(\sum_{n=0}^{\infty} b_n (x-a)^n \right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_k b_{n-k} \right) (x-a)^n$$

whenever $|x-a| < r \equiv \min(r_1, r_2)$.

Proof: By Theorem 10.8.3 both series converge absolutely if $|x-a| < r$. Therefore, by Theorem 10.6.8

$$\begin{aligned} & \left(\sum_{n=0}^{\infty} a_n (x-a)^n \right) \left(\sum_{n=0}^{\infty} b_n (x-a)^n \right) = \\ & \sum_{n=0}^{\infty} \sum_{k=0}^n a_k (x-a)^k b_{n-k} (x-a)^{n-k} = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n a_k b_{n-k} \right) (x-a)^n. \end{aligned}$$

This proves the theorem.

The significance of this theorem in terms of applications is that it states you can multiply power series just as you would multiply polynomials and everything will be all right on the common interval of convergence.

This theorem can be used to find Taylor series which would perhaps be hard to find without it. Here is an example.

Example 10.8.14 Find the Taylor series for $e^x \sin x$ centered at $x = 0$.

Using Problems 11 - 13 on Page 256 or Example 10.8.8 on Page 279, and Example 10.5.9 on Page 267, all that is required is to multiply

$$\left(\overbrace{1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} \cdots}^{e^x} \right) \left(\overbrace{x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots}^{\sin x} \right)$$

From the above theorem the result should be

$$x + x^2 + \left(-\frac{1}{3!} + \frac{1}{2!} \right) x^3 + \cdots$$

$$= x + x^2 + \frac{1}{3}x^3 + \cdots$$

You can continue this way and get the following to a few more terms.

$$x + x^2 + \frac{1}{3}x^3 - \frac{1}{30}x^5 - \frac{1}{90}x^6 - \frac{1}{630}x^7 + \cdots$$

I don't see a pattern in these coefficients but I can go on generating them as long as I want. (In practice this tends to not be very long.) I also know the resulting power series will converge for all x because both the series for e^x and the one for $\sin x$ converge for all x .

Example 10.8.15 Find the Taylor series for $\tan x$ centered at $x = 0$.

Lets suppose it has a Taylor series $a_0 + a_1x + a_2x^2 + \cdots$. Then

$$(a_0 + a_1x + a_2x^2 + \cdots) \left(\overbrace{1 - \frac{x^2}{2} + \frac{x^4}{4!} + \cdots}^{\cos x} \right) = \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots \right).$$

Using the above, $a_0 = 0$, $a_1x = x$ so $a_1 = 1$, $(0 - \frac{1}{2}) + a_2)x^2 = 0$ so $a_2 = 0$.

$$(a_3 - \frac{a_1}{2})x^3 = \frac{-1}{3!}x^3$$

so $a_3 - \frac{1}{2} = -\frac{1}{6}$ so $a_3 = \frac{1}{3}$. Clearly one can continue in this manner. Thus the first several terms of the power series for $\tan x$ are

$$\tan x = x + \frac{1}{3}x^3 + \cdots$$

You can go on calculating these terms and find the next two yielding

$$\tan x = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + \frac{17}{315}x^7 + \cdots$$

This is a very significant technique because, as you see, there does not appear to be a very simple pattern for the coefficients of the power series for $\tan x$. Of course there are some issues here about whether $\tan x$ even has a power series, but if it does, the above must be it. In fact, $\tan(x)$ will have a power series valid on some interval centered at 0 and this becomes completely obvious when one uses methods from complex analysis but it isn't too obvious at this point. If you are interested in this issue, read the last section of the chapter. Note also that what has been accomplished is to divide the power series for $\sin x$ by the power series for $\cos x$ just like they were polynomials.

10.9 Exercises

1. Find the radius of convergence of the following.

- (a) $\sum_{k=1}^{\infty} \left(\frac{x}{2}\right)^n$
- (b) $\sum_{k=1}^{\infty} \sin\left(\frac{1}{n}\right) 3^n x^n$
- (c) $\sum_{k=0}^{\infty} k! x^k$
- (d) $\sum_{n=0}^{\infty} \frac{(3n)^n}{(3n)!} x^n$
- (e) $\sum_{n=0}^{\infty} \frac{(2n)^n}{(2n)!} x^n$

2. Find $\sum_{k=1}^{\infty} k2^{-k}$.
3. Find $\sum_{k=1}^{\infty} k^2 3^{-k}$.
4. Find $\sum_{k=1}^{\infty} \frac{2^{-k}}{k}$.
5. Find $\sum_{k=1}^{\infty} \frac{3^{-k}}{k}$.
6. Find where the series $\sum_{k=1}^{\infty} \frac{1-e^{-kx}}{k}$ converges.
7. Find the power series centered at 0 for the function $1/(1+x^2)$ and give the radius of convergence. Where does the function make sense? Where does the power series equal the function?
8. Find a power series for the function, $f(x) \equiv \frac{\sin(\sqrt{x})}{\sqrt{x}}$ for $x > 0$. Where does $f(x)$ make sense? Where does the power series you found converge?
9. Use the power series technique which was applied in Example 10.8.10 to consider the initial value problem $y' = y, y(0) = 1$. This yields another way to obtain the power series for e^x .
10. Use the power series technique on the initial value problem $y' + y = 0, y(0) = 1$. What is the solution to this initial value problem?
11. Use the power series technique to find the first several nonzero terms in the power series solution to the initial value problem

$$y'' + xy = 0, y(0) = 0, y'(0) = 1.$$

Tell where your solution gives a valid description of a solution for the initial value problem. **Hint:** This is a little different but you proceed the same way as in Example 10.8.10. The main difference is you have to do two differentiations of the power series instead of one.

12. Suppose the function, e^x is defined in terms of a power series, $e^x \equiv \sum_{k=0}^{\infty} \frac{x^k}{k!}$. Use Theorem 10.6.8 on Page 272 to show directly the usual law of exponents,

$$e^{x+y} = e^x e^y.$$

Be sure to check all the hypotheses.

13. Define the following function²:

$$f(x) \equiv \begin{cases} e^{-(1/x^2)} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

Show that $f^{(k)}(x)$ exists for all k and for all x . Show also that $f^{(k)}(0) = 0$ for all $k \in \mathbb{N}$. Therefore, the power series for $f(x)$ is of the form $\sum_{k=0}^{\infty} 0x^k$ and it converges for all values of x . However, it fails to converge to $f(x)$ except at the single point, $x = 0$.

²Surprisingly, this function is very important to those who use modern techniques to study differential equations. One needs to consider test functions which have the property they have infinitely many derivatives but vanish outside of some interval. The theory of complex variables can be used to show there are no examples of such functions if they have a valid power series expansion. It even becomes a little questionable whether such strange functions even exist at all. Nevertheless, they do, there are enough of them, and it is this very example which is used to show this.

14. Let $f_n(x) \equiv \left(\frac{1}{n} + x^2\right)^{1/2}$. Show that for all x ,

$$||x| - f_n(x)| \leq \frac{1}{\sqrt{n}}.$$

Now show $f'_n(0) = 0$ for all n and so $f'_n(0) \rightarrow 0$. However, the function, $f(x) \equiv |x|$ has no derivative at $x = 0$. Thus even though $f_n(x) \rightarrow f(x)$ for all x , you cannot say that $f'_n(0) \rightarrow f'(0)$.

15. Let the functions, $f_n(x)$ be given in Problem 14 and consider

$$g_1(x) = f_1(x), \quad g_n(x) = f_n(x) - f_{n-1}(x) \text{ if } n > 1.$$

Show that for all x ,

$$\sum_{k=1}^{\infty} g_k(x) = |x|$$

and that $g'_k(0) = 0$ for all k . Therefore, you can't differentiate the series term by term and get the right answer³.

16. Use the theorem about the binomial series to give a proof of the binomial theorem

$$(a+b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$$

whenever n is a positive integer.

17. You know $\int_0^x \frac{1}{t+1} dt = \ln|1+x|$. Use this and Theorem 10.8.12 to find the power series for $\ln|1+x|$ centered at 0. Where does this power series converge? Where does it converge to the function, $\ln|1+x|$?
18. You know $\int_0^x \frac{1}{t^2+1} dt = \arctan x$. Use this and Theorem 10.8.12 to find the power series for $\arctan x$ centered at 0. Where does this power series converge? Where does it converge to the function, $\arctan x$?
19. Find the power series for $\sin(x^2)$ by plugging in x^2 where ever there is an x in the power series for $\sin x$. How do you know this is the power series for $\sin(x^2)$?
20. Find the first several terms of the power series for $\sin^2(x)$ by multiplying the power series for $\sin(x)$. Next use the trig. identity, $\sin^2(x) = \frac{1-\cos(2x)}{2}$ and the power series for $\cos(2x)$ to find the power series.
21. Find the power series for $f(x) = \frac{1}{\sqrt{1-x^2}}$.
22. It is hard to find $\int_0^1 e^{x^2} dx$ because you don't have a convenient antiderivative for the integrand. Replace e^{x^2} with an appropriate power series and estimate this integral.
23. Do the same as the previous problem for $\int_0^1 \sin(x^2) dx$.
24. Find $\lim_{x \rightarrow 0} \frac{\tan(\sin x) - \sin(\tan x)}{x^7} \cdot 4$.

³How bad can this get? It can be much worse than this. In fact, there are functions which are continuous everywhere and differentiable nowhere. We typically don't have names for them but they are there just the same. Every such function can be written as an infinite sum of polynomials which of course have derivatives at every point. Thus it is nonsense to differentiate an infinite sum term by term without a theorem of some sort.

⁴This is a wonderful example. You should plug in small values of x using a calculator and see what you get using modern technology.

25. Consider the function, $S(x) \equiv \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^{2n-1}}{(2n-1)!}$. This is the power series for $\sin(x)$ but pretend you don't know this. Show that the series for $S(x)$ converges for all $x \in \mathbb{R}$. Also show that S satisfies the initial value problem $y'' + y = 0, y(0) = 0, y'(0) = 1$.
26. Consider the function, $C(x) \equiv \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$. This is the power series for $\cos(x)$ but pretend you don't know this. Show that the series for $C(x)$ converges for all $x \in \mathbb{R}$. Also show that S satisfies the initial value problem $y'' + y = 0, y(0) = 1, y'(0) = 0$.
27. Show there is at most one solution to the initial value problem, $y'' + y = 0, y(0) = a, y'(0) = b$ and find the solution to this problem in terms of $C(x)$ and $S(x)$. Also show directly from the series descriptions for $C(x)$ and $S(x)$ that $S'(x) = C(x)$ and $C'(x) = -S(x)$.
28. * Using problem 27 about uniqueness of the initial value problem, show that

$$C(x+y) = C(x)C(y) - S(x)S(y)$$

and

$$S(x+y) = S(x)C(y) + S(y)C(x).$$

Do this in the following way: Fix y and consider the function $f(x) \equiv C(x+y)$ and

$$g(x) = C(x)C(y) - S(x)S(y).$$

Then show both f and g satisfy the same initial value problem and so they must be equal. Do the other identity the same way. Also show $S(-x) = -S(x)$ and $C(-x) = C(x)$ and $S(x)^2 + C(x)^2 = 1$. This last claim is really easy. Just take the derivative and see $S^2 + C^2$ must be constant.

29. *You know $S(0) = 0$ and $C(0) = 1$. Show there exists $T > 0$ such that on $(0, T)$ both $S(x)$ and $C(x)$ are positive but $C(T) = 0$ while $S(T) = 1$. (We usually refer to T as $\frac{\pi}{2}$.) To do this, note that $S'(0) > 0$ and so S is an increasing function on some interval. Therefore, C is a decreasing function on that interval because of $S^2 + C^2 = 1$. If C is bounded below by some positive number, then S must be unbounded because $S' = C$. However this would contradict $S^2 + C^2 = 1$. Therefore, $C(T) = 0$ for some T . Let T be the first time this occurs. You fill in the mathematical details of this argument. Next show that on $(T, 2T)$, $S(x) > 0$ and $C(x) < 0$ and on $(2T, 3T)$, both $C(x)$ and $S(x)$ are negative. Finally, show that on $(3T, 4T)$, $C(x) > 0$ and $S(x) < 0$. Also show $C(x+2T) = C(x)$ and $S(x+2T) = S(x)$. Do all this without resorting to identifying $S(x)$ with $\sin x$ and $C(x)$ with $\cos x$. Finally explain why $\sin x = S(x)$ for all x and $C(x) = \cos x$ for all x .

Note: Problems 25 - 29 outline a way to define the circular functions with no reference to plane geometry. The job is not finished because these circular functions were defined as the x and y coordinates of a point on the unit circle where the angle was measured in terms of arc length and I have not yet tied it in to arc length. This is very easy to do later. Taking the Pythagorean theorem as the **definition** of length, a precise description of what is meant by arc length in terms of integrals can be presented. When this is done, it is possible to **define** $\sin x$ and $\cos x$ in terms of these power series described above and totally eliminate all references to plane geometry. This approach is vastly superior to the traditional approach presented earlier in this book. Calculus is different than geometry and so it is desirable to obtain descriptions of the important functions which are free of geometry.

10.10 Some Other Theorems

First recall Theorem 10.6.8 on Page 272. For convenience, the version of this theorem which is of interest here is listed below.

Theorem 10.10.1 Suppose $\sum_{i=0}^{\infty} a_i$ and $\sum_{j=0}^{\infty} b_j$ both converge absolutely. Then

$$\left(\sum_{i=0}^{\infty} a_i \right) \left(\sum_{j=0}^{\infty} b_j \right) = \sum_{n=0}^{\infty} c_n$$

where

$$c_n = \sum_{k=0}^n a_k b_{n-k}.$$

Furthermore, $\sum_{n=0}^{\infty} c_n$ converges absolutely.

Proof: It only remains to verify the last series converges absolutely. By Theorem 10.6.5 on Page 270 and letting p_{nk} be as defined there,

$$\begin{aligned} \sum_{n=0}^{\infty} |c_n| &= \sum_{n=0}^{\infty} \left| \sum_{k=0}^n a_k b_{n-k} \right| \\ &\leq \sum_{n=0}^{\infty} \sum_{k=0}^n |a_k| |b_{n-k}| = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} p_{nk} |a_k| |b_{n-k}| \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} p_{nk} |a_k| |b_{n-k}| = \sum_{k=0}^{\infty} \sum_{n=k}^{\infty} |a_k| |b_{n-k}| \\ &= \sum_{k=0}^{\infty} |a_k| \sum_{n=0}^{\infty} |b_n| < \infty. \end{aligned}$$

This proves the theorem.

The theorem is about multiplying two series. What if you wanted to consider

$$\left(\sum_{n=0}^{\infty} a_n \right)^p$$

where p is a positive integer maybe larger than 2? Is there a similar theorem to the above?

Definition 10.10.2 Define

$$\sum_{k_1 + \cdots + k_p = m} a_{k_1} a_{k_2} \cdots a_{k_p}$$

as follows. Consider all ordered lists of nonnegative integers $k_1, \dots, k_p$ which have the property that $\sum_{i=1}^p k_i = m$. For each such list of integers, form the product, $a_{k_1} a_{k_2} \cdots a_{k_p}$ and then add all these products.

Note that

$$\sum_{k=0}^n a_k a_{n-k} = \sum_{k_1 + k_2 = n} a_{k_1} a_{k_2}$$

Therefore, from the above theorem, if $\sum a_i$ converges absolutely, it follows

$$\left(\sum_{i=0}^{\infty} a_i\right)^2 = \sum_{n=0}^{\infty} \left(\sum_{k_1+k_2=n} a_{k_1} a_{k_2}\right).$$

It turns out a similar theorem holds for replacing 2 with p .

Theorem 10.10.3 Suppose $\sum_{n=0}^{\infty} a_n$ converges absolutely. Then

$$\left(\sum_{n=0}^{\infty} a_n\right)^p = \sum_{m=0}^{\infty} c_{mp}$$

where

$$c_{mp} \equiv \sum_{k_1+\dots+k_p=m} a_{k_1} \cdots a_{k_p}.$$

Proof: First note this is obviously true if $p = 1$ and is also true if $p = 2$ from the above theorem. Now suppose this is true for p and consider $(\sum_{n=0}^{\infty} a_n)^{p+1}$. By the induction hypothesis and the above theorem on the Cauchy product,

$$\begin{aligned} \left(\sum_{n=0}^{\infty} a_n\right)^{p+1} &= \left(\sum_{n=0}^{\infty} a_n\right)^p \left(\sum_{n=0}^{\infty} a_n\right) \\ &= \left(\sum_{m=0}^{\infty} c_{mp}\right) \left(\sum_{n=0}^{\infty} a_n\right) \\ &= \sum_{n=0}^{\infty} \left(\sum_{k=0}^n c_{kp} a_{n-k}\right) \\ &= \sum_{n=0}^{\infty} \sum_{k=0}^n \sum_{k_1+\dots+k_p=k} a_{k_1} \cdots a_{k_p} a_{n-k} \\ &= \sum_{n=0}^{\infty} \sum_{k_1+\dots+k_{p+1}=n} a_{k_1} \cdots a_{k_{p+1}} \end{aligned}$$

and this proves the theorem.

This theorem implies the following corollary for power series.

Corollary 10.10.4 Let

$$\sum_{n=0}^{\infty} a_n (x-a)^n$$

be a power series having radius of convergence, $r > 0$. Then if $|x-a| < r$,

$$\left(\sum_{n=0}^{\infty} a_n (x-a)^n\right)^p = \sum_{n=0}^{\infty} b_{np} (x-a)^n$$

where

$$b_{np} \equiv \sum_{k_1+\dots+k_p=n} a_{k_1} \cdots a_{k_p}.$$

Proof: Since $|x - a| < r$, the series, $\sum_{n=0}^{\infty} a_n (x - a)^n$, converges absolutely. Therefore, the above theorem applies and

$$\begin{aligned} \left(\sum_{n=0}^{\infty} a_n (x - a)^n \right)^p &= \\ \sum_{n=0}^{\infty} \left(\sum_{k_1 + \dots + k_p = n} a_{k_1} (x - a)^{k_1} \dots a_{k_p} (x - a)^{k_p} \right) &= \\ \sum_{n=0}^{\infty} \left(\sum_{k_1 + \dots + k_p = n} a_{k_1} \dots a_{k_p} \right) (x - a)^n. \end{aligned}$$

With this theorem it is possible to consider the question raised in Example 10.8.15 on Page 283 about the existence of the power series for $\tan x$. This question is clearly included in the more general question of when

$$\left(\sum_{n=0}^{\infty} a_n (x - a)^n \right)^{-1}$$

has a power series.

Lemma 10.10.5 *Let $f(x) = \sum_{n=0}^{\infty} a_n (x - a)^n$, a power series having radius of convergence $r > 0$. Suppose also that $f(a) = 1$. Then there exists $r_1 > 0$ and $\{b_n\}$ such that for all $|x - a| < r_1$,*

$$\frac{1}{f(x)} = \sum_{n=0}^{\infty} b_n (x - a)^n.$$

Proof: By continuity, there exists $r_1 > 0$ such that if $|x - a| < r_1$, then

$$\sum_{n=1}^{\infty} |a_n| |x - a|^n < 1.$$

Now pick such an x . Then

$$\begin{aligned} \frac{1}{f(x)} &= \frac{1}{1 + \sum_{n=1}^{\infty} a_n (x - a)^n} \\ &= \frac{1}{1 + \sum_{n=0}^{\infty} c_n (x - a)^n} \end{aligned}$$

where $c_n = a_n$ if $n > 0$ and $c_0 = 0$. Then

$$\left| \sum_{n=1}^{\infty} a_n (x - a)^n \right| \leq \sum_{n=1}^{\infty} |a_n| |x - a|^n < 1 \quad (10.22)$$

and so from the formula for the sum of a geometric series,

$$\frac{1}{f(x)} = \sum_{p=0}^{\infty} \left(\sum_{n=0}^{\infty} c_n (x - a)^n \right)^p.$$

By Corollary 10.10.4, this equals

$$\sum_{p=0}^{\infty} \sum_{n=0}^{\infty} b_{np} (x - a)^n \quad (10.23)$$

where

$$b_{np} = \sum_{k_1 + \dots + k_p = n} c_{k_1} \cdots c_{k_p}.$$

Thus $|b_{np}| \leq \sum_{k_1 + \dots + k_p = n} |c_{k_1}| \cdots |c_{k_p}| \equiv B_{np}$ and so by Theorem 10.10.3,

$$\begin{aligned} \sum_{p=0}^{\infty} \sum_{n=0}^{\infty} |b_{np}| |x-a|^n &\leq \sum_{p=0}^{\infty} \sum_{n=0}^{\infty} B_{np} |x-a|^n \\ &= \sum_{p=0}^{\infty} \left(\sum_{n=0}^{\infty} |c_n| |x-a|^n \right)^p < \infty \end{aligned}$$

by 10.22 and the formula for the sum of a geometric series. Since the series of 10.23 converges absolutely, Theorem 10.6.5 on Page 270 implies the series in 10.23 equals

$$\sum_{n=0}^{\infty} \left(\sum_{p=0}^{\infty} b_{np} \right) (x-a)^n$$

and so, letting $\sum_{p=0}^{\infty} b_{np} \equiv b_n$, this proves the lemma.

With this lemma, the following theorem is easy to obtain.

Theorem 10.10.6 *Let $f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n$, a power series having radius of convergence $r > 0$. Suppose also that $f(a) \neq 0$. Then there exists $r_1 > 0$ and $\{b_n\}$ such that for all $|x-a| < r_1$,*

$$\frac{1}{f(x)} = \sum_{n=0}^{\infty} b_n (x-a)^n.$$

Proof: Let $g(x) \equiv f(x)/f(a)$ so that $g(x)$ satisfies the conditions of the above lemma. Then by that lemma, there exists $r_1 > 0$ and a sequence, $\{b_n\}$ such that

$$\frac{f(a)}{f(x)} = \sum_{n=0}^{\infty} b_n (x-a)^n$$

for all $|x-a| < r_1$. Then

$$\frac{1}{f(x)} = \sum_{n=0}^{\infty} \tilde{b}_n (x-a)^n$$

where $\tilde{b}_n = b_n/f(a)$. This proves the theorem.

There is a very interesting question related to r_1 in this theorem. One might think that if $|x-a| < r$, the radius of convergence of $f(x)$ and if $f(x) \neq 0$ it should be possible to write $1/f(x)$ as a power series centered at a . Unfortunately this is not true. Consider $f(x) = 1+x^2$. In this case $r = \infty$ but the power series for $1/f(x)$ converges only if $|x| < 1$. What happens is this, $1/f(x)$ will have a power series that will converge for $|x-a| < r_1$ where r_1 is the distance between a and the nearest singularity or zero of $f(x)$ in the complex plane. In the case of $f(x) = 1+x^2$ this function has a zero at $x = \pm i$. This is just another instance of why the natural setting for the study of power series is the complex plane. To read more on power series, you should see the book by Apostol [3] or any text on complex variable.

Part II

Basic Linear Algebra

Fundamentals

11.0.1 Outcomes

1. Describe $\mathbb{R}^n$ and do algebra with vectors in $\mathbb{R}^n$.
2. Evaluate the distance between two points in $\mathbb{R}^n$.
3. Represent a line in 3 space by a vector parametrization, a set of scalar parametric equations or using symmetric form.
4. Find a parametrization of a line given information about
 - (a) a point of the line and the direction of the line
 - (b) two points contained in the line
5. Determine the direction of a line given its parametrization.

11.1 $\mathbb{R}^n$

The notation, $\mathbb{R}^n$ refers to the collection of ordered lists of n real numbers. More precisely, consider the following definition.

Definition 11.1.1 *Define*

$$\mathbb{R}^n \equiv \{(x_1, \dots, x_n) : x_j \in \mathbb{R} \text{ for } j = 1, \dots, n\}.$$

$(x_1, \dots, x_n) = (y_1, \dots, y_n)$ if and only if for all $j = 1, \dots, n$, $x_j = y_j$. When

$$(x_1, \dots, x_n) \in \mathbb{R}^n,$$

it is conventional to denote $(x_1, \dots, x_n)$ by the single bold face letter, $\mathbf{x}$. The numbers, x_j are called the **coordinates**. The set

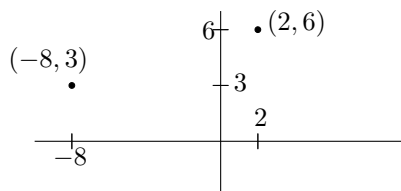
$$\{(0, \dots, 0, t, 0, \dots, 0) : t \in \mathbb{R}\}$$

for t in the i^{th} slot is called the i^{th} coordinate axis **coordinate axis**, the x_i axis for short. The point $\mathbf{0} \equiv (0, \dots, 0)$ is called the **origin**. Points in $\mathbb{R}^n$ are also called **vectors**.

Thus $(1, 2, 4) \in \mathbb{R}^3$ and $(2, 1, 4) \in \mathbb{R}^3$ but $(1, 2, 4) \neq (2, 1, 4)$ because, even though the same numbers are involved, they don't match up. In particular, the first entries are not equal.

Why would anyone be interested in such a thing? First consider the case when $n = 1$. Then from the definition, $\mathbb{R}^1 = \mathbb{R}$. Recall that $\mathbb{R}$ is identified with the points of

a line. Look at the number line again. Observe that this amounts to identifying a point on this line with a real number. In other words a real number determines where you are on this line. Now suppose $n = 2$ and consider two lines which intersect each other at right angles as shown in the following picture.



Notice how you can identify a point shown in the plane with the ordered pair, $(2, 6)$. You go to the right a distance of 2 and then up a distance of 6. Similarly, you can identify another point in the plane with the ordered pair $(-8, 3)$. Go to the left a distance of 8 and then up a distance of 3. The reason you go to the left is that there is a $-$ sign on the eight. From this reasoning, every ordered pair determines a unique point in the plane. Conversely, taking a point in the plane, you could draw two lines through the point, one vertical and the other horizontal and determine unique points, x_1 on the horizontal line in the above picture and x_2 on the vertical line in the above picture, such that the point of interest is identified with the ordered pair, (x_1, x_2) . In short, points in the plane can be identified with ordered pairs similar to the way that points on the real line are identified with real numbers. Now suppose $n = 3$. As just explained, the first two coordinates determine a point in a plane. Letting the third component determine how far up or down you go, depending on whether this number is positive or negative, this determines a point in space. Thus, $(1, 4, -5)$ would mean to determine the point in the plane that goes with $(1, 4)$ and then to go below this plane a distance of 5 to obtain a unique point in space. You see that the ordered triples correspond to points in space just as the ordered pairs correspond to points in a plane and single real numbers correspond to points on a line.

You can't stop here and say that you are only interested in $n \leq 3$. What if you were interested in the motion of two objects? You would need three coordinates to describe where the first object is and you would need another three coordinates to describe where the other object is located. Therefore, you would need to be considering $\mathbb{R}^6$. If the two objects moved around, you would need a time coordinate as well. As another example, consider a hot object which is cooling and suppose you want the temperature of this object. How many coordinates would be needed? You would need one for the temperature, three for the position of the point in the object and one more for the time. Thus you would need to be considering $\mathbb{R}^5$. Many other examples can be given. Sometimes n is very large. This is often the case in applications to business when they are trying to maximize profit subject to constraints. It also occurs in numerical analysis when people try to solve hard problems on a computer.

There are other ways to identify points in space with three numbers but the one presented is the most basic. In this case, the coordinates are known as **Cartesian coordinates** after Descartes¹ who invented this idea in the first half of the seventeenth century. I will often not bother to draw a distinction between the point in n dimensional space and its Cartesian coordinates.

¹René Descartes 1596-1650 is often credited with inventing analytic geometry although it seems the ideas were actually known much earlier. He was interested in many different subjects, physiology, chemistry, and physics being some of them. He also wrote a large book in which he tried to explain the book of Genesis scientifically. Descartes ended up dying in Sweden.

11.2 Algebra in $\mathbb{R}^n$

There are two algebraic operations done with points of $\mathbb{R}^n$. One is addition and the other is multiplication by numbers, called scalars.

Definition 11.2.1 If $\mathbf{x} \in \mathbb{R}^n$ and a is a number, also called a **scalar**, then $a\mathbf{x} \in \mathbb{R}^n$ is defined by

$$a\mathbf{x} = a(x_1, \dots, x_n) \equiv (ax_1, \dots, ax_n). \quad (11.1)$$

This is known as **scalar multiplication**. If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ then $\mathbf{x} + \mathbf{y} \in \mathbb{R}^n$ and is defined by

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (x_1, \dots, x_n) + (y_1, \dots, y_n) \\ &\equiv (x_1 + y_1, \dots, x_n + y_n) \end{aligned} \quad (11.2)$$

An element of $\mathbb{R}^n$, $\mathbf{x} \equiv (x_1, \dots, x_n)$ is often called a **vector**. The above definition is known as **vector addition**.

With this definition, the algebraic properties satisfy the conclusions of the following theorem.

Theorem 11.2.2 For $\mathbf{v}, \mathbf{w}$ vectors in $\mathbb{R}^n$ and α, β scalars, (real numbers), the following hold.

$$\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}, \quad (11.3)$$

the commutative law of addition,

$$(\mathbf{v} + \mathbf{w}) + \mathbf{z} = \mathbf{v} + (\mathbf{w} + \mathbf{z}), \quad (11.4)$$

the associative law for addition,

$$\mathbf{v} + \mathbf{0} = \mathbf{v}, \quad (11.5)$$

the existence of an additive identity,

$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0}, \quad (11.6)$$

the existence of an additive inverse, Also

$$\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}, \quad (11.7)$$

$$(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}, \quad (11.8)$$

$$\alpha(\beta\mathbf{v}) = \alpha\beta(\mathbf{v}), \quad (11.9)$$

$$1\mathbf{v} = \mathbf{v}. \quad (11.10)$$

In the above $\mathbf{0} = (0, \dots, 0)$.

You should verify these properties all hold. For example, consider 11.7

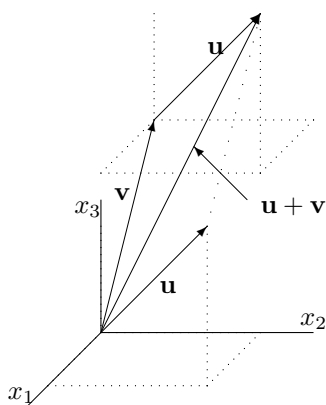
$$\begin{aligned} \alpha(\mathbf{v} + \mathbf{w}) &= \alpha(v_1 + w_1, \dots, v_n + w_n) \\ &= (\alpha(v_1 + w_1), \dots, \alpha(v_n + w_n)) \\ &= (\alpha v_1 + \alpha w_1, \dots, \alpha v_n + \alpha w_n) \\ &= (\alpha v_1, \dots, \alpha v_n) + (\alpha w_1, \dots, \alpha w_n) \\ &= \alpha\mathbf{v} + \alpha\mathbf{w}. \end{aligned}$$

As usual subtraction is defined as $\mathbf{x} - \mathbf{y} \equiv \mathbf{x} + (-\mathbf{y})$.

11.3 Geometric Meaning Of Vector Addition In $\mathbb{R}^3$

It was explained earlier that an element of $\mathbb{R}^n$ is an n tuple of numbers and it was also shown that this can be used to determine a point in three dimensional space in the case where $n = 3$ and in two dimensional space, in the case where $n = 2$. This point was specified relative to some coordinate axes.

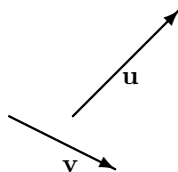
Consider the case where $n = 3$ for now. If you draw an arrow from the point in three dimensional space determined by $(0, 0, 0)$ to the point (a, b, c) with its tail sitting at the point $(0, 0, 0)$ and its point at the point (a, b, c) , this arrow is called the **position vector** of the point determined by $\mathbf{u} \equiv (a, b, c)$. One way to get to this point is to start at $(0, 0, 0)$ and move in the direction of the x_1 axis to $(a, 0, 0)$ and then in the direction of the x_2 axis to $(a, b, 0)$ and finally in the direction of the x_3 axis to (a, b, c) . It is evident that the same arrow (vector) would result if you began at the point, $\mathbf{v} \equiv (d, e, f)$, moved in the direction of the x_1 axis to $(d + a, e, f)$, then in the direction of the x_2 axis to $(d + a, e + b, f)$, and finally in the x_3 direction to $(d + a, e + b, f + c)$ only this time, the arrow would have its tail sitting at the point determined by $\mathbf{v} \equiv (d, e, f)$ and its point at $(d + a, e + b, f + c)$. It is said to be the same arrow (vector) because it will point in the same direction and have the same length. It is like you took an actual arrow, the sort of thing you shoot with a bow, and moved it from one location to another keeping it pointing the same direction. This is illustrated in the following picture in which $\mathbf{v} + \mathbf{u}$ is illustrated. Note the parallelogram determined in the picture by the vectors $\mathbf{u}$ and $\mathbf{v}$.



Thus the geometric significance of $(d, e, f) + (a, b, c) = (d + a, e + b, f + c)$ is this. You start with the position vector of the point (d, e, f) and at its point, you place the vector determined by (a, b, c) with its tail at (d, e, f) . Then the point of this last vector will be $(d + a, e + b, f + c)$. This is the geometric significance of vector addition. Also, as shown in the picture, $\mathbf{u} + \mathbf{v}$ is the directed diagonal of the parallelogram determined by the two vectors $\mathbf{u}$ and $\mathbf{v}$.

The following example is art.

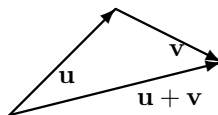
Exercise 11.3.1 Here is a picture of two vectors, $\mathbf{u}$ and $\mathbf{v}$.



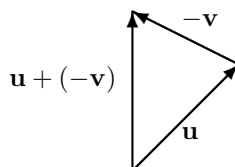
Sketch a picture of $\mathbf{u} + \mathbf{v}$, $\mathbf{u} - \mathbf{v}$, and $\mathbf{u} + 2\mathbf{v}$.

First here is a picture of $\mathbf{u} + \mathbf{v}$. You first draw $\mathbf{u}$ and then at the point of $\mathbf{u}$ you

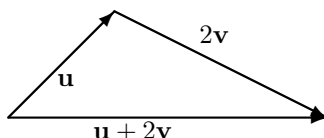
place the tail of $\mathbf{v}$ as shown. Then $\mathbf{u} + \mathbf{v}$ is the vector which results which is drawn in the following pretty picture.



Next consider $\mathbf{u} - \mathbf{v}$. This means $\mathbf{u} + (-\mathbf{v})$. From the above geometric description of vector addition, $-\mathbf{v}$ is the vector which has the same length but which points in the opposite direction to $\mathbf{v}$. Here is a picture.



Finally consider the vector $\mathbf{u} + 2\mathbf{v}$. Here is a picture of this one also.

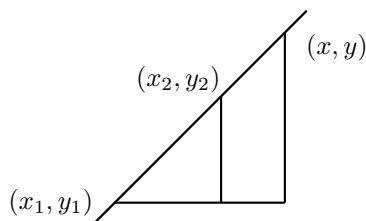


11.4 Lines

To begin with consider the case $n = 1, 2$. In the case where $n = 1$, the only line is just $\mathbb{R}^1 = \mathbb{R}$. Therefore, if x_1 and x_2 are two different points in $\mathbb{R}$, consider

$$x = x_1 + t(x_2 - x_1)$$

where $t \in \mathbb{R}$ and the totality of all such points will give $\mathbb{R}$. You see that you can always solve the above equation for t , showing that every point on $\mathbb{R}$ is of this form. Now consider the plane. Does a similar formula hold? Let (x_1, y_1) and (x_2, y_2) be two different points in $\mathbb{R}^2$ which are contained in a line, l . Suppose that $x_1 \neq x_2$. Then if (x, y) is an arbitrary point on l ,



Now by similar triangles,

$$m \equiv \frac{y_2 - y_1}{x_2 - x_1} = \frac{y - y_1}{x - x_1}$$

and so the point slope form of the line, l , is given as

$$y - y_1 = m(x - x_1).$$

If t is defined by

$$x = x_1 + t(x_2 - x_1),$$

you obtain this equation along with

$$\begin{aligned} y &= y_1 + mt(x_2 - x_1) \\ &= y_1 + t(y_2 - y_1). \end{aligned}$$

Therefore,

$$(x, y) = (x_1, y_1) + t(x_2 - x_1, y_2 - y_1).$$

If $x_1 = x_2$, then in place of the point slope form above, $x = x_1$. Since the two given points are different, $y_1 \neq y_2$ and so you still obtain the above formula for the line. Because of this, the following is the definition of a line in $\mathbb{R}^n$.

Definition 11.4.1 A line in $\mathbb{R}^n$ containing the two different points, $\mathbf{x}^1$ and $\mathbf{x}^2$ is the collection of points of the form

$$\mathbf{x} = \mathbf{x}^1 + t(\mathbf{x}^2 - \mathbf{x}^1)$$

where $t \in \mathbb{R}$. This is known as a **parametric equation** and the variable t is called the **parameter**.

Often t denotes time in applications to Physics. Note this definition agrees with the usual notion of a line in two dimensions and so this is consistent with earlier concepts.

Lemma 11.4.2 Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ with $\mathbf{a} \neq \mathbf{0}$. Then $\mathbf{x} = t\mathbf{a} + \mathbf{b}$, $t \in \mathbb{R}$, is a line.

Proof: Let $\mathbf{x}^1 = \mathbf{b}$ and let $\mathbf{x}^2 - \mathbf{x}^1 = \mathbf{a}$ so that $\mathbf{x}^2 \neq \mathbf{x}^1$. Then $t\mathbf{a} + \mathbf{b} = \mathbf{x}^1 + t(\mathbf{x}^2 - \mathbf{x}^1)$ and so $\mathbf{x} = t\mathbf{a} + \mathbf{b}$ is a line containing the two different points, $\mathbf{x}^1$ and $\mathbf{x}^2$. This proves the lemma.

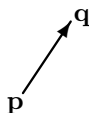
Definition 11.4.3 The vector $\mathbf{a}$ in the above lemma is called a **direction vector** for the line.

Definition 11.4.4 Let $\mathbf{p}$ and $\mathbf{q}$ be two points in $\mathbb{R}^n$, $\mathbf{p} \neq \mathbf{q}$. The **directed line segment** from $\mathbf{p}$ to $\mathbf{q}$, denoted by $\overrightarrow{\mathbf{pq}}$, is defined to be the collection of points,

$$\mathbf{x} = \mathbf{p} + t(\mathbf{q} - \mathbf{p}), \quad t \in [0, 1]$$

with the direction corresponding to increasing t . In the definition, when $t = 0$, the point $\mathbf{p}$ is obtained and as t increases other points on this line segment are obtained until when $t = 1$, you get the point, $\mathbf{q}$. This is what is meant by saying the direction corresponds to increasing t .

Think of $\overrightarrow{\mathbf{pq}}$ as an arrow whose point is on $\mathbf{q}$ and whose base is at $\mathbf{p}$ as shown in the following picture.



This line segment is a part of a line from the above Definition.

Example 11.4.5 Find a parametric equation for the line through the points $(1, 2, 0)$ and $(2, -4, 6)$.

Use the definition of a line given above to write

$$(x, y, z) = (1, 2, 0) + t(1, -6, 6), \quad t \in \mathbb{R}.$$

The vector $(1, -6, 6)$ is obtained by $(2, -4, 6) - (1, 2, 0)$ as indicated above.

The reason for the word, “a”, rather than the word, “the” is there are infinitely many different parametric equations for the same line. To see this replace t with $3s$. Then you obtain a parametric equation for the same line because the same set of points is obtained. The difference is they are obtained from different values of the parameter. What happens is this: The line is a set of points but the parametric description gives more information than that. It tells how the set of points are obtained. Obviously, there are many ways to trace out a given set of points and each of these ways corresponds to a different parametric equation for the line.

Example 11.4.6 Find a parametric equation for the line which contains the point $(1, 2, 0)$ and has direction vector, $(1, 2, 1)$.

From the above this is just

$$(x, y, z) = (1, 2, 0) + t(1, 2, 1), \quad t \in \mathbb{R}. \quad (11.11)$$

Sometimes people elect to write a line like the above in the form

$$x = 1 + t, \quad y = 2 + 2t, \quad z = t, \quad t \in \mathbb{R}. \quad (11.12)$$

This is a set of scalar parametric equations which amounts to the same thing as 11.11.

There is one other form for a line which is sometimes considered useful. It is the so called symmetric form. Consider the line of 11.12. You can solve for the parameter, t to write

$$t = x - 1, \quad t = \frac{y - 2}{2}, \quad t = z.$$

Therefore,

$$x - 1 = \frac{y - 2}{2} = z.$$

This is the symmetric form of the line.

Example 11.4.7 Suppose the *symmetric form of a line* is

$$\frac{x - 2}{3} = \frac{y - 1}{2} = z + 3.$$

Find the line in parametric form.

Let $t = \frac{x-2}{3}$, $t = \frac{y-1}{2}$ and $t = z + 3$. Then solving for x, y, z , you get

$$x = 3t + 2, \quad y = 2t + 1, \quad z = t - 3, \quad t \in \mathbb{R}.$$

Written in terms of vectors this is

$$(2, 1, -3) + t(3, 2, 1) = (x, y, z), \quad t \in \mathbb{R}.$$

11.5 Distance in $\mathbb{R}^n$

How is distance between two points in $\mathbb{R}^n$ defined?

Definition 11.5.1 Let $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ be two points in $\mathbb{R}^n$. Then $|\mathbf{x} - \mathbf{y}|$ indicates the distance between these points and is defined as

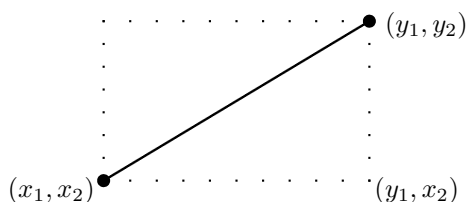
$$\text{distance between } \mathbf{x} \text{ and } \mathbf{y} \equiv |\mathbf{x} - \mathbf{y}| \equiv \left(\sum_{k=1}^n |x_k - y_k|^2 \right)^{1/2}.$$

This is called the **distance formula**. Thus $|\mathbf{x}| \equiv |\mathbf{x} - \mathbf{0}|$. The symbol, $B(\mathbf{a}, r)$ is defined by

$$B(\mathbf{a}, r) \equiv \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{a}| < r\}.$$

This is called an **open ball** of radius r centered at $\mathbf{a}$. It gives all the points in $\mathbb{R}^n$ which are closer to $\mathbf{a}$ than r .

First of all note this is a generalization of the notion of distance in $\mathbb{R}$. There the distance between two points, x and y was given by the absolute value of their difference. Thus $|x - y|$ is equal to the distance between these two points on $\mathbb{R}$. Now $|x - y| = \left((x - y)^2 \right)^{1/2}$ where the square root is always the positive square root. Thus it is the same formula as the above definition except there is only one term in the sum. Geometrically, this is the right way to define distance which is seen from the Pythagorean theorem. Consider the following picture in the case that $n = 2$.

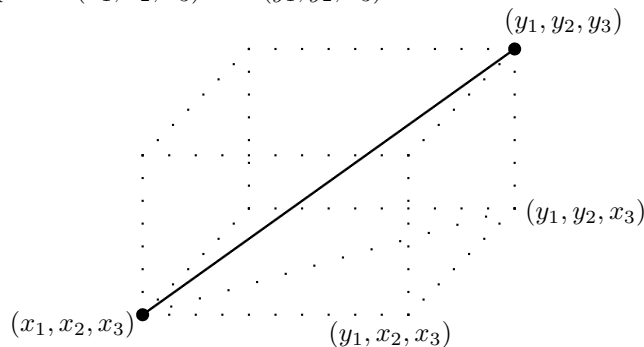


There are two points in the plane whose Cartesian coordinates are (x_1, x_2) and (y_1, y_2) respectively. Then the solid line joining these two points is the hypotenuse of a right triangle which is half of the rectangle shown in dotted lines. What is its length? Note the lengths of the sides of this triangle are $|y_1 - x_1|$ and $|y_2 - x_2|$. Therefore, the Pythagorean theorem implies the length of the hypotenuse equals

$$\left(|y_1 - x_1|^2 + |y_2 - x_2|^2 \right)^{1/2} = \left((y_1 - x_1)^2 + (y_2 - x_2)^2 \right)^{1/2}$$

which is just the formula for the distance given above.

Now suppose $n = 3$ and let (x_1, x_2, x_3) and (y_1, y_2, y_3) be two points in $\mathbb{R}^3$. Consider the following picture in which one of the solid lines joins the two points and a dotted line joins the points (x_1, x_2, x_3) and (y_1, y_2, x_3) .



By the Pythagorean theorem, the length of the dotted line joining (x_1, x_2, x_3) and (y_1, y_2, x_3) equals

$$\left((y_1 - x_1)^2 + (y_2 - x_2)^2\right)^{1/2}$$

while the length of the line joining (y_1, y_2, x_3) to (y_1, y_2, y_3) is just $|y_3 - x_3|$. Therefore, by the Pythagorean theorem again, the length of the line joining the points (x_1, x_2, x_3) and (y_1, y_2, y_3) equals

$$\begin{aligned} & \left\{ \left[\left((y_1 - x_1)^2 + (y_2 - x_2)^2 \right)^{1/2} \right]^2 + (y_3 - x_3)^2 \right\}^{1/2} \\ &= \left((y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2 \right)^{1/2}, \end{aligned}$$

which is again just the distance formula above.

This completes the argument that the above definition is reasonable. Of course you cannot continue drawing pictures in ever higher dimensions but there is no problem with the formula for distance in any number of dimensions. Here is an example.

Example 11.5.2 Find the distance between the points in $\mathbb{R}^4$,

$$\mathbf{a} = (1, 2, -4, 6), \quad \mathbf{b} = (2, 3, -1, 0)$$

Use the distance formula and write

$$|\mathbf{a} - \mathbf{b}|^2 = (1 - 2)^2 + (2 - 3)^2 + (-4 - (-1))^2 + (6 - 0)^2 = 47$$

Therefore, $|\mathbf{a} - \mathbf{b}| = \sqrt{47}$.

All this amounts to defining the distance between two points as the length of a straight line joining these two points. However, there is nothing sacred about using straight lines. One could define the distance to be the length of some other sort of line joining these points. It won't be done in this book but sometimes this sort of thing is done.

Another convention which is usually followed, especially in $\mathbb{R}^2$ and $\mathbb{R}^3$ is to denote the first component of a point in $\mathbb{R}^2$ by x and the second component by y . In $\mathbb{R}^3$ it is customary to denote the first and second components as just described while the third component is called z .

Example 11.5.3 Describe the points which are at the same distance between $(1, 2, 3)$ and $(0, 1, 2)$.

Let (x, y, z) be such a point. Then

$$\sqrt{(x-1)^2 + (y-2)^2 + (z-3)^2} = \sqrt{x^2 + (y-1)^2 + (z-2)^2}.$$

Squaring both sides

$$(x-1)^2 + (y-2)^2 + (z-3)^2 = x^2 + (y-1)^2 + (z-2)^2$$

and so

$$x^2 - 2x + 14 + y^2 - 4y + z^2 - 6z = x^2 + y^2 - 2y + 5 + z^2 - 4z$$

which implies

$$-2x + 14 - 4y - 6z = -2y + 5 - 4z$$

and so

$$2x + 2y + 2z = -9. \tag{11.13}$$

Since these steps are reversible, the set of points which is at the same distance from the two given points consists of the points, (x, y, z) such that 11.13 holds.

The following lemma is fundamental. It is a form of the Cauchy Schwarz inequality.

Lemma 11.5.4 *Let $\mathbf{x} = (x_1, \dots, x_n)$ and $\mathbf{y} = (y_1, \dots, y_n)$ be two points in $\mathbb{R}^n$. Then*

$$\left| \sum_{i=1}^n x_i y_i \right| \leq |\mathbf{x}| |\mathbf{y}|. \quad (11.14)$$

Proof: Let θ be either 1 or -1 such that

$$\theta \sum_{i=1}^n x_i y_i = \sum_{i=1}^n x_i (\theta y_i) = \left| \sum_{i=1}^n x_i y_i \right|$$

and consider $p(t) \equiv \sum_{i=1}^n (x_i + t\theta y_i)^2$. Then for all $t \in \mathbb{R}$,

$$\begin{aligned} 0 &\leq p(t) = \sum_{i=1}^n x_i^2 + 2t \sum_{i=1}^n x_i \theta y_i + t^2 \sum_{i=1}^n y_i^2 \\ &= |\mathbf{x}|^2 + 2t \sum_{i=1}^n x_i \theta y_i + t^2 |\mathbf{y}|^2 \end{aligned}$$

If $|\mathbf{y}| = 0$ then 11.14 is obviously true because both sides equal zero. Therefore, assume $|\mathbf{y}| \neq 0$ and then $p(t)$ is a polynomial of degree two whose graph opens up. Therefore, it either has no zeroes, two zeroes or one repeated zero. If it has two zeroes, the above inequality must be violated because in this case the graph must dip below the x axis. Therefore, it either has no zeroes or exactly one. From the quadratic formula this happens exactly when

$$4 \left(\sum_{i=1}^n x_i \theta y_i \right)^2 - 4 |\mathbf{x}|^2 |\mathbf{y}|^2 \leq 0$$

and so

$$\sum_{i=1}^n x_i \theta y_i = \left| \sum_{i=1}^n x_i y_i \right| \leq |\mathbf{x}| |\mathbf{y}|$$

as claimed. This proves the inequality.

There are certain properties of the distance which are obvious. Two of them which follow directly from the definition are

$$|\mathbf{x} - \mathbf{y}| = |\mathbf{y} - \mathbf{x}|,$$

$$|\mathbf{x} - \mathbf{y}| \geq 0 \text{ and equals } 0 \text{ only if } \mathbf{y} = \mathbf{x}.$$

The third fundamental property of distance is known as the triangle inequality. Recall that in any triangle the sum of the lengths of two sides is always at least as large as the third side. The following corollary is equivalent to this simple statement.

Corollary 11.5.5 *Let $\mathbf{x}, \mathbf{y}$ be points of $\mathbb{R}^n$. Then*

$$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|.$$

Proof: Using the Cauchy Schwarz inequality, Lemma 11.5.4,

$$\begin{aligned} |\mathbf{x} + \mathbf{y}|^2 &\equiv \sum_{i=1}^n (x_i + y_i)^2 \\ &= \sum_{i=1}^n x_i^2 + 2 \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2 \\ &\leq |\mathbf{x}|^2 + 2 |\mathbf{x}| |\mathbf{y}| + |\mathbf{y}|^2 \\ &= (|\mathbf{x}| + |\mathbf{y}|)^2 \end{aligned}$$

and so upon taking square roots of both sides,

$$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$$

and this proves the corollary.

11.6 Geometric Meaning Of Scalar Multiplication In $\mathbb{R}^3$

As discussed earlier, $\mathbf{x} = (x_1, x_2, x_3)$ determines a vector. You draw the line from $\mathbf{0}$ to $\mathbf{x}$ placing the point of the vector on $\mathbf{x}$. What is the length of this vector? The length of this vector is defined to equal $|\mathbf{x}|$ as in Definition 11.5.1. Thus the length of $\mathbf{x}$ equals $\sqrt{x_1^2 + x_2^2 + x_3^2}$. When you multiply $\mathbf{x}$ by a scalar, α , you get $(\alpha x_1, \alpha x_2, \alpha x_3)$ and the length of this vector is defined as $\sqrt{((\alpha x_1)^2 + (\alpha x_2)^2 + (\alpha x_3)^2)} = |\alpha| \sqrt{x_1^2 + x_2^2 + x_3^2}$. Thus the following holds.

$$|\alpha \mathbf{x}| = |\alpha| |\mathbf{x}|.$$

In other words, multiplication by a scalar magnifies the length of the vector. What about the direction? You should convince yourself by drawing a picture that if α is negative, it causes the resulting vector to point in the opposite direction while if $\alpha > 0$ it preserves the direction the vector points. One way to see this is to first observe that if $\alpha \neq 1$, then $\mathbf{x}$ and $\alpha \mathbf{x}$ are both points on the same line.

11.7 Exercises

- Verify all the properties 11.3-11.10.
- Compute the following
 - $5(1, 2, 3, -2) + 6(2, 1, -2, 7)$
 - $5(1, 2, -2) - 6(2, 1, -2)$
 - $-3(1, 0, 3, -2) + (2, 0, -2, 1)$
 - $-3(1, -2, -3, -2) - 2(2, -1, -2, 7)$
 - $-(2, 2, -3, -2) + 2(2, 4, -2, 7)$
- Find symmetric equations for the line through the points $(2, 2, 4)$ and $(-2, 3, 1)$.
- Find symmetric equations for the line through the points $(1, 2, 4)$ and $(-2, 1, 1)$.
- Symmetric equations for a line are given. Find parametric equations of the line.
 - $\frac{x+1}{3} = \frac{2y+3}{2} = z + 7$
 - $\frac{2x-1}{3} = \frac{2y+3}{6} = z - 7$
 - $\frac{x+1}{3} = 2y + 3 = 2z - 1$
 - $\frac{1-2x}{3} = \frac{3-2y}{2} = z + 1$
 - $\frac{x-1}{3} = \frac{2y-3}{5} = z + 2$
 - $\frac{x+1}{3} = \frac{3-y}{5} = z + 1$
- Parametric equations for a line are given. Find symmetric equations for the line if possible. If it is not possible to do it explain why.
 - $x = 1 + 2t, y = 3 - t, z = 5 + 3t$

- (b) $x = 1 + t, y = 3 - t, z = 5 - 3t$
 (c) $x = 1 + 2t, y = 3 + t, z = 5 + 3t$
 (d) $x = 1 - 2t, y = 1, z = 1 + t$
 (e) $x = 1 - t, y = 3 + 2t, z = 5 - 3t$
 (f) $x = t, y = 3 - t, z = 1 + t$
7. The first point given is a point contained in the line. The second point given is a direction vector for the line. Find parametric equations for the line determined by this information.
- (a) $(1, 2, 1), (2, 0, 3)$
 (b) $(1, 0, 1), (1, 1, 3)$
 (c) $(1, 2, 0), (1, 1, 0)$
 (d) $(1, 0, -6), (-2, -1, 3)$
 (e) $(-1, -2, -1), (2, 1, -1)$
 (f) $(0, 0, 0), (2, -3, 1)$
8. Parametric equations for a line are given. Determine a direction vector for this line.
- (a) $x = 1 + 2t, y = 3 - t, z = 5 + 3t$
 (b) $x = 1 + t, y = 3 + 3t, z = 5 - t$
 (c) $x = 7 + t, y = 3 + 4t, z = 5 - 3t$
 (d) $x = 2t, y = -3t, z = 3t$
 (e) $x = 2t, y = 3 + 2t, z = 5 + t$
 (f) $x = t, y = 3 + 3t, z = 5 + t$
9. A line contains the given two points. Find parametric equations for this line. Identify the direction vector.
- (a) $(0, 1, 0), (2, 1, 2)$
 (b) $(0, 1, 1), (2, 5, 0)$
 (c) $(1, 1, 0), (0, 1, 2)$
 (d) $(0, 1, 3), (0, 3, 0)$
 (e) $(0, 1, 0), (0, 6, 2)$
 (f) $(0, 1, 2), (2, 0, 2)$
10. Draw a picture of the points in $\mathbb{R}^2$ which are determined by the following ordered pairs.
- (a) $(1, 2)$
 (b) $(-2, -2)$
 (c) $(-2, 3)$
 (d) $(2, -5)$
11. Does it make sense to write $(1, 2) + (2, 3, 1)$? Explain.
12. Draw a picture of the points in $\mathbb{R}^3$ which are determined by the following ordered triples.
- (a) $(1, 2, 0)$
 (b) $(-2, -2, 1)$
 (c) $(-2, 3, -2)$
13. You are given two points in $\mathbb{R}^3, (4, 5, -4)$ and $(2, 3, 0)$. Show the distance from the point, $(3, 4, -2)$ to the first of these points is the same as the distance from this point to the second of the original pair of points. Note that $3 = \frac{4+2}{2}, 4 = \frac{5+3}{2}$. Obtain a theorem which will be valid for general pairs of points, (x, y, z) and (x_1, y_1, z_1) and prove your theorem using the distance formula.
14. A sphere is the set of all points which are at a given distance from a single given point. Find an equation for the sphere which is the set of all points that are at a distance of 4 from the point $(1, 2, 3)$ in $\mathbb{R}^3$.
15. A parabola is the set of all points (x, y) in the plane such that the distance from the point (x, y) to a given point, (x_0, y_0) equals the distance from (x, y) to a given line. The point, (x_0, y_0) is called the **focus** and the line is called the **directrix**. Find the equation of the parabola which results from the line $y = l$ and (x_0, y_0) a given focus with $y_0 < l$. Repeat for $y_0 > l$.
16. A sphere centered at the point $(x_0, y_0, z_0) \in \mathbb{R}^3$ having radius r consists of all points, (x, y, z) whose distance to (x_0, y_0, z_0) equals r . Write an equation for this sphere in $\mathbb{R}^3$.
17. Suppose the distance between (x, y) and (x', y') were defined to equal the larger of the two numbers $|x - x'|$

and $|y - y'|$. Draw a picture of the sphere centered at the point, $(0, 0)$ if this notion of distance is used.

18. Repeat the same problem except this time let the distance between the two points be $|x - x'| + |y - y'|$.
19. If (x_1, y_1, z_1) and (x_2, y_2, z_2) are two points such that $|(x_i, y_i, z_i)| = 1$ for $i = 1, 2$, show that in terms of the usual distance, $\left|\left(\frac{x_1+x_2}{2}, \frac{y_1+y_2}{2}, \frac{z_1+z_2}{2}\right)\right| < 1$. What would happen if you used the way of measuring distance given in Problem 17 ($|(x, y, z)| = \text{maximum of } |z|, |x|, |y|$)?
20. Give a simple description using the distance formula of the set of points which are at an equal distance between the two points (x_1, y_1, z_1) and (x_2, y_2, z_2) .
21. Suppose you are given two points, $(-a, 0)$ and $(a, 0)$ in $\mathbb{R}^2$ and a number, $r > 2a$. The set of points described by

$$\{(x, y) \in \mathbb{R}^2 : |(x, y) - (-a, 0)| + |(x, y) - (a, 0)| = r\}$$

is known as an ellipse. The two given points are known as the **focus points** of the ellipse. Simplify this to the form $\left(\frac{x}{\alpha}\right)^2 + \left(\frac{y}{\beta}\right)^2 = 1$. This is a nice exercise in messy algebra.

22. Suppose you are given two points, $(-a, 0)$ and $(a, 0)$ in $\mathbb{R}^2$ and a number, $r < 2a$. The set of points described by

$$\{(x, y) \in \mathbb{R}^2 : |(x, y) - (-a, 0)| - |(x, y) - (a, 0)| = r\}$$

is known as **hyperbola**. The two given points are known as the **focus points** of the hyperbola. Simplify this to the form $\left(\frac{x}{\alpha}\right)^2 - \left(\frac{y}{\beta}\right)^2 = 1$. This is a nice exercise in messy algebra.

23. Let (x_1, y_1) and (x_2, y_2) be two points in $\mathbb{R}^2$. Give a simple description using the distance formula of the perpendicular bisector of the line segment joining these two points. Thus you want all points, (x, y) such that $|(x, y) - (x_1, y_1)| = |(x, y) - (x_2, y_2)|$.

Matrices And Linear Transformations

12.0.1 Outcomes

1. Perform the basic matrix operations of matrix addition, scalar multiplication, transposition and matrix multiplication. Identify when these operations are not defined. Represent the basic operations in terms of double subscript notation.
2. Recall and prove algebraic properties for matrix addition, scalar multiplication, transposition, and matrix multiplication. Apply these properties to manipulate an algebraic expression involving matrices.
3. Evaluate the inverse of a matrix using Gauss Jordan elimination.
4. Recall the cancellation laws for matrix multiplication. Demonstrate when cancellation laws do not apply.
5. Recall and prove identities involving matrix inverses.
6. Understand the relationship between linear transformations and matrices.

12.1 Matrix Arithmetic

12.1.1 Addition And Scalar Multiplication Of Matrices

Numbers are also called **scalars**. In this book, scalars will be real numbers or complex numbers.

A **matrix** is a rectangular array of numbers. Several of them are referred to as **matrices**. For example, here is a matrix.

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{pmatrix}$$

The size or dimension of a matrix is defined as $m \times n$ where m is the number of rows and n is the number of columns. The above matrix is a 3×4 matrix because there are three rows and four columns. The first row is $(1 \ 2 \ 3 \ 4)$, the second row is $(5 \ 2 \ 8 \ 7)$ and

so forth. The first column is $\begin{pmatrix} 1 \\ 5 \\ 6 \end{pmatrix}$. When specifying the size of a matrix, you always

list the number of rows before the number of columns. Also, you can remember the columns are like columns in a Greek temple. They stand upright while the rows just lay

there like rows made by a tractor in a plowed field. Entries of the matrix are identified according to position in the matrix. For example, 8 is in position 2, 3 because it is in the second row and the third column. You might remember that you always list the rows before the columns by using the phrase **Row**man **C**atholic. The symbol, (a_{ij}) refers to a matrix. The entry in the i^{th} row and the j^{th} column of this matrix is denoted by a_{ij} . Using this notation on the above matrix, $a_{23} = 8$, $a_{32} = -9$, $a_{12} = 2$, etc.

There are various operations which are done on matrices. Matrices can be added, multiplied by a scalar, and multiplied by other matrices. To illustrate scalar multiplication, consider the following example in which a matrix is being multiplied by the scalar, 3.

$$3 \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 6 & 9 & 12 \\ 15 & 6 & 24 & 21 \\ 18 & -27 & 3 & 6 \end{pmatrix}.$$

The new matrix is obtained by multiplying every entry of the original matrix by the given scalar. If A is an $m \times n$ matrix, $-A$ equals $(-1)A$ because $A + (-1)A = 0$.

Two matrices must be the same size to be added. The sum of two matrices is a matrix which is obtained by adding the corresponding entries. Thus

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 2 \end{pmatrix} + \begin{pmatrix} -1 & 4 \\ 2 & 8 \\ 6 & -4 \end{pmatrix} = \begin{pmatrix} 0 & 6 \\ 5 & 12 \\ 11 & -2 \end{pmatrix}.$$

Two matrices are equal exactly when they are the same size and the corresponding entries are identical. Thus

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

because they are different sizes. As noted above, you write (c_{ij}) for the matrix C whose ij^{th} entry is c_{ij} . In doing arithmetic with matrices you must define what happens in terms of the c_{ij} sometimes called the **entries** of the matrix or the **components** of the matrix.

The above discussion stated for general matrices is given in the following definition.

Definition 12.1.1 (Scalar Multiplication) If $A = (a_{ij})$ and k is a scalar, then $kA = (ka_{ij})$.

Example 12.1.2 $7 \begin{pmatrix} 2 & 0 \\ 1 & -4 \end{pmatrix} = \begin{pmatrix} 14 & 0 \\ 7 & -28 \end{pmatrix}.$

Definition 12.1.3 (Addition) If $A = (a_{ij})$ and $B = (b_{ij})$ are two $m \times n$ matrices. Then $A + B = C$ where

$$C = (c_{ij})$$

for $c_{ij} = a_{ij} + b_{ij}$.

Example 12.1.4

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 2 & 3 \\ -6 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 4 & 6 \\ -5 & 2 & 5 \end{pmatrix}$$

To save on notation, we will often use A_{ij} to refer to the ij^{th} entry of the matrix, A .

Definition 12.1.5 (The zero matrix) The $m \times n$ zero matrix is the $m \times n$ matrix having every entry equal to zero. It is denoted by 0 .

Example 12.1.6 The 2×3 zero matrix is $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Note there are 2×3 zero matrices, 3×4 zero matrices, etc. In fact there is a zero matrix for every size.

Definition 12.1.7 (Equality of matrices) Let A and B be two matrices. Then $A = B$ means that the two matrices are of the same size and for $A = (a_{ij})$ and $B = (b_{ij})$, $a_{ij} = b_{ij}$ for all $1 \leq i \leq m$ and $1 \leq j \leq n$.

The following properties of matrices can be easily verified. You should do so.

- Commutative Law Of Addition.

$$A + B = B + A, \quad (12.1)$$

- Associative Law for Addition.

$$(A + B) + C = A + (B + C), \quad (12.2)$$

- Existence of an Additive Identity

$$A + 0 = A, \quad (12.3)$$

- Existence of an Additive Inverse

$$A + (-A) = 0, \quad (12.4)$$

Also for α, β scalars, the following additional properties hold.

- Distributive law over Matrix Addition.

$$\alpha(A + B) = \alpha A + \alpha B, \quad (12.5)$$

- Distributive law over Scalar Addition

$$(\alpha + \beta)A = \alpha A + \beta A, \quad (12.6)$$

- Associative law for Scalar Multiplication

$$\alpha(\beta A) = \alpha\beta(A), \quad (12.7)$$

- Rule for Multiplication by 1.

$$1A = A. \quad (12.8)$$

As an example, consider the Commutative Law of Addition. Let $A + B = C$ and $B + A = D$. Why is $D = C$?

$$C_{ij} = A_{ij} + B_{ij} = B_{ij} + A_{ij} = D_{ij}.$$

Therefore, $C = D$ because the ij^{th} entries are the same. Note that the conclusion follows from the commutative law of addition of numbers.

12.1.2 Multiplication Of Matrices

Definition 12.1.8 *Matrices which are $n \times 1$ or $1 \times n$ are called **vectors** and are often denoted by a bold letter. Thus the $n \times 1$ matrix*

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

*is also called a **column vector**. The $1 \times n$ matrix*

$$(x_1 \cdots x_n)$$

*is called a **row vector**.*

Although the following description of matrix multiplication may seem strange, it is in fact the most important and useful of the matrix operations. To begin with consider the case where a matrix is multiplied by a column vector. We will illustrate the general definition by first considering a special case.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = ?$$

This equals

$$7 \begin{pmatrix} 1 \\ 4 \end{pmatrix} + 8 \begin{pmatrix} 2 \\ 5 \end{pmatrix} + 9 \begin{pmatrix} 3 \\ 6 \end{pmatrix} = \begin{pmatrix} 50 \\ 122 \end{pmatrix}$$

In more general terms,

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} + x_3 \begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix} = \begin{pmatrix} x_1 a_{11} + x_2 a_{12} + x_3 a_{13} \\ x_1 a_{21} + x_2 a_{22} + x_3 a_{23} \end{pmatrix}$$

Thus you take x_1 times the first column, add to x_2 times the second column, and finally x_3 times the third column. In general, here is the definition of how to multiply an $(m \times n)$ matrix times a $(n \times 1)$ matrix.

Definition 12.1.9 *Let $A = (A_{ij})$ be an $m \times n$ matrix,*

$$A = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{pmatrix}$$

where the i^{th} column of A is denoted by $\mathbf{a}_i$, and let $\mathbf{v}$ be an $n \times 1$ matrix,

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

Then $A\mathbf{v}$ is an $m \times 1$ matrix equal to

$$v_1 \mathbf{a}_1 + v_2 \mathbf{a}_2 + \cdots + v_n \mathbf{a}_n = \sum_{k=1}^n v_k \mathbf{a}_k$$

It follows from the observation that the j^{th} column of A is

$$\begin{pmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{mj} \end{pmatrix},$$

that the above sum is of the form

$$v_1 \begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \end{pmatrix} + v_2 \begin{pmatrix} A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \end{pmatrix} + \cdots + v_k \begin{pmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{mn} \end{pmatrix}$$

so the i^{th} entry of the $m \times 1$ matrix or column vector which results is

$$\sum_{j=1}^n A_{ij} v_j$$

Here is another example.

Example 12.1.10 Compute

$$\begin{pmatrix} 1 & 2 & 1 & 3 \\ 0 & 2 & 1 & -2 \\ 2 & 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}.$$

First of all this is of the form $(3 \times 4)(4 \times 1)$ and so the result should be a (3×1) . Note how the inside numbers cancel. Then this equals

$$1 \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + 2 \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix} + 0 \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} + 1 \begin{pmatrix} 3 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \\ 5 \end{pmatrix}$$

The next task is to multiply an $m \times n$ matrix times an $n \times p$ matrix. Before doing so, the following may be helpful.

For A and B matrices, in order to form the product, AB the number of columns of A must equal the number of rows of B .

$$\begin{matrix} & \text{these must match!} \\ (m \times & n) \end{matrix} \begin{matrix} \text{ } \\ \widehat{ } \\ (n \times p) \end{matrix} = m \times p$$

Note the two outside numbers give the size of the product. Remember:

If the two middle numbers don't match,

you can't multiply the matrices!

Definition 12.1.11 When the number of columns of A equals the number of rows of B the two matrices are said to be **conformable** and the product, AB is obtained as follows. Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix. Then B is of the form

$$B = (\mathbf{b}_1 \quad \cdots \quad \mathbf{b}_p)$$

where $\mathbf{b}_k$ is an $n \times 1$ matrix or column vector. Then the $m \times p$ matrix, AB is defined as follows:

$$AB \equiv (\mathbf{Ab}_1 \quad \cdots \quad \mathbf{Ab}_p) \quad (12.9)$$

where $\mathbf{Ab}_k$ is an $m \times 1$ matrix or column vector which gives the k^{th} column of AB .

Example 12.1.12 Multiply the following.

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix}$$

The first thing you need to check before doing anything else is whether it is possible to do the multiplication. The first matrix is a 2×3 and the second matrix is a 3×3 . Therefore, is it possible to multiply these matrices. According to the above discussion it should be a 2×3 matrix of the form

$$\left(\overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix}}^{\text{First column}} \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}, \overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix}}^{\text{Second column}} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix}}^{\text{Third column}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right)$$

You know how to multiply a matrix times a vector and so you do so to obtain each of the three columns. Thus

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 9 & 3 \\ -2 & 7 & 3 \end{pmatrix}.$$

Example 12.1.13 Multiply the following.

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix}$$

First check if it is possible. This is of the form $(3 \times 3)(2 \times 3)$. The inside numbers do not match and so you can't do this multiplication. This means that anything you write will be absolute nonsense because it is impossible to multiply these matrices in this order. Aren't they the same two matrices considered in the previous example? Yes they are. It is just that here they are in a different order. This shows something you must always remember about matrix multiplication.

Order Matters!

Matrix Multiplication Is Not Commutative!

This is very different than multiplication of numbers!

12.1.3 The ij^{th} Entry Of A Product

It is important to describe matrix multiplication in terms of entries of the matrices. What is the ij^{th} entry of AB ? It would be the i^{th} entry of the j^{th} column of AB . Thus it would be the i^{th} entry of $\mathbf{Ab}_j$. Now

$$\mathbf{b}_j = \begin{pmatrix} B_{1j} \\ \vdots \\ B_{nj} \end{pmatrix}$$

and from the above definition, the i^{th} entry is

$$\sum_{k=1}^n A_{ik} B_{kj}. \quad (12.10)$$

In terms of pictures of the matrix, you are doing

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \vdots & \vdots & & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{np} \end{pmatrix}$$

Then as explained above, the j^{th} column is of the form

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{nj} \end{pmatrix}$$

which is a $m \times 1$ matrix or column vector which equals

$$\begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \end{pmatrix} B_{1j} + \begin{pmatrix} A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \end{pmatrix} B_{2j} + \cdots + \begin{pmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{mn} \end{pmatrix} B_{nj}.$$

The second entry of this $m \times 1$ matrix is

$$A_{21}B_{1j} + A_{22}B_{2j} + \cdots + A_{2n}B_{nj} = \sum_{k=1}^n A_{2k}B_{kj}.$$

Similarly, the i^{th} entry of this $m \times 1$ matrix is

$$A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj} = \sum_{k=1}^n A_{ik}B_{kj}.$$

This shows the following definition for matrix multiplication in terms of the ij^{th} entries of the product coincides with Definition 12.1.11.

Definition 12.1.14 Let $A = (A_{ij})$ be an $m \times n$ matrix and let $B = (B_{ij})$ be an $n \times p$ matrix. Then AB is an $m \times p$ matrix and

$$(AB)_{ij} = \sum_{k=1}^n A_{ik}B_{kj}. \quad (12.11)$$

Example 12.1.15 Multiply if possible $\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \end{pmatrix}$.

First check to see if this is possible. It is of the form $(3 \times 2)(2 \times 3)$ and since the inside numbers match, the two matrices are conformable and it is possible to do the multiplication. The result should be a 3×3 matrix. The answer is of the form

$$\left(\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 \\ 7 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right)$$

where the commas separate the columns in the resulting product. Thus the above product equals

$$\begin{pmatrix} 16 & 15 & 5 \\ 13 & 15 & 5 \\ 46 & 42 & 14 \end{pmatrix},$$

a 3×3 matrix as desired. In terms of the ij^{th} entries and the above definition, the entry in the third row and second column of the product should equal

$$\begin{aligned} \sum_j a_{3j}b_{j2} &= a_{31}b_{12} + a_{32}b_{22} \\ &= 2 \times 3 + 6 \times 6 = 42. \end{aligned}$$

You should try a few more such examples to verify the above definition in terms of the ij^{th} entries works for other entries.

Example 12.1.16 *Multiply if possible* $\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \\ 0 & 0 & 0 \end{pmatrix}$.

This is not possible because it is of the form $(3 \times 2)(3 \times 3)$ and the middle numbers don't match. In other words the two matrices are not conformable in the indicated order.

Example 12.1.17 *Multiply if possible* $\begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix}$.

This is possible because in this case it is of the form $(3 \times 3)(3 \times 2)$ and the middle numbers do match so the matrices are conformable. When the multiplication is done it equals

$$\begin{pmatrix} 13 & 13 \\ 29 & 32 \\ 0 & 0 \end{pmatrix}.$$

Check this and be sure you come up with the same answer.

Example 12.1.18 *Multiply if possible* $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} (1 \ 2 \ 1 \ 0)$.

In this case you are trying to do $(3 \times 1)(1 \times 4)$. The inside numbers match so you can do it. Verify

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} (1 \ 2 \ 1 \ 0) = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 2 & 4 & 2 & 0 \\ 1 & 2 & 1 & 0 \end{pmatrix}$$

12.1.4 Properties Of Matrix Multiplication

As pointed out above, sometimes it is possible to multiply matrices in one order but not in the other order. What if it makes sense to multiply them in either order? Will the two products be equal then?

Example 12.1.19 *Compare* $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ *and* $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

The first product is

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}.$$

The second product is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}.$$

You see these are not equal. Again you cannot conclude that $AB = BA$ for matrix multiplication even when multiplication is defined in both orders. However, there are some properties which do hold.

Proposition 12.1.20 *If all multiplications and additions make sense, the following hold for matrices, A, B, C and a, b scalars.*

$$A(aB + bC) = a(AB) + b(AC) \quad (12.12)$$

$$(B + C)A = BA + CA \quad (12.13)$$

$$A(BC) = (AB)C \quad (12.14)$$

Proof: Using Definition 12.1.14,

$$\begin{aligned} (A(aB + bC))_{ij} &= \sum_k A_{ik}(aB + bC)_{kj} \\ &= \sum_k A_{ik}(aB_{kj} + bC_{kj}) \\ &= a \sum_k A_{ik}B_{kj} + b \sum_k A_{ik}C_{kj} \\ &= a(AB)_{ij} + b(AC)_{ij} \\ &= (a(AB) + b(AC))_{ij}. \end{aligned}$$

Thus $A(B + C) = AB + AC$ as claimed. Formula 12.13 is entirely similar.

Formula 12.14 is the associative law of multiplication. Using Definition 12.1.14,

$$\begin{aligned} (A(BC))_{ij} &= \sum_k A_{ik}(BC)_{kj} \\ &= \sum_k A_{ik} \sum_l B_{kl}C_{lj} \\ &= \sum_l (AB)_{il}C_{lj} \\ &= ((AB)C)_{ij}. \end{aligned}$$

This proves 12.14.

12.1.5 The Transpose

Another important operation on matrices is that of taking the **transpose**. The following example shows what is meant by this operation, denoted by placing a T as an exponent on the matrix.

$$\begin{pmatrix} 1 & 4 \\ 3 & 1 \\ 2 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 & 2 \\ 4 & 1 & 6 \end{pmatrix}$$

What happened? The first column became the first row and the second column became the second row. Thus the 3×2 matrix became a 2×3 matrix. The number 3 was in the second row and the first column and it ended up in the first row and second column. Here is the definition.

Definition 12.1.21 Let A be an $m \times n$ matrix. Then A^T denotes the $n \times m$ matrix which is defined as follows.

$$(A^T)_{ij} = A_{ji}$$

Example 12.1.22

$$\begin{pmatrix} 1 & 2 & -6 \\ 3 & 5 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 5 \\ -6 & 4 \end{pmatrix}.$$

The transpose of a matrix has the following important properties.

Lemma 12.1.23 Let A be an $m \times n$ matrix and let B be a $n \times p$ matrix. Then

$$(AB)^T = B^T A^T \quad (12.15)$$

and if α and β are scalars,

$$(\alpha A + \beta B)^T = \alpha A^T + \beta B^T \quad (12.16)$$

Proof: From the definition,

$$\begin{aligned} ((AB)^T)_{ij} &= (AB)_{ji} \\ &= \sum_k A_{jk} B_{ki} \\ &= \sum_k (B^T)_{ik} (A^T)_{kj} \\ &= (B^T A^T)_{ij} \end{aligned}$$

The proof of Formula 12.16 is left as an exercise and this proves the lemma.

Definition 12.1.24 An $n \times n$ matrix, A is said to be **symmetric** if $A = A^T$. It is said to be **skew symmetric** if $A = -A^T$.

Example 12.1.25 Let

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 5 & -3 \\ 3 & -3 & 7 \end{pmatrix}.$$

Then A is symmetric.

Example 12.1.26 Let

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & 0 & 2 \\ -3 & -2 & 0 \end{pmatrix}$$

Then A is skew symmetric.

12.1.6 The Identity And Inverses

There is a special matrix called I and referred to as the identity matrix. It is always a square matrix, meaning the number of rows equals the number of columns and it has the property that there are ones down the main diagonal and zeroes elsewhere. Here are some identity matrices of various sizes.

$$(1), \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The first is the 1×1 identity matrix, the second is the 2×2 identity matrix, the third is the 3×3 identity matrix, and the fourth is the 4×4 identity matrix. By extension, you can likely see what the $n \times n$ identity matrix would be. It is so important that there is a special symbol to denote the ij^{th} entry of the identity matrix

$$I_{ij} = \delta_{ij}$$

where δ_{ij} is the **Kronecker symbol** defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

It is called the **identity matrix** because it is a **multiplicative identity** in the following sense.

Lemma 12.1.27 *Suppose A is an $m \times n$ matrix and I_n is the $n \times n$ identity matrix. Then $AI_n = A$. If I_m is the $m \times m$ identity matrix, it also follows that $I_mA = A$.*

Proof:

$$\begin{aligned} (AI_n)_{ij} &= \sum_k A_{ik} \delta_{kj} \\ &= A_{ij} \end{aligned}$$

and so $AI_n = A$. The other case is left as an exercise for you.

Definition 12.1.28 *An $n \times n$ matrix, A has an **inverse**, A^{-1} if and only if $AA^{-1} = A^{-1}A = I$. Such a matrix is called **invertible**.*

It is very important to observe that the inverse of a matrix, if it exists, is unique. Another way to think of this is that if it acts like the inverse, then it is the inverse.

Theorem 12.1.29 *Suppose A^{-1} exists and $AB = BA = I$. Then $B = A^{-1}$.*

Proof:

$$A^{-1} = A^{-1}I = A^{-1}(AB) = (A^{-1}A)B = IB = B.$$

Unlike ordinary multiplication of numbers, it can happen that $A \neq 0$ but A may fail to have an inverse. This is illustrated in the following example.

Example 12.1.30 *Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Does A have an inverse?*

One might think A would have an inverse because it does not equal zero. However,

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and if A^{-1} existed, this could not happen because you could write

$$\begin{aligned} \begin{pmatrix} 0 \\ 0 \end{pmatrix} &= A^{-1} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \right) = A^{-1} \left(A \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right) = \\ &= (A^{-1}A) \begin{pmatrix} -1 \\ 1 \end{pmatrix} = I \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \end{aligned}$$

a contradiction. Thus the answer is that A does not have an inverse.

Example 12.1.31 Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Show $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ is the inverse of A .

To check this, multiply

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

showing that this matrix is indeed the inverse of A .

There are various ways of finding the inverse of a matrix. One way will be presented in the discussion on determinants. You can also find them directly from the definition provided they exist.

In the last example, how would you find A^{-1} ? You wish to find a matrix, $\begin{pmatrix} x & z \\ y & w \end{pmatrix}$ such that

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x & z \\ y & w \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This requires the solution of the systems of equations,

$$x + y = 1, x + 2y = 0$$

and

$$z + w = 0, z + 2w = 1.$$

The first pair of equations has the solution $y = -1$ and $x = 2$. The second pair of equations has the solution $w = 1$, $z = -1$. Therefore, from the definition of the inverse,

$$A^{-1} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

To be sure it is the inverse, you should multiply on both sides of the original matrix. It turns out that if it works on one side, it will always work on the other. The consideration of this as well as a more detailed treatment of inverses is a good topic for a linear algebra course. However, see the appendix on determinants which discusses this topic from the point of view of determinants.

12.2 Linear Transformations

An $m \times n$ matrix can be used to transform vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$ through the use of matrix multiplication.

Example 12.2.1 Consider the matrix, $\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix}$. Think of it as a function which takes vectors in $\mathbb{R}^3$ and makes them into vectors in $\mathbb{R}^2$ as follows. For $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ a vector in $\mathbb{R}^3$, multiply on the left by the given matrix to obtain the vector in $\mathbb{R}^2$. Here are some numerical examples.

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} -3 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ 5 \\ -3 \end{pmatrix} = \begin{pmatrix} 20 \\ 25 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 7 \\ 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 7 \end{pmatrix},$$

More generally,

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + 2y \\ 2x + y \end{pmatrix}$$

The idea is to define a function which takes vectors in $\mathbb{R}^3$ and delivers new vectors in $\mathbb{R}^2$.

This is an example of something called a linear transformation.

Definition 12.2.2 Let X and Y be vector spaces and let $T : X \rightarrow Y$ be a function. Thus for each $\mathbf{x} \in X$, $T\mathbf{x} \in Y$. Then T is a **linear transformation** if whenever α, β are scalars and $\mathbf{x}_1$ and $\mathbf{x}_2$ are vectors in X ,

$$T(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2) = \alpha T\mathbf{x}_1 + \beta T\mathbf{x}_2.$$

In words, linear transformations distribute across $+$ and allow you to factor out scalars. At this point, recall the properties of matrix multiplication. The pertinent property is 12.13 on Page 315. Recall it states that for a and b scalars,

$$A(aB + bC) = aAB + bAC$$

In particular, for A an $m \times n$ matrix and B and C , $n \times 1$ matrices (column vectors) the above formula holds which is nothing more than the statement that matrix multiplication gives an example of a linear transformation.

Definition 12.2.3 A linear transformation is called **one to one** (often written as $1 - 1$) if it never takes two different vectors to the same vector. Thus T is one to one if whenever $\mathbf{x} \neq \mathbf{y}$

$$T\mathbf{x} \neq T\mathbf{y}.$$

Equivalently, if $T(\mathbf{x}) = T(\mathbf{y})$, then $\mathbf{x} = \mathbf{y}$.

In the case that a linear transformation comes from matrix multiplication, it is common usage to refer to the matrix as a one to one matrix when the linear transformation it determines is one to one.

Definition 12.2.4 A linear transformation mapping X to Y is called **onto** if whenever $\mathbf{y} \in Y$ there exists $\mathbf{x} \in X$ such that $T(\mathbf{x}) = \mathbf{y}$.

Thus T is onto if everything in Y gets hit. In the case that a linear transformation comes from matrix multiplication, it is common to refer to the matrix as onto when the linear transformation it determines is onto. Also it is common usage to write TX , $T(X)$, or $\text{Im}(T)$ as the set of vectors of Y which are of the form $T\mathbf{x}$ for some $\mathbf{x} \in X$. In the case that T is obtained from multiplication by an $m \times n$ matrix, A , it is standard to simply write $A(\mathbb{R}^n)$ or $A\mathbb{R}^n$, or $\text{Im}(A)$ to denote those vectors in $\mathbb{R}^m$ which are obtained in the form $A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$.

12.3 Constructing The Matrix Of A Linear Transformation

It turns out that if T is any linear transformation which maps $\mathbb{R}^n$ to $\mathbb{R}^m$, there is always an $m \times n$ matrix, A with the property that

$$A\mathbf{x} = T\mathbf{x} \quad (12.17)$$

for all $\mathbf{x} \in \mathbb{R}^n$. Here is why. Suppose $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear transformation and you want to find the matrix defined by this linear transformation as described in 12.17. Then if $\mathbf{x} \in \mathbb{R}^n$ it follows

$$\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

where as implied above, $\mathbf{e}_i$ is the vector which has zeros in every slot but the i^{th} and a 1 in this slot. Then since T is linear,

$$\begin{aligned} T\mathbf{x} &= \sum_{i=1}^n x_i T(\mathbf{e}_i) \\ &= \begin{pmatrix} | & & | \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ &\equiv A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \end{aligned}$$

and so you see that the matrix desired is obtained from letting the i^{th} column equal $T(\mathbf{e}_i)$. This yields the following theorem.

Theorem 12.3.1 Let T be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then the matrix, A satisfying 12.17 is given by

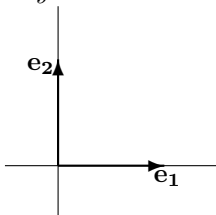
$$\begin{pmatrix} | & & | \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ | & & | \end{pmatrix}$$

where $T\mathbf{e}_i$ is the i^{th} column of A .

Sometimes you need to find a matrix which represents a given linear transformation which is described in geometrical terms. The idea is to produce a matrix which you can multiply a vector by to get the same thing as some geometrical description. A good example of this is the problem of rotation of vectors.

Example 12.3.2 Determine the matrix which represents the linear transformation defined by rotating every vector through an angle of θ .

Let $\mathbf{e}_1 \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{e}_2 \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. These identify the geometric vectors which point along the positive x axis and positive y axis as shown.



From the above, you only need to find $T\mathbf{e}_1$ and $T\mathbf{e}_2$, the first being the first column of the desired matrix, A and the second being the second column. From drawing a picture and doing a little geometry, you see that

$$T\mathbf{e}_1 = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}, T\mathbf{e}_2 = \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}.$$

Therefore, from Theorem 12.3.1,

$$A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

Example 12.3.3 Find the matrix of the linear transformation which is obtained by first rotating all vectors through an angle of ϕ and then through an angle θ . Thus you want the linear transformation which rotates all angles through an angle of $\theta + \phi$.

Let $T_{\theta+\phi}$ denote the linear transformation which rotates every vector through an angle of $\theta + \phi$. Then to get $T_{\theta+\phi}$, you could first do T_ϕ and then do T_θ where T_ϕ is the linear transformation which rotates through an angle of ϕ and T_θ is the linear transformation which rotates through an angle of θ . Denoting the corresponding matrices by $A_{\theta+\phi}$, A_ϕ , and A_θ , you must have for every $\mathbf{x}$

$$A_{\theta+\phi}\mathbf{x} = T_{\theta+\phi}\mathbf{x} = T_\theta T_\phi \mathbf{x} = A_\theta A_\phi \mathbf{x}.$$

Consequently, you must have

$$\begin{aligned} A_{\theta+\phi} &= \begin{pmatrix} \cos(\theta + \phi) & -\sin(\theta + \phi) \\ \sin(\theta + \phi) & \cos(\theta + \phi) \end{pmatrix} = A_\theta A_\phi \\ &= \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}. \end{aligned}$$

You know how to multiply matrices. Do so to the pair on the right. This yields

$$\begin{aligned} &\begin{pmatrix} \cos(\theta + \phi) & -\sin(\theta + \phi) \\ \sin(\theta + \phi) & \cos(\theta + \phi) \end{pmatrix} \\ &= \begin{pmatrix} \cos \theta \cos \phi - \sin \theta \sin \phi & -\cos \theta \sin \phi - \sin \theta \cos \phi \\ \sin \theta \cos \phi + \cos \theta \sin \phi & \cos \theta \cos \phi - \sin \theta \sin \phi \end{pmatrix}. \end{aligned}$$

Don't these look familiar? They are the usual trig. identities for the sum of two angles derived here using linear algebra concepts.

You do not have to stop with two dimensions. You can consider rotations and other geometric concepts in any number of dimensions. This is one of the major advantages of linear algebra. You can break down a difficult geometrical procedure into small steps, each corresponding to multiplication by an appropriate matrix. Then by multiplying the matrices, you can obtain a single matrix which can give you numerical information on the results of applying the given sequence of simple procedures. That which you could never visualize can still be understood to the extent of finding exact numerical answers. The following is a more routine example quite typical of what will be important in the calculus of several variables.

Example 12.3.4 Let $T(x_1, x_2) = \begin{pmatrix} x_1 + 3x_2 \\ x_1 - x_2 \\ x_1 \\ 3x_2 + 5x_1 \end{pmatrix}$. Thus $T: \mathbb{R}^2 \rightarrow \mathbb{R}^4$. Explain why

T is a linear transformation and write $T(x_1, x_2)$ in the form $A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ where A is an appropriate matrix.

From the definition of matrix multiplication,

$$T(x_1, x_2) = \begin{pmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 0 \\ 5 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Since $T\mathbf{x}$ is of the form $A\mathbf{x}$ for A a matrix, it follows T is a linear transformation. You could also verify directly that $T(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha T(\mathbf{x}) + \beta T(\mathbf{y})$.

12.4 Exercises

1. Here are some matrices:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 7 \end{pmatrix},$$

$$B = \begin{pmatrix} 3 & -1 & 2 \\ -3 & 2 & 1 \end{pmatrix},$$

$$C = \begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix},$$

$$D = \begin{pmatrix} -1 & 2 \\ 2 & -3 \end{pmatrix}, E = \begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Find if possible $-3A, 3B - A, AC, CB, AE, EA$. If it is not possible explain why.

2. Here are some matrices:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{pmatrix},$$

$$B = \begin{pmatrix} 2 & -5 & 2 \\ -3 & 2 & 1 \end{pmatrix},$$

$$C = \begin{pmatrix} 1 & 2 \\ 5 & 0 \end{pmatrix},$$

$$D = \begin{pmatrix} -1 & 1 \\ 4 & -3 \end{pmatrix}, E = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

Find if possible $-3A, 3B - A, AC, CA, AE, EA, BE, DE$. If it is not possible explain why.

3. Here are some matrices:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \\ 1 & -1 \end{pmatrix},$$

$$B = \begin{pmatrix} 2 & -5 & 2 \\ -3 & 2 & 1 \end{pmatrix},$$

$$C = \begin{pmatrix} 1 & 2 \\ 5 & 0 \end{pmatrix},$$

$$D = \begin{pmatrix} -1 & 1 \\ 4 & -3 \end{pmatrix}, E = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

Find if possible $-3A^T, 3B - A^T, AC, CA, AE, E^T B, BE, DE$,

EE^T, E^TE . If it is not possible explain why.

4. Let $A = \begin{pmatrix} 1 & 1 \\ -2 & -1 \\ 1 & 2 \end{pmatrix}$,
 $B = \begin{pmatrix} 1 & -1 & -2 \\ 2 & 1 & -2 \end{pmatrix}$, and
 $C = \begin{pmatrix} 1 & 1 & -3 \\ -1 & 2 & 0 \\ -3 & -1 & 0 \end{pmatrix}$. Find if possible.

- (a) AB
 (b) BA
 (c) AC
 (d) CA
 (e) CB
 (f) BC

5. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 2 \\ 3 & k \end{pmatrix}$.
 Is it possible to choose k such that $AB = BA$? If so, what should k equal?

6. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 1 & 2 \\ 1 & k \end{pmatrix}$.
 Is it possible to choose k such that $AB = BA$? If so, what should k equal?

7. Let $\mathbf{x} = (-1, -1, 1)$ and $\mathbf{y} = (0, 1, 2)$.
 Find $\mathbf{x}^T\mathbf{y}$ and $\mathbf{xy}^T$ if possible.

8. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/4$.

9. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $-\pi/3$.

10. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $2\pi/3$.

11. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $\pi/12$.
Hint: Note that $\pi/12 = \pi/3 - \pi/4$.

12. Let $T(x_1, x_2) = \begin{pmatrix} x_1 + 4x_2 \\ x_2 + 2x_1 \end{pmatrix}$.
 Thus $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. Explain why T is a linear transformation and write $T(x_1, x_2)$ in the form $A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ where A is an appropriate matrix.

13. Let $T(x_1, x_2) = \begin{pmatrix} x_1 - x_2 \\ x_1 \\ 3x_2 + x_1 \\ 3x_2 + 5x_1 \end{pmatrix}$.
 Thus $T : \mathbb{R}^2 \rightarrow \mathbb{R}^4$. Explain why T is a linear transformation and write $T(x_1, x_2)$ in the form $A \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ where A is an appropriate matrix.

14. Let $T(x_1, x_2, x_3, x_4) = \begin{pmatrix} x_1 - x_2 + 2x_3 \\ 2x_3 + x_1 \\ 3x_3 \\ 3x_4 + 3x_2 + x_1 \end{pmatrix}$. Thus
 $T : \mathbb{R}^4 \rightarrow \mathbb{R}^4$. Explain why T is a linear transformation and write $T(x_1, x_2, x_3, x_4)$ in the form $A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ where A is an appropriate matrix.

15. Let $T(x_1, x_2) = \begin{pmatrix} x_1^2 + 4x_2 \\ x_2 + 2x_1 \end{pmatrix}$.
 Thus $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$. Explain why T cannot possibly be a linear transformation.

16. Suppose A and B are square matrices of the same size. Which of the following are correct?

- (a) $(A - B)^2 = A^2 - 2AB + B^2$
 (b) $(AB)^2 = A^2B^2$
 (c) $(A + B)^2 = A^2 + 2AB + B^2$
 (d) $(A + B)^2 = A^2 + AB + BA + B^2$
 (e) $A^2B^2 = A(AB)B$
 (f) $(A + B)^3 = A^3 + 3A^2B + 3AB^2 + B^3$
 (g) $(A + B)(A - B) = A^2 - B^2$

17. Let $A = \begin{pmatrix} -1 & -1 \\ 3 & 3 \end{pmatrix}$. Find $\boxed{\text{all}}$ 2×2 matrices, B such that $AB = 0$.

18. In 12.1 - 12.8 describe $-A$ and 0 .
19. Let A be an $n \times n$ matrix. Show A equals the sum of a symmetric and a skew symmetric matrix. **Hint:** Consider the matrix $\frac{1}{2}(A + A^T)$. Is this matrix symmetric?
20. If A is a skew symmetric matrix, what can be concluded about A^n where $n = 1, 2, 3, \dots$?
21. Show every skew symmetric matrix has all zeros down the main diagonal. The main diagonal consists of every entry of the matrix which is of the form a_{ii} . It runs from the upper left down to the lower right.
22. Using only the properties 12.1 - 12.8 show $-A$ is unique.
23. Using only the properties 12.1 - 12.8 show 0 is unique.
24. Using only the properties 12.1 - 12.8 show $0A = 0$. Here the 0 on the left is the scalar 0 and the 0 on the right is the zero for $m \times n$ matrices.
25. Using only the properties 12.1 - 12.8 and previous problems show $(-1)A = -A$.
26. Prove 12.16.
27. Prove that $I_m A = A$ where A is an $m \times n$ matrix.
28. Let

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$
 Find A^{-1} if possible. If A^{-1} does not exist, determine why.
29. Let

$$A = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}.$$
 Find A^{-1} if possible. If A^{-1} does not exist, determine why.
30. Let

$$A = \begin{pmatrix} 3 & 2 \\ 2 & 1 \end{pmatrix}.$$
 Find A^{-1} if possible. If A^{-1} does not exist, determine why.
31. Give an example of matrices, A, B, C such that $B \neq C$, $A \neq 0$, and yet $AB = AC$.
32. Suppose $AB = AC$ and A is an invertible $n \times n$ matrix. Does it follow that $B = C$? Explain why or why not. What if A were a non invertible $n \times n$ matrix?
33. Find your own examples:
 - (a) 2×2 matrices, A and B such that $A \neq 0, B \neq 0$ with $AB \neq BA$.
 - (b) 2×2 matrices, A and B such that $A \neq 0, B \neq 0$, but $AB = 0$.
 - (c) 2×2 matrices, A, D , and C such that $A \neq 0, C \neq D$, but $AC = AD$.
34. Explain why if $BA = CA$ and A^{-1} exists, then $B = C$.
35. Give an example of a matrix, A such that $A^2 = I$ and yet $A \neq I$ and $A \neq -I$.
36. Show that if A^{-1} exists for an $n \times n$ matrix, then it is unique. That is, if $BA = I$ and $AB = I$, then $B = A^{-1}$.
37. Show $(AB)^{-1} = B^{-1}A^{-1}$.
38. Show that if A is an invertible $n \times n$ matrix, then so is A^T and $(A^T)^{-1} = (A^{-1})^T$.
39. Show that if A is an $n \times n$ invertible matrix and $\mathbf{x}$ is a $n \times 1$ matrix such that $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b}$ an $n \times 1$ matrix, then $\mathbf{x} = A^{-1}\mathbf{b}$.
40. Prove that if A^{-1} exists and $A\mathbf{x} = \mathbf{0}$ then $\mathbf{x} = \mathbf{0}$.
41. Show that $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$ by verifying that

$$\begin{aligned} (ABC)(C^{-1}B^{-1}A^{-1}) \\ = (C^{-1}B^{-1}A^{-1})(ABC) = I. \end{aligned}$$

Determinants

13.0.1 Outcomes

1. Evaluate the determinant of a square matrix by applying
 - (a) the cofactor formula or
 - (b) row operations.
2. Recall the general properties of determinants.
3. Recall that the determinant of a product of matrices is the product of the determinants. Recall that the determinant of a matrix is equal to the determinant of its transpose.
4. Apply Cramer's Rule to solve a 2×2 or a 3×3 linear system.
5. Use determinants to determine whether a matrix has an inverse.
6. Evaluate the inverse of a matrix using cofactors.

13.1 Basic Techniques And Properties

13.1.1 Cofactors And 2×2 Determinants

Let A be an $n \times n$ matrix. The **determinant** of A , denoted as $\det(A)$ is a number. If the matrix is a 2×2 matrix, this number is very easy to find.

Definition 13.1.1 Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then

$$\det(A) \equiv ad - cb.$$

The determinant is also often denoted by enclosing the matrix with two vertical lines. Thus

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

Example 13.1.2 Find $\det \begin{pmatrix} 2 & 4 \\ -1 & 6 \end{pmatrix}$.

From the definition this is just $(2)(6) - (-1)(4) = 16$.

Having defined what is meant by the determinant of a 2×2 matrix, what about a 3×3 matrix?

Definition 13.1.3 Suppose A is a 3×3 matrix. The ij^{th} **minor**, denoted as $\text{minor}(A)_{ij}$, is the determinant of the 2×2 matrix which results from deleting the i^{th} row and the j^{th} column.

Example 13.1.4 Consider the matrix,

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

The $(1, 2)$ minor is the determinant of the 2×2 matrix which results when you delete the first row and the second column. This minor is therefore

$$\det \begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} = -2.$$

The $(2, 3)$ minor is the determinant of the 2×2 matrix which results when you delete the second row and the third column. This minor is therefore

$$\det \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = -4.$$

Definition 13.1.5 Suppose A is a 3×3 matrix. The ij^{th} **cofactor** is defined to be $(-1)^{i+j} \times (ij^{\text{th}} \text{ minor})$. In words, you multiply $(-1)^{i+j}$ times the ij^{th} minor to get the ij^{th} cofactor. The cofactors of a matrix are so important that special notation is appropriate when referring to them. The ij^{th} cofactor of a matrix, A will be denoted by $\text{cof}(A)_{ij}$. It is also convenient to refer to the cofactor of an entry of a matrix as follows. For a_{ij} an entry of the matrix, its cofactor is just $\text{cof}(A)_{ij}$. Thus the cofactor of the ij^{th} entry is just the ij^{th} cofactor.

Example 13.1.6 Consider the matrix,

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

The $(1, 2)$ minor is the determinant of the 2×2 matrix which results when you delete the first row and the second column. This minor is therefore

$$\det \begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} = -2.$$

It follows

$$\text{cof}(A)_{12} = (-1)^{1+2} \det \begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} = (-1)^{1+2} (-2) = 2$$

The $(2, 3)$ minor is the determinant of the 2×2 matrix which results when you delete the second row and the third column. This minor is therefore

$$\det \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = -4.$$

Therefore,

$$\text{cof}(A)_{23} = (-1)^{2+3} \det \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = (-1)^{2+3} (-4) = 4.$$

Similarly,

$$\text{cof}(A)_{22} = (-1)^{2+2} \det \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} = -8.$$

Definition 13.1.7 The determinant of a 3×3 matrix, A , is obtained by picking a row (column) and taking the product of each entry in that row (column) with its cofactor and adding these terms. This process, when applied to the i^{th} row (column), is known as expanding the determinant along the i^{th} row (column).

Example 13.1.8 Find the determinant of

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

Here is how it is done by “expanding along the first column”.

$$\overbrace{1(-1)^{1+1} \begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix}}^{\text{cof}(A)_{11}} + \overbrace{4(-1)^{2+1} \begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix}}^{\text{cof}(A)_{21}} + \overbrace{3(-1)^{3+1} \begin{vmatrix} 2 & 3 \\ 3 & 2 \end{vmatrix}}^{\text{cof}(A)_{31}} = 0.$$

You see, we just followed the rule in the above definition. We took the 1 in the first column and multiplied it by its cofactor, the 4 in the first column and multiplied it by its cofactor, and the 3 in the first column and multiplied it by its cofactor. Then we added these numbers together.

You could also expand the determinant along the second row as follows.

$$\overbrace{4(-1)^{2+1} \begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix}}^{\text{cof}(A)_{21}} + \overbrace{3(-1)^{2+2} \begin{vmatrix} 1 & 3 \\ 3 & 1 \end{vmatrix}}^{\text{cof}(A)_{22}} + \overbrace{2(-1)^{2+3} \begin{vmatrix} 1 & 2 \\ 3 & 2 \end{vmatrix}}^{\text{cof}(A)_{23}} = 0.$$

Observe this gives the same number. You should try expanding along other rows and columns. If you don't make any mistakes, you will always get the same answer.

What about a 4×4 matrix? You know now how to find the determinant of a 3×3 matrix. The pattern is the same.

Definition 13.1.9 Suppose A is a 4×4 matrix. The ij^{th} **minor** is the determinant of the 3×3 matrix you obtain when you delete the i^{th} row and the j^{th} column. The ij^{th} **cofactor**, $\text{cof}(A)_{ij}$ is defined to be $(-1)^{i+j} \times (ij^{\text{th}} \text{ minor})$. In words, you multiply $(-1)^{i+j}$ times the ij^{th} minor to get the ij^{th} cofactor.

Definition 13.1.10 The determinant of a 4×4 matrix, A , is obtained by picking a row (column) and taking the product of each entry in that row (column) with its cofactor and adding these terms. This process when applied to the i^{th} row (column) is known as expanding the determinant along the i^{th} row (column).

Example 13.1.11 Find $\det(A)$ where

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 4 & 2 & 3 \\ 1 & 3 & 4 & 5 \\ 3 & 4 & 3 & 2 \end{pmatrix}$$

As in the case of a 3×3 matrix, you can expand this along any row or column. Lets pick the third column. $\det(A) =$

$$3(-1)^{1+3} \begin{vmatrix} 5 & 4 & 3 \\ 1 & 3 & 5 \\ 3 & 4 & 2 \end{vmatrix} + 2(-1)^{2+3} \begin{vmatrix} 1 & 2 & 4 \\ 1 & 3 & 5 \\ 3 & 4 & 2 \end{vmatrix} +$$

$$4(-1)^{3+3} \begin{vmatrix} 1 & 2 & 4 \\ 5 & 4 & 3 \\ 3 & 4 & 2 \end{vmatrix} + 3(-1)^{4+3} \begin{vmatrix} 1 & 2 & 4 \\ 5 & 4 & 3 \\ 1 & 3 & 5 \end{vmatrix}.$$

Now you know how to expand each of these 3×3 matrices along a row or a column. If you do so, you will get -12 assuming you make no mistakes. You could expand this matrix along any row or any column and assuming you make no mistakes, you will always get the same thing which is defined to be the determinant of the matrix, A . This method of evaluating a determinant by expanding along a row or a column is called the **method of Laplace expansion**.

Note that each of the four terms above involves three terms consisting of determinants of 2×2 matrices and each of these will need 2 terms. Therefore, there will be $4 \times 3 \times 2 = 24$ terms to evaluate in order to find the determinant using the method of Laplace expansion. Suppose now you have a 10×10 matrix and you follow the above pattern for evaluating determinants. By analogy to the above, there will be $10! = 3,628,800$ terms involved in the evaluation of such a determinant by Laplace expansion along a row or column. This is a lot of terms.

In addition to the difficulties just discussed, you should regard the above claim that you always get the same answer by picking any row or column with considerable skepticism. It is incredible and not at all obvious. However, it requires a little effort to establish it. This is done in the appendix on the theory of the determinant. The above examples motivate the following definitions, the second of which is incredible.

Definition 13.1.12 Let $A = (a_{ij})$ be an $n \times n$ matrix and suppose the determinant of a $(n-1) \times (n-1)$ matrix has been defined. Then a new matrix called the **cofactor matrix**, $\text{cof}(A)$ is defined by $\text{cof}(A) = (c_{ij})$ where to obtain c_{ij} delete the i^{th} row and the j^{th} column of A , take the determinant of the $(n-1) \times (n-1)$ matrix which results, (This is called the ij^{th} **minor** of A .) and then multiply this number by $(-1)^{i+j}$. Thus $(-1)^{i+j} \times (\text{the } ij^{\text{th}} \text{ minor})$ equals the ij^{th} cofactor. To make the formulas easier to remember, $\text{cof}(A)_{ij}$ will denote the ij^{th} entry of the cofactor matrix.

With this definition of the cofactor matrix, here is how to define the determinant of an $n \times n$ matrix.

Definition 13.1.13 Let A be an $n \times n$ matrix where $n \geq 2$ and suppose the determinant of an $(n-1) \times (n-1)$ has been defined. Then

$$\det(A) = \sum_{j=1}^n a_{ij} \text{cof}(A)_{ij} = \sum_{i=1}^n a_{ij} \text{cof}(A)_{ij}. \quad (13.1)$$

The first formula consists of expanding the determinant along the i^{th} row and the second expands the determinant along the j^{th} column. This is called the method of Laplace expansion.

Theorem 13.1.14 Expanding the $n \times n$ matrix along any row or column always gives the same answer so the above definition is a good definition.

13.1.2 The Determinant Of A Triangular Matrix

Notwithstanding the difficulties involved in using the method of Laplace expansion, certain types of matrices are very easy to deal with.

Definition 13.1.15 A matrix M , is upper triangular if $M_{ij} = 0$ whenever $i > j$. Thus such a matrix equals zero below the main diagonal, the entries of the form

M_{ii} , as shown.

$$\begin{pmatrix} * & * & \cdots & * \\ 0 & * & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & * \end{pmatrix}$$

A lower triangular matrix is defined similarly as a matrix for which all entries above the main diagonal are equal to zero.

You should verify the following using the above theorem on Laplace expansion.

Corollary 13.1.16 *Let M be an upper (lower) triangular matrix. Then $\det(M)$ is obtained by taking the product of the entries on the main diagonal.*

Example 13.1.17 *Let*

$$A = \begin{pmatrix} 1 & 2 & 3 & 77 \\ 0 & 2 & 6 & 7 \\ 0 & 0 & 3 & 33.7 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Find $\det(A)$.

From the above corollary, it suffices to take the product of the diagonal entries. Thus $\det(A) = 1 \times 2 \times 3 \times (-1) = -6$. Without using the corollary, you could expand along the first column. This gives

$$\begin{aligned} & 1 \begin{vmatrix} 2 & 6 & 7 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix} + 0(-1)^{2+1} \begin{vmatrix} 2 & 3 & 77 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix} + \\ & 0(-1)^{3+1} \begin{vmatrix} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 0 & -1 \end{vmatrix} + 0(-1)^{4+1} \begin{vmatrix} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 3 & 33.7 \end{vmatrix} \end{aligned}$$

and the only nonzero term in the expansion is

$$1 \begin{vmatrix} 2 & 6 & 7 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix}.$$

Now expand this along the first column to obtain

$$\begin{aligned} & 1 \times \left(2 \times \begin{vmatrix} 3 & 33.7 \\ 0 & -1 \end{vmatrix} + 0(-1)^{2+1} \begin{vmatrix} 6 & 7 \\ 0 & -1 \end{vmatrix} + 0(-1)^{3+1} \begin{vmatrix} 6 & 7 \\ 3 & 33.7 \end{vmatrix} \right) \\ & = 1 \times 2 \times \begin{vmatrix} 3 & 33.7 \\ 0 & -1 \end{vmatrix} \end{aligned}$$

Next expand this last determinant along the first column to obtain the above equals

$$1 \times 2 \times 3 \times (-1) = -6$$

which is just the product of the entries down the main diagonal of the original matrix.

13.1.3 Properties Of Determinants

There are many properties satisfied by determinants. Some of these properties have to do with row operations which are defined below.

Definition 13.1.18 *The **row operations** consist of the following*

1. *Switch two rows.*
2. *Multiply a row by a nonzero number.*
3. *Replace a row by itself added to a multiple of another row.*

Theorem 13.1.19 *Let A be an $n \times n$ matrix and let A_1 be a matrix which results from multiplying some row of A by a scalar, c . Then $c \det(A) = \det(A_1)$.*

Example 13.1.20 *Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $A_1 = \begin{pmatrix} 2 & 4 \\ 3 & 4 \end{pmatrix}$. $\det(A) = -2$, $\det(A_1) = -4$.*

Theorem 13.1.21 *Let A be an $n \times n$ matrix and let A_1 be a matrix which results from switching two rows of A . Then $\det(A) = -\det(A_1)$. Also, if one row of A is a multiple of another row of A , then $\det(A) = 0$.*

Example 13.1.22 *Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and let $A_1 = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}$. $\det A = -2$, $\det(A_1) = 2$.*

Theorem 13.1.23 *Let A be an $n \times n$ matrix and let A_1 be a matrix which results from applying row operation 3. That is you replace some row by a multiple of another row added to itself. Then $\det(A) = \det(A_1)$.*

Example 13.1.24 *Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and let $A_1 = \begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix}$. Thus the second row of A_1 is one times the first row added to the second row. $\det(A) = -2$ and $\det(A_1) = -2$.*

Theorem 13.1.25 *In Theorems 13.1.19 - 13.1.23 you can replace the word, "row" with the word "column".*

There are two other major properties of determinants which do not involve row operations.

Theorem 13.1.26 *Let A and B be two $n \times n$ matrices. Then*

$$\det(AB) = \det(A) \det(B).$$

Also,

$$\det(A) = \det(A^T).$$

Example 13.1.27 *Compare $\det(AB)$ and $\det(A) \det(B)$ for*

$$A = \begin{pmatrix} 1 & 2 \\ -3 & 2 \end{pmatrix}, B = \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}.$$

First

$$AB = \begin{pmatrix} 1 & 2 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 11 & 4 \\ -1 & -4 \end{pmatrix}$$

and so

$$\det(AB) = \det \begin{pmatrix} 11 & 4 \\ -1 & -4 \end{pmatrix} = -40.$$

Now

$$\det(A) = \det \begin{pmatrix} 1 & 2 \\ -3 & 2 \end{pmatrix} = 8$$

and

$$\det(B) = \det \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix} = -5.$$

Thus $\det(A) \det(B) = 8 \times (-5) = -40$.

13.1.4 Finding Determinants Using Row Operations

Theorems 13.1.23 - 13.1.25 can be used to find determinants using row operations. As pointed out above, the method of Laplace expansion will not be practical for any matrix of large size. Here is an example in which all the row operations are used.

Example 13.1.28 Find the determinant of the matrix,

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 1 & 2 & 3 \\ 4 & 5 & 4 & 3 \\ 2 & 2 & -4 & 5 \end{pmatrix}$$

Replace the second row by (-5) times the first row added to it. Then replace the third row by (-4) times the first row added to it. Finally, replace the fourth row by (-2) times the first row added to it. This yields the matrix,

$$B = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -9 & -13 & -17 \\ 0 & -3 & -8 & -13 \\ 0 & -2 & -10 & -3 \end{pmatrix}$$

and from Theorem 13.1.23, it has the same determinant as A . Now using other row operations, $\det(B) = \left(\frac{-1}{3}\right) \det(C)$ where

$$C = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 11 & 22 \\ 0 & -3 & -8 & -13 \\ 0 & 6 & 30 & 9 \end{pmatrix}.$$

The second row was replaced by (-3) times the third row added to the second row. By Theorem 13.1.23 this didn't change the value of the determinant. Then the last row was multiplied by (-3) . By Theorem 13.1.19 the resulting matrix has a determinant which is (-3) times the determinant of the unmultiplied matrix. Therefore, we multiplied by $-1/3$ to retain the correct value. Now replace the last row with 2 times the third added to it. This does not change the value of the determinant by Theorem 13.1.23. Finally

switch the third and second rows. This causes the determinant to be multiplied by (-1) . Thus $\det(C) = -\det(D)$ where

$$D = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -3 & -8 & -13 \\ 0 & 0 & 11 & 22 \\ 0 & 0 & 14 & -17 \end{pmatrix}$$

You could do more row operations or you could note that this can be easily expanded along the first column followed by expanding the 3×3 matrix which results along its first column. Thus

$$\det(D) = 1(-3) \begin{vmatrix} 11 & 22 \\ 14 & -17 \end{vmatrix} = 1485$$

and so $\det(C) = -1485$ and $\det(A) = \det(B) = \left(\frac{-1}{3}\right)(-1485) = 495$.

Example 13.1.29 Find the determinant of the matrix

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & -3 & 2 & 1 \\ 2 & 1 & 2 & 5 \\ 3 & -4 & 1 & 2 \end{pmatrix}$$

Replace the second row by (-1) times the first row added to it. Next take -2 times the first row and add to the third and finally take -3 times the first row and add to the last row. This yields

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -1 & -1 \\ 0 & -3 & -4 & 1 \\ 0 & -10 & -8 & -4 \end{pmatrix}.$$

By Theorem 13.1.23 this matrix has the same determinant as the original matrix. Remember you can work with the columns also. Take -5 times the last column and add to the second column. This yields

$$\begin{pmatrix} 1 & -8 & 3 & 2 \\ 0 & 0 & -1 & -1 \\ 0 & -8 & -4 & 1 \\ 0 & 10 & -8 & -4 \end{pmatrix}$$

By Theorem 13.1.25 this matrix has the same determinant as the original matrix. Now take (-1) times the third row and add to the top row. This gives.

$$\begin{pmatrix} 1 & 0 & 7 & 1 \\ 0 & 0 & -1 & -1 \\ 0 & -8 & -4 & 1 \\ 0 & 10 & -8 & -4 \end{pmatrix}$$

which by Theorem 13.1.23 has the same determinant as the original matrix. Let's expand it now along the first column. This yields the following for the determinant of the original matrix.

$$\det \begin{pmatrix} 0 & -1 & -1 \\ -8 & -4 & 1 \\ 10 & -8 & -4 \end{pmatrix}$$

which equals

$$8 \det \begin{pmatrix} -1 & -1 \\ -8 & -4 \end{pmatrix} + 10 \det \begin{pmatrix} -1 & -1 \\ -4 & 1 \end{pmatrix} = -82$$

Do not try to be fancy in using row operations. That is, stick mostly to the one which replaces a row or column with a multiple of another row or column added to it. Also note there is no way to check your answer other than working the problem more than one way. To be sure you have gotten it right you must do this.

13.2 Applications

13.2.1 A Formula For The Inverse

The definition of the determinant in terms of Laplace expansion along a row or column also provides a way to give a formula for the inverse of a matrix. Recall the definition of the inverse of a matrix in Definition 12.1.28 on Page 317. Also recall the definition of the cofactor matrix given in Definition 13.1.12 on Page 328. This cofactor matrix was just the matrix which results from replacing the ij^{th} entry of the matrix with the ij^{th} cofactor.

The following theorem says that to find the inverse, take the transpose of the cofactor matrix and divide by the determinant. The transpose of the cofactor matrix is called the **adjugate** or sometimes the **classical adjoint** of the matrix A . In other words, A^{-1} is equal to one divided by the determinant of A times the adjugate matrix of A . This is what the following theorem says with more precision.

Theorem 13.2.1 A^{-1} exists if and only if $\det(A) \neq 0$. If $\det(A) \neq 0$, then $A^{-1} = (a_{ij}^{-1})$ where

$$a_{ij}^{-1} = \det(A)^{-1} \operatorname{cof}(A)_{ji}$$

for $\operatorname{cof}(A)_{ij}$ the ij^{th} cofactor of A .

Example 13.2.2 Find the inverse of the matrix,

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 0 & 1 \\ 1 & 2 & 1 \end{pmatrix}$$

First find the determinant of this matrix. Using Theorems 13.1.23 - 13.1.25 on Page 330, the determinant of this matrix equals the determinant of the matrix,

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & -6 & -8 \\ 0 & 0 & -2 \end{pmatrix}$$

which equals 12. The cofactor matrix of A is

$$\begin{pmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{pmatrix}.$$

Each entry of A was replaced by its cofactor. Therefore, from the above theorem, the inverse of A should equal

$$\frac{1}{12} \begin{pmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{pmatrix}^T = \begin{pmatrix} -\frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}.$$

Does it work? You should check to see if it does. When the matrices are multiplied

$$\begin{pmatrix} -\frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 0 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so it is correct.

Example 13.2.3 Find the inverse of the matrix,

$$A = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{2} \\ -\frac{5}{6} & \frac{2}{3} & -\frac{1}{2} \end{pmatrix}$$

First find its determinant. This determinant is $\frac{1}{6}$. The inverse is therefore equal to

$$6 \begin{pmatrix} \begin{vmatrix} \frac{1}{3} & -\frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} -\frac{1}{6} & -\frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} -\frac{1}{6} & \frac{1}{3} \\ -\frac{5}{6} & \frac{2}{3} \end{vmatrix} \\ -\begin{vmatrix} 0 & \frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} \frac{1}{2} & 0 \\ -\frac{5}{6} & \frac{2}{3} \end{vmatrix} \\ \begin{vmatrix} 0 & \frac{1}{2} \\ \frac{1}{3} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{6} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} \frac{1}{2} & 0 \\ -\frac{1}{6} & \frac{1}{3} \end{vmatrix} \end{pmatrix}^T.$$

Expanding all the 2×2 determinants this yields

$$6 \begin{pmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{6} & -\frac{1}{3} \\ -\frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}^T = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & 1 \end{pmatrix}$$

Always check your work.

$$\begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{2} \\ -\frac{5}{6} & \frac{2}{3} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so it is correct. If the result of multiplying these matrices had been something other than the identity matrix, you would know there was an error. When this happens, you need to search for the mistake if you are interested in getting the right answer. A common mistake is to forget to take the transpose of the cofactor matrix.

Proof of Theorem 13.2.1: From the definition of the determinant in terms of expansion along a column, and letting $(a_{ir}) = A$, if $\det(A) \neq 0$,

$$\sum_{i=1}^n a_{ir} \operatorname{cof}(A)_{ir} \det(A)^{-1} = \det(A) \det(A)^{-1} = 1.$$

Now consider

$$\sum_{i=1}^n a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

when $k \neq r$. Replace the k^{th} column with the r^{th} column to obtain a matrix, B_k whose determinant equals zero by Theorem 13.1.21. However, expanding this matrix, B_k along the k^{th} column yields

$$0 = \det(B_k) \det(A)^{-1} = \sum_{i=1}^n a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

Summarizing,

$$\sum_{i=1}^n a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1} = \delta_{rk} \equiv \begin{cases} 1 & \text{if } r = k \\ 0 & \text{if } r \neq k \end{cases}.$$

Now

$$\sum_{i=1}^n a_{ir} \operatorname{cof}(A)_{ik} = \sum_{i=1}^n a_{ir} \operatorname{cof}(A)_{ki}^T$$

which is the kr^{th} entry of $\operatorname{cof}(A)^T A$. Therefore,

$$\frac{\operatorname{cof}(A)^T}{\det(A)} A = I. \quad (13.2)$$

Using the other formula in Definition 13.1.13, and similar reasoning,

$$\sum_{j=1}^n a_{rj} \operatorname{cof}(A)_{kj} \det(A)^{-1} = \delta_{rk}$$

Now

$$\sum_{j=1}^n a_{rj} \operatorname{cof}(A)_{kj} = \sum_{j=1}^n a_{rj} \operatorname{cof}(A)_{jk}^T$$

which is the rk^{th} entry of $A \operatorname{cof}(A)^T$. Therefore,

$$A \frac{\operatorname{cof}(A)^T}{\det(A)} = I, \quad (13.3)$$

and it follows from 13.2 and 13.3 that $A^{-1} = (a_{ij}^{-1})$, where

$$a_{ij}^{-1} = \operatorname{cof}(A)_{ji} \det(A)^{-1}.$$

In other words,

$$A^{-1} = \frac{\operatorname{cof}(A)^T}{\det(A)}.$$

Now suppose A^{-1} exists. Then by Theorem 13.1.26,

$$1 = \det(I) = \det(AA^{-1}) = \det(A) \det(A^{-1})$$

so $\det(A) \neq 0$. This proves the theorem.

This way of finding inverses is especially useful in the case where it is desired to find the inverse of a matrix whose entries are functions.

Example 13.2.4 Suppose

$$A(t) = \begin{pmatrix} e^t & 0 & 0 \\ 0 & \cos t & \sin t \\ 0 & -\sin t & \cos t \end{pmatrix}$$

Show that $A(t)^{-1}$ exists and then find it.

First note $\det(A(t)) = e^t \neq 0$ so $A(t)^{-1}$ exists. The cofactor matrix is

$$C(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{pmatrix}$$

and so the inverse is

$$\frac{1}{e^t} \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{pmatrix}^T = \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}.$$

13.2.2 Cramer's Rule

This formula for the inverse also implies a famous procedure known as **Cramer's rule**. Cramer's rule gives a formula for the solutions, $\mathbf{x}$, to a system of equations, $A\mathbf{x} = \mathbf{y}$ in the special case that A is a square matrix. Note this rule does not apply if you have a system of equations in which there is a different number of equations than variables.

In case you are solving a system of equations, $A\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$, it follows that if A^{-1} exists,

$$\mathbf{x} = (A^{-1}A)\mathbf{x} = A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{y}$$

thus solving the system. Now in the case that A^{-1} exists, there is a formula for A^{-1} given above. Using this formula,

$$x_i = \sum_{j=1}^n a_{ij}^{-1} y_j = \sum_{j=1}^n \frac{1}{\det(A)} \operatorname{cof}(A)_{ji} y_j.$$

By the formula for the expansion of a determinant along a column,

$$x_i = \frac{1}{\det(A)} \det \begin{pmatrix} * & \cdots & y_1 & \cdots & * \\ \vdots & & \vdots & & \vdots \\ * & \cdots & y_n & \cdots & * \end{pmatrix},$$

where here the i^{th} column of A is replaced with the column vector, $(y_1, \dots, y_n)^T$, and the determinant of this modified matrix is taken and divided by $\det(A)$. This formula is known as Cramer's rule.

Procedure 13.2.5 Suppose A is an $n \times n$ matrix and it is desired to solve the system $A\mathbf{x} = \mathbf{y}$, $\mathbf{y} = (y_1, \dots, y_n)^T$ for $\mathbf{x} = (x_1, \dots, x_n)^T$. Then Cramer's rule says

$$x_i = \frac{\det A_i}{\det A}$$

where A_i is obtained from A by replacing the i^{th} column of A with the column $(y_1, \dots, y_n)^T$.

Example 13.2.6 Find x, y if

$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

From Cramer's rule,

$$x = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 2 & 2 & 1 \\ 3 & -3 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = \frac{1}{2}$$

Now to find y, z

$$y = \frac{\begin{vmatrix} 1 & 1 & 1 \\ 3 & 2 & 1 \\ 2 & 3 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = -\frac{1}{7}, \quad z = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 2 \\ 2 & -3 & 3 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = \frac{11}{14}$$

You see the pattern. For large systems Cramer's rule is less than useful if you want to find an answer. This is because to use it you must evaluate determinants. However, you have no practical way to evaluate determinants for large matrices other than row operations and if you are using row operations, you might just as well use them to solve the system to begin with. It will be a lot less trouble. Nevertheless, there are situations in which Cramer's rule is useful.

Example 13.2.7 Solve for z if

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix}$$

You could do it by row operations but it might be easier in this case to use Cramer's rule because the matrix of coefficients does not consist of numbers but of functions. Thus

$$z = \frac{\begin{vmatrix} 1 & 0 & 1 \\ 0 & e^t \cos t & t \\ 0 & -e^t \sin t & t^2 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{vmatrix}} = t((\cos t)t + \sin t)e^{-t}.$$

You end up doing this sort of thing sometimes in ordinary differential equations in the method of variation of parameters.

13.3 Exercises

1. Find the determinants of the following matrices.

(a) $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 2 \\ 0 & 9 & 8 \end{pmatrix}$ (The answer is 31.)

(b) $\begin{pmatrix} 4 & 3 & 2 \\ 1 & 7 & 8 \\ 3 & -9 & 3 \end{pmatrix}$ (The answer is 375.)

(c) $\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 3 & 2 & 3 \\ 4 & 1 & 5 & 0 \\ 1 & 2 & 1 & 2 \end{pmatrix}$, (The answer is -2 .)

2. Find the following determinant by expanding along the first row and second column.

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix}$$

3. Find the following determinant by expanding along the first column and third row.

$$\begin{vmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{vmatrix}$$

4. Find the following determinant by expanding along the second row and first column.

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 2 & 1 & 1 \end{vmatrix}$$

5. Compute the determinant by cofactor expansion. Pick the easiest row or column to use.

$$\begin{vmatrix} 1 & 0 & 0 & 1 \\ 2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 2 \\ 2 & 1 & 3 & 1 \end{vmatrix}$$

6. Find the determinant using row operations.

$$\begin{vmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ -4 & 1 & 2 \end{vmatrix}$$

7. Find the determinant using row operations.

$$\begin{vmatrix} 2 & 1 & 3 \\ 2 & 4 & 2 \\ 1 & 4 & -5 \end{vmatrix}$$

8. Find the determinant using row operations.

$$\begin{vmatrix} 1 & 2 & 1 & 2 \\ 3 & 1 & -2 & 3 \\ -1 & 0 & 3 & 1 \\ 2 & 3 & 2 & -2 \end{vmatrix}$$

9. Find the determinant using row operations.

$$\begin{vmatrix} 1 & 4 & 1 & 2 \\ 3 & 2 & -2 & 3 \\ -1 & 0 & 3 & 3 \\ 2 & 1 & 2 & -2 \end{vmatrix}$$

10. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

11. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} c & d \\ a & b \end{pmatrix}$$

12. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a & b \\ a+c & b+d \end{pmatrix}$$

13. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a & b \\ 2c & 2d \end{pmatrix}$$

14. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} b & a \\ d & c \end{pmatrix}$$

15. Tell whether the statement is true or false.

- (a) If A is a 3×3 matrix with a zero determinant, then one column must be a multiple of some other column.
- (b) If any two columns of a square matrix are equal, then the determinant of the matrix equals zero.
- (c) For A and B two $n \times n$ matrices, $\det(A + B) = \det(A) + \det(B)$.
- (d) For A an $n \times n$ matrix, $\det(3A) = 3 \det(A)$
- (e) If A^{-1} exists then $\det(A^{-1}) = \det(A)^{-1}$.

- (f) If B is obtained by multiplying a single row of A by 4 then $\det(B) = 4 \det(A)$.
 - (g) For A an $n \times n$ matrix, $\det(-A) = (-1)^n \det(A)$.
 - (h) If A is a real $n \times n$ matrix, then $\det(A^T A) \geq 0$.
 - (i) Cramer's rule is useful for finding solutions to systems of linear equations in which there is an infinite set of solutions.
 - (j) If $A^k = 0$ for some positive integer, k , then $\det(A) = 0$.
 - (k) If $A\mathbf{x} = \mathbf{0}$ for some $\mathbf{x} \neq \mathbf{0}$, then $\det(A) = 0$.
16. Verify an example of each property of determinants found in Theorems 13.1.23 - 13.1.25 for 2×2 matrices.
17. A matrix is said to be **orthogonal** if $A^T A = I$. Thus the inverse of an orthogonal matrix is just its transpose. What are the possible values of $\det(A)$ if A is an orthogonal matrix?
18. Fill in the missing entries to make the matrix orthogonal as in Problem 17.

$$\begin{pmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{\sqrt{12}}{6} \\ \frac{1}{\sqrt{2}} & - & - \\ - & \frac{\sqrt{6}}{3} & - \end{pmatrix}.$$

19. If A^{-1} exist, what is the relationship between $\det(A)$ and $\det(A^{-1})$. Explain your answer.
20. Is it true that $\det(A + B) = \det(A) + \det(B)$? If this is so, explain why it is so and if it is not so, give a counter example.
21. Let A be an $r \times r$ matrix and suppose there are $r - 1$ rows (columns) such that all rows (columns) are linear combinations of these $r - 1$ rows (columns). Show $\det(A) = 0$.
22. Show $\det(aA) = a^n \det(A)$ where here A is an $n \times n$ matrix and a is a scalar.
23. Suppose A is an upper triangular matrix. Show that A^{-1} exists if and only if all entries of the main diagonal are non zero. Is it true that A^{-1} will also be upper triangular? Explain. Is everything the same for lower triangular matrices?
24. Let A and B be two $n \times n$ matrices. $A \sim B$ (A is **similar** to B) means there exists an invertible matrix, S such that $A = S^{-1}BS$. Show that if $A \sim B$, then $B \sim A$. Show also that $A \sim A$ and that if $A \sim B$ and $B \sim C$, then $A \sim C$.
25. In the context of Problem 24 show that if $A \sim B$, then $\det(A) = \det(B)$.
26. Two $n \times n$ matrices, A and B , are similar if $B = S^{-1}AS$ for some invertible $n \times n$ matrix, S . Show that if two matrices are similar, they have the same characteristic polynomials. The characteristic polynomial of an $n \times n$ matrix, M is the polynomial, $\det(\lambda I - M)$.
27. Prove by doing computations that $\det(AB) = \det(A) \det(B)$ if A and B are 2×2 matrices.
28. Illustrate with an example of 2×2 matrices that the determinant of a product equals the product of the determinants.

29. An $n \times n$ matrix is called **nilpotent** if for some positive integer, k it follows $A^k = 0$. If A is a nilpotent matrix and k is the smallest possible integer such that $A^k = 0$, what are the possible values of $\det(A)$?

30. Use Cramer's rule to find the solution to

$$\begin{aligned}x + 2y &= 1 \\ 2x - y &= 2\end{aligned}$$

31. Use Cramer's rule to find the solution to

$$\begin{aligned}x + 2y + z &= 1 \\ 2x - y - z &= 2 \\ x + z &= 1\end{aligned}$$

32. Here is a matrix,

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 3 & 1 & 0 \end{pmatrix}$$

Determine whether the matrix has an inverse by finding whether the determinant is non zero.

33. Here is a matrix,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}$$

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

34. Here is a matrix,

$$\begin{pmatrix} 1 & t & t^2 \\ 0 & 1 & 2t \\ t & 0 & 2 \end{pmatrix}$$

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

35. Here is a matrix,

$$\begin{pmatrix} e^t & e^{-t} \cos t & e^{-t} \sin t \\ e^t & -e^{-t} \cos t - e^{-t} \sin t & -e^{-t} \sin t + e^{-t} \cos t \\ e^t & 2e^{-t} \sin t & -2e^{-t} \cos t \end{pmatrix}$$

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

36. Here is a matrix,

$$\begin{pmatrix} e^t & \cosh t & \sinh t \\ e^t & \sinh t & \cosh t \\ e^t & \cosh t & \sinh t \end{pmatrix}$$

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

37. Use the formula for the inverse in terms of the cofactor matrix to find if possible the inverses of the matrices

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 4 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

If it is not possible to take the inverse, explain why.

38. Use the formula for the inverse in terms of the cofactor matrix to find the inverse of the matrix,

$$A = \begin{pmatrix} e^t & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & e^t \cos t - e^t \sin t & e^t \cos t + e^t \sin t \end{pmatrix}.$$

39. Find the inverse if it exists of the matrix,

$$\begin{pmatrix} e^t & \cos t & \sin t \\ e^t & -\sin t & \cos t \\ e^t & -\cos t & -\sin t \end{pmatrix}.$$

40. Let $F(t) = \det \begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix}$. Verify

$$F'(t) = \det \begin{pmatrix} a'(t) & b'(t) \\ c(t) & d(t) \end{pmatrix} + \det \begin{pmatrix} a(t) & b(t) \\ c'(t) & d'(t) \end{pmatrix}.$$

Now suppose

$$F(t) = \det \begin{pmatrix} a(t) & b(t) & c(t) \\ d(t) & e(t) & f(t) \\ g(t) & h(t) & i(t) \end{pmatrix}.$$

Use Laplace expansion and the first part to verify $F'(t) =$

$$\begin{aligned} & \det \begin{pmatrix} a'(t) & b'(t) & c'(t) \\ d(t) & e(t) & f(t) \\ g(t) & h(t) & i(t) \end{pmatrix} + \det \begin{pmatrix} a(t) & b(t) & c(t) \\ d'(t) & e'(t) & f'(t) \\ g(t) & h(t) & i(t) \end{pmatrix} \\ & + \det \begin{pmatrix} a(t) & b(t) & c(t) \\ d(t) & e(t) & f(t) \\ g'(t) & h'(t) & i'(t) \end{pmatrix}. \end{aligned}$$

Conjecture a general result valid for $n \times n$ matrices and explain why it will be true. Can a similar thing be done with the columns?

41. Let $Ly = y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_1(x)y' + a_0(x)y$ where the a_i are given continuous functions defined on a closed interval, (a, b) and y is some function which has n derivatives so it makes sense to write Ly . Suppose $Ly_k = 0$ for $k = 1, 2, \dots, n$. The **Wronskian** of these functions, y_i is defined as

$$W(y_1, \dots, y_n)(x) \equiv \det \begin{pmatrix} y_1(x) & \cdots & y_n(x) \\ y_1'(x) & \cdots & y_n'(x) \\ \vdots & & \vdots \\ y_1^{(n-1)}(x) & \cdots & y_n^{(n-1)}(x) \end{pmatrix}$$

Show that for $W(x) = W(y_1, \dots, y_n)(x)$ to save space,

$$W'(x) = \det \begin{pmatrix} y_1(x) & \cdots & y_n(x) \\ y_1'(x) & \cdots & y_n'(x) \\ \vdots & & \vdots \\ y_1^{(n)}(x) & \cdots & y_n^{(n)}(x) \end{pmatrix}.$$

Now use the differential equation, $Ly = 0$ which is satisfied by each of these functions, y_i and properties of determinants presented above to verify that $W' +$

$a_{n-1}(x)W = 0$. Give an explicit solution of this linear differential equation, **Abel's formula**, and use your answer to verify that the Wronskian of these solutions to the equation, $Ly = 0$ either vanishes identically on (a, b) or never. **Hint:** To solve the differential equation, let $A'(x) = a_{n-1}(x)$ and multiply both sides of the differential equation by $e^{A(x)}$ and then argue the left side is the derivative of something.

Spectral Theory

14.0.1 Outcomes

1. Describe the eigenvalue problem geometrically and algebraically.
2. Evaluate the spectrum and eigenvectors for a square matrix.

14.1 Eigenvalues And Eigenvectors Of A Matrix

Spectral Theory refers to the study of eigenvalues and eigenvectors of a matrix. It is of fundamental importance in many areas. Row operations will no longer be such a useful tool in this subject.

14.1.1 Definition Of Eigenvectors And Eigenvalues

In this section, the field of scalars, $\mathbb{F}$ will equal $\mathbb{C}$, the complex numbers, if nothing is stated to the contrary.

To illustrate the idea behind what will be discussed, consider the following example.

Example 14.1.1 *Here is a matrix.*

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix}.$$

Multiply this matrix by the vector

$$\begin{pmatrix} -5 \\ -4 \\ 3 \end{pmatrix}$$

and see what happens. Then multiply it by

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

and see what happens. Does this matrix act this way for some other vector?

First

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix} \begin{pmatrix} -5 \\ -4 \\ 3 \end{pmatrix} = \begin{pmatrix} -50 \\ -40 \\ 30 \end{pmatrix} = 10 \begin{pmatrix} -5 \\ -4 \\ 3 \end{pmatrix}.$$

Next

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

When you multiply the first vector by the given matrix, it stretched the vector, multiplying it by 10. When you multiplied the matrix by the second vector it sent it to the zero vector. Now consider

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -5 \\ 38 \\ -11 \end{pmatrix}.$$

In this case, multiplication by the matrix did not result in merely multiplying the vector by a number.

In the above example, the first two vectors were called eigenvectors and the numbers, 10 and 0 are called eigenvalues. Not every number is an eigenvalue and not every vector is an eigenvector.

Definition 14.1.2 Let M be an $n \times n$ matrix and let $\mathbf{x} \in \mathbb{C}^n$ be a nonzero vector for which

$$M\mathbf{x} = \lambda\mathbf{x} \quad (14.1)$$

for some scalar, λ . Then $\mathbf{x}$ is called an **eigenvector** and λ is called an **eigenvalue** (*characteristic value*) of the matrix, M .

Note: Eigenvectors are never equal to zero!

The set of all eigenvalues of an $n \times n$ matrix, M , is denoted by $\sigma(M)$ and is referred to as the **spectrum** of M .

The eigenvectors of a matrix M are those vectors, $\mathbf{x}$ for which multiplication by M results in a scalar multiple of $\mathbf{x}$. Since the zero vector, $\mathbf{0}$ has no direction this would make no sense for the zero vector. As noted above, $\mathbf{0}$ is never allowed to be an eigenvector. How can eigenvectors be identified? Suppose $\mathbf{x}$ satisfies 14.1. Then

$$(M - \lambda I)\mathbf{x} = \mathbf{0}$$

for some $\mathbf{x} \neq \mathbf{0}$. (Equivalently, you could write $(\lambda I - M)\mathbf{x} = \mathbf{0}$.) Sometimes we will use

$$(\lambda I - M)\mathbf{x} = \mathbf{0}$$

and sometimes

$$(M - \lambda I)\mathbf{x} = \mathbf{0}.$$

It makes absolutely no difference and you should use whichever you like better. Therefore, the matrix $M - \lambda I$ cannot have an inverse because if it did, the equation could be solved,

$$\mathbf{x} = \left((M - \lambda I)^{-1} (M - \lambda I) \right) \mathbf{x} = (M - \lambda I)^{-1} ((M - \lambda I)\mathbf{x}) = (M - \lambda I)^{-1} \mathbf{0} = \mathbf{0},$$

and this would require $\mathbf{x} = \mathbf{0}$, contrary to the requirement that $\mathbf{x} \neq \mathbf{0}$. By Theorem 13.2.1 on Page 333,

$$\det(M - \lambda I) = 0. \quad (14.2)$$

(Equivalently you could write $\det(\lambda I - M) = 0$.) The expression, $\det(\lambda I - M)$ or equivalently, $\det(M - \lambda I)$ is a polynomial called the **characteristic polynomial** and the above equation is called the characteristic equation. For M an $n \times n$ matrix, it follows from the theorem on expanding a matrix by its cofactor that $\det(M - \lambda I)$ is a polynomial of degree n . As such, the equation, 14.2 has a solution, $\lambda \in \mathbb{C}$ by the fundamental theorem of algebra. Is it actually an eigenvalue? The answer is yes from Theorem 13.2.1 which implies that $M - \lambda I$ has no inverse. From linear algebra, it follows $M - \lambda I$ is not one to one and so there exists a nonzero vector $\mathbf{x}$ such that $(M - \lambda I)\mathbf{x} = \mathbf{0}$. If you are interested in the proof of this assertion, see Theorem C.2.20 in the appendix. This proves the following corollary.

Corollary 14.1.3 *Let M be an $n \times n$ matrix and $\det(M - \lambda I) = 0$. Then there exists a nonzero vector, $\mathbf{x} \in \mathbb{C}^n$ such that $(M - \lambda I)\mathbf{x} = \mathbf{0}$.*

14.1.2 Finding Eigenvectors And Eigenvalues

As an example, consider the following.

Example 14.1.4 *Find the eigenvalues and eigenvectors for the matrix,*

$$A = \begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix}.$$

You first need to identify the eigenvalues. Recall this requires the solution of the equation

$$\det(A - \lambda I) = 0.$$

In this case this equation is

$$\det \left(\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) = 0$$

When you expand this determinant and simplify, you find the equation you need to solve is

$$(\lambda - 5)(\lambda^2 - 20\lambda + 100) = 0$$

and so the eigenvalues are

$$5, 10, 10.$$

10 is listed twice because it is a zero of multiplicity two due to

$$\lambda^2 - 20\lambda + 100 = (\lambda - 10)^2.$$

Having found the eigenvalues, it only remains to find the eigenvectors. First find the eigenvectors for $\lambda = 5$. As explained above, this requires you to solve the equation,

$$\left(\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} - 5 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

That is you need to find the solution to

$$\begin{pmatrix} 0 & -10 & -5 \\ 2 & 9 & 2 \\ -4 & -8 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

By now this is an old problem. You set up the augmented matrix and row reduce to get the solution. Thus the matrix you must row reduce is

$$\left(\begin{array}{ccc|c} 0 & -10 & -5 & 0 \\ 2 & 9 & 2 & 0 \\ -4 & -8 & 1 & 0 \end{array} \right). \quad (14.3)$$

A reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & 0 & -\frac{5}{4} & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

and so the solution is any vector of the form

$$\begin{pmatrix} \frac{5}{4}t \\ -\frac{1}{2}t \\ t \end{pmatrix} = t \begin{pmatrix} \frac{5}{4} \\ -\frac{1}{2} \\ 1 \end{pmatrix}$$

where $t \in \mathbb{F}$. You would obtain the same collection of vectors if you replaced t with $4t$. Thus a simpler description for the solutions to this system of equations whose augmented matrix is in 14.3 is

$$t \begin{pmatrix} 5 \\ -2 \\ 4 \end{pmatrix} \quad (14.4)$$

where $t \in \mathbb{F}$. Now you need to remember that you can't take $t = 0$ because this would result in the zero vector and

Eigenvectors are never equal to zero!

Other than this value, every other choice of z in 14.4 results in an eigenvector. It is a good idea to check your work! To do so, we will take the original matrix and multiply by this vector and see if we get 5 times this vector.

$$\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} \begin{pmatrix} 5 \\ -2 \\ 4 \end{pmatrix} = \begin{pmatrix} 25 \\ -10 \\ 20 \end{pmatrix} = 5 \begin{pmatrix} 5 \\ -2 \\ 4 \end{pmatrix}$$

so it appears this is correct. Always check your work on these problems if you care about getting the answer right.

The parameter, t is sometimes called a **free variable**. The set of vectors in 14.4 is called the **eigenspace** and it equals $\ker(A - \lambda I)$. You should observe that in this case the eigenspace has dimension 1 because the eigenspace is the span of a single vector. In general, you obtain the solution from the row echelon form and the number of different free variables gives you the dimension of the eigenspace. Just remember that not every vector in the eigenspace is an eigenvector. The vector, $\mathbf{0}$ is not an eigenvector although it is in the eigenspace because

Eigenvectors are never equal to zero!

Next consider the eigenvectors for $\lambda = 10$. These vectors are solutions to the equation,

$$\left(\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} - 10 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

That is you must find the solutions to

$$\begin{pmatrix} -5 & -10 & -5 \\ 2 & 4 & 2 \\ -4 & -8 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

which reduces to consideration of the augmented matrix,

$$\left(\begin{array}{ccc|c} -5 & -10 & -5 & 0 \\ 2 & 4 & 2 & 0 \\ -4 & -8 & -4 & 0 \end{array} \right)$$

A reduced echelon form for this matrix is

$$\begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and so the eigenvectors are of the form

$$\begin{pmatrix} -2s - t \\ s \\ t \end{pmatrix} = s \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

You can't pick t and s both equal to zero because this would result in the zero vector and

Eigenvectors are never equal to zero!

However, every other choice of t and s does result in an eigenvector for the eigenvalue $\lambda = 10$. As in the case for $\lambda = 5$ you should check your work if you care about getting it right.

$$\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -10 \\ 0 \\ 10 \end{pmatrix} = 10 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

so it worked. The other vector will also work. Check it.

14.1.3 A Warning

The above example shows how to find eigenvectors and eigenvalues algebraically. You may have noticed it is a bit long. Sometimes students try to first row reduce the matrix before looking for eigenvalues. This is a **terrible idea** because row operations destroy the eigenvalues. The eigenvalue problem is really not about row operations.

The general eigenvalue problem is the hardest problem in algebra and people still do research on ways to find eigenvalues and their eigenvectors. If you are doing anything which would yield a way to find eigenvalues and eigenvectors for general matrices without too much trouble, the thing you are doing will certainly be wrong. The problems you will see here are not too hard because they are cooked up to be easy. General methods to compute eigenvalues and eigenvectors numerically are available. These methods work even when the problem is not cooked up to be easy. To see this, see a linear algebra book.

If you are so fortunate as to find the eigenvalues as in the above example, then finding the eigenvectors does reduce to row operations and this part of the problem is easy. However, finding the eigenvalues along with the eigenvectors is anything but easy because for an $n \times n$ matrix, it involves solving a polynomial equation of degree n . If you only find a good approximation to the eigenvalue, it won't work. It either is or is

not an eigenvalue and if it is not, the only solution to the equation, $(M - \lambda I) \mathbf{x} = \mathbf{0}$ will be the zero solution as explained above and

Eigenvectors are never equal to zero!

Here is another example.

Example 14.1.5 *Let*

$$A = \begin{pmatrix} 2 & 2 & -2 \\ 1 & 3 & -1 \\ -1 & 1 & 1 \end{pmatrix}$$

First find the eigenvalues.

$$\det \left(\begin{pmatrix} 2 & 2 & -2 \\ 1 & 3 & -1 \\ -1 & 1 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) = 0$$

This reduces to $\lambda^3 - 6\lambda^2 + 8\lambda = 0$ and the solutions are 0, 2, and 4.

0 Can be an Eigenvalue!

Now find the eigenvectors. For $\lambda = 0$ the augmented matrix for finding the solutions is

$$\left(\begin{array}{ccc|c} 2 & 2 & -2 & 0 \\ 1 & 3 & -1 & 0 \\ -1 & 1 & 1 & 0 \end{array} \right)$$

and a reduced echelon form is

$$\left(\begin{array}{cccc} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

Therefore, the eigenvectors are of the form

$$t \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

where $t \neq 0$.

Next find the eigenvectors for $\lambda = 2$. The augmented matrix for the system of equations needed to find these eigenvectors is

$$\left(\begin{array}{ccc|c} 0 & 2 & -2 & 0 \\ 1 & 1 & -1 & 0 \\ -1 & 1 & -1 & 0 \end{array} \right)$$

and a reduced echelon form is

$$\left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

and so the eigenvectors are of the form

$$t \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

where $t \neq 0$.

Finally find the eigenvectors for $\lambda = 4$. The augmented matrix for the system of equations needed to find these eigenvectors is

$$\left(\begin{array}{ccc|c} -2 & 2 & -2 & 0 \\ 1 & -1 & -1 & 0 \\ -1 & 1 & -3 & 0 \end{array} \right)$$

and a reduced echelon form is

$$\left(\begin{array}{cccc} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right).$$

Therefore, the eigenvectors are of the form

$$t \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

where $t \neq 0$.

14.1.4 Complex Eigenvalues

Sometimes you have to consider eigenvalues which are complex numbers. This occurs in differential equations for example. You do these problems exactly the same way as you do the ones in which the eigenvalues are real. Here is an example.

Example 14.1.6 Find the eigenvalues and eigenvectors of the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix}.$$

You need to find the eigenvalues. Solve

$$\det \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) = 0.$$

This reduces to $(\lambda - 1)(\lambda^2 - 4\lambda + 5) = 0$. The solutions are $\lambda = 1, \lambda = 2 + i, \lambda = 2 - i$.

There is nothing new about finding the eigenvectors for $\lambda = 1$ so consider the eigenvalue $\lambda = 2 + i$. You need to solve

$$\left((2 + i) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

In other words, you must consider the augmented matrix,

$$\left(\begin{array}{ccc|c} 1+i & 0 & 0 & 0 \\ 0 & i & 1 & 0 \\ 0 & -1 & i & 0 \end{array} \right)$$

for the solution. Divide the top row by $(1+i)$ and then take $-i$ times the second row and add to the bottom. This yields

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & i & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

Now multiply the second row by $-i$ to obtain

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & -i & 0 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

Therefore, the eigenvectors are of the form

$$t \begin{pmatrix} 0 \\ i \\ 1 \end{pmatrix}.$$

You should find the eigenvectors for $\lambda = 2 - i$. These are

$$t \begin{pmatrix} 0 \\ -i \\ 1 \end{pmatrix}.$$

As usual, if you want to get it right you had better check it.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ -i \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 - 2i \\ 2 - i \end{pmatrix} = (2 - i) \begin{pmatrix} 0 \\ -i \\ 1 \end{pmatrix}$$

so it worked.

14.2 Exercises

1. State the eigenvalue problem from an algebraic perspective.
2. State the eigenvalue problem from a geometric perspective.
3. Suppose T is a nonzero linear transformation and it satisfies $T^2 = T$. Show that the eigenvalues are either 1 or 0. Then show that $\lambda = 1$ is an eigenvalue.
4. Is it possible for a nonzero matrix to have only 0 as an eigenvalue?
5. Show that if $A\mathbf{x} = \lambda\mathbf{x}$ and $A\mathbf{y} = \lambda\mathbf{y}$, then whenever a, b are scalars,

$$A(a\mathbf{x} + b\mathbf{y}) = \lambda(a\mathbf{x} + b\mathbf{y}).$$

Does this imply that $a\mathbf{x} + b\mathbf{y}$ is an eigenvector? Explain.

6. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} -1 & -1 & 7 \\ -1 & 0 & 4 \\ -1 & -1 & 5 \end{pmatrix}.$$

7. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} -3 & -7 & 19 \\ -2 & -1 & 8 \\ -2 & -3 & 10 \end{pmatrix}.$$

8. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} -7 & -12 & 30 \\ -3 & -7 & 15 \\ -3 & -6 & 14 \end{pmatrix}.$$

9. Find the eigenvalues and eigenvectors of the matrix

tors of the matrix

$$\begin{pmatrix} 7 & -2 & 0 \\ 8 & -1 & 0 \\ -2 & 4 & 6 \end{pmatrix}.$$

10. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 3 & -2 & -1 \\ 0 & 5 & 1 \\ 0 & 2 & 4 \end{pmatrix}.$$

11. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 6 & 8 & -23 \\ 4 & 5 & -16 \\ 3 & 4 & -12 \end{pmatrix}$$

12. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 5 & 2 & -5 \\ 12 & 3 & -10 \\ 12 & 4 & -11 \end{pmatrix}.$$

13. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 20 & 9 & -18 \\ 6 & 5 & -6 \\ 30 & 14 & -27 \end{pmatrix}.$$

14. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 1 & 26 & -17 \\ 4 & -4 & 4 \\ -9 & -18 & 9 \end{pmatrix}.$$

15. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 3 & -1 & -2 \\ 11 & 3 & -9 \\ 8 & 0 & -6 \end{pmatrix}.$$

16. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} -2 & 1 & 2 \\ -11 & -2 & 9 \\ -8 & 0 & 7 \end{pmatrix}.$$

17. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 2 & 1 & -1 \\ 2 & 3 & -2 \\ 2 & 2 & -1 \end{pmatrix}.$$

18. Find the complex eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} 4 & -2 & -2 \\ 0 & 2 & -2 \\ 2 & 0 & 2 \end{pmatrix}.$$

19. Let T be the linear transformation which reflects vectors about the x axis. Find a matrix for T and then find its eigenvalues and eigenvectors.

20. Let T be the linear transformation which rotates all vectors in $\mathbb{R}^2$ counterclockwise through an angle of $\pi/2$. Find a matrix of T and then find eigenvalues and eigenvectors.

21. Let T be the linear transformation which reflects all vectors in $\mathbb{R}^3$ through the plane determined by $z = 0$. Find a matrix for T and then obtain its eigenvalues and eigenvectors.

22. Let A be the 2×2 matrix of a linear transformation which rotates all vectors through an angle of $\pi/3$. Explain from geometric reasoning why this matrix cannot have any real eigenvalues. For which angles will such a rotation matrix have a real eigenvalue?

Part III

Vectors In $\mathbb{R}^n$

Vectors And Points In $\mathbb{R}^n$

15.0.1 Outcomes

1. Be able to represent a vector in each of the following ways for $n = 2, 3$
 - (a) as a directed arrow in n space
 - (b) as an ordered n tuple
 - (c) as a linear combination of unit coordinate vectors
2. Carry out the vector operations:
 - (a) addition
 - (b) scalar multiplication
 - (c) find magnitude (norm or length)
 - (d) Find the vector of unit length in the direction of a given vector.
3. Represent the operations of vector addition, scalar multiplication and norm geometrically.
4. Recall and apply the basic properties of vector addition, scalar multiplication and norm.
5. Model and solve application problems using vectors.
6. Describe an open ball in $\mathbb{R}^n$.
7. Determine whether a set in $\mathbb{R}^n$ is open, closed, or neither.

15.1 Open And Closed Sets

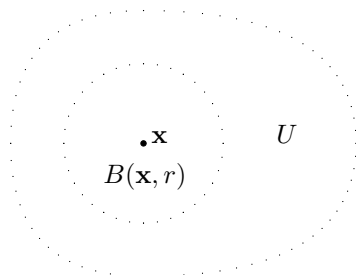
Eventually, one must consider functions which are defined on subsets of $\mathbb{R}^n$ and their properties. The next definition will end up being quite important. It describes a type of subset of $\mathbb{R}^n$ with the property that if $\mathbf{x}$ is in this set, then so is $\mathbf{y}$ whenever $\mathbf{y}$ is close enough to $\mathbf{x}$.

Definition 15.1.1 *Let $U \subseteq \mathbb{R}^n$. U is an **open set** if whenever $\mathbf{x} \in U$, there exists $r > 0$ such that $B(\mathbf{x}, r) \subseteq U$. More generally, if U is any subset of $\mathbb{R}^n$, $\mathbf{x} \in U$ is an **interior point** of U if there exists $r > 0$ such that $\mathbf{x} \in B(\mathbf{x}, r) \subseteq U$. In other words U is an open set exactly when every point of U is an interior point of U .*

If there is something called an open set, surely there should be something called a closed set and here is the definition of one.

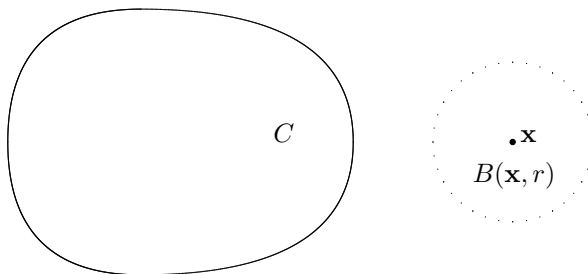
Definition 15.1.2 A subset, C , of $\mathbb{R}^n$ is called a **closed set** if $\mathbb{R}^n \setminus C$ is an open set. The symbol, $\mathbb{R}^n \setminus C$ denotes everything in $\mathbb{R}^n$ which is not in C . It is also called the **complement** of C . The symbol, S^C is a short way of writing $\mathbb{R}^n \setminus S$.

To illustrate this definition, consider the following picture.



You see in this picture how the edges are dotted. This is because an open set, can not include the edges or the set would fail to be open. For example, consider what would happen if you picked a point out on the edge of U in the above picture. Every open ball centered at that point would have in it some points which are outside U . Therefore, such a point would violate the above definition. You also see the edges of $B(\mathbf{x}, r)$ dotted suggesting that $B(\mathbf{x}, r)$ ought to be an open set. This is intuitively clear but does require a proof. This will be done in the next theorem and will give examples of open sets. Also, you can see that if $\mathbf{x}$ is close to the edge of U , you might have to take r to be very small.

It is roughly the case that open sets don't have their skins while closed sets do. Here is a picture of a closed set, C .



Note that $\mathbf{x} \notin C$ and since $\mathbb{R}^n \setminus C$ is open, there exists a ball, $B(\mathbf{x}, r)$ contained entirely in $\mathbb{R}^n \setminus C$. If you look at $\mathbb{R}^n \setminus C$, what would be its skin? It can't be in $\mathbb{R}^n \setminus C$ and so it must be in C . This is a rough heuristic explanation of what is going on with these definitions. Also note that $\mathbb{R}^n$ and $\emptyset$ are both open and closed. Here is why. If $\mathbf{x} \in \emptyset$, then there must be a ball centered at $\mathbf{x}$ which is also contained in $\emptyset$. This must be considered to be true because there is nothing in $\emptyset$ so there can be no example to show it false¹. Therefore, from the definition, it follows $\emptyset$ is open. It is also closed because

¹To a mathematician, the statement: Whenever a pig is born with wings it can fly must be taken as true. We do not consider biological or aerodynamic considerations in such statements. There is no such thing as a winged pig and therefore, all winged pigs must be superb flyers since there can be no example of one which is not. On the other hand we would also consider the statement: Whenever a pig is born with wings it can't possibly fly, as equally true. The point is, you can say anything you want about the elements of the empty set and no one can gainsay your statement. Therefore, such statements are considered as true by default. You may say this is a very strange way of thinking about truth and ultimately this is because mathematics is not about truth. It is more about consistency and logic.

if $\mathbf{x} \notin \emptyset$, then $B(\mathbf{x}, 1)$ is also contained in $\mathbb{R}^n \setminus \emptyset = \mathbb{R}^n$. Therefore, $\emptyset$ is both open and closed. From this, it follows $\mathbb{R}^n$ is also both open and closed.



Theorem 15.1.3 Let $\mathbf{x} \in \mathbb{R}^n$ and let $r \geq 0$. Then $B(\mathbf{x}, r)$ is an open set. Also,

$$D(\mathbf{x}, r) \equiv \{\mathbf{y} \in \mathbb{R}^n : |\mathbf{y} - \mathbf{x}| \leq r\}$$

is a closed set.

Proof: Suppose $\mathbf{y} \in B(\mathbf{x}, r)$. It is necessary to show there exists $r_1 > 0$ such that $B(\mathbf{y}, r_1) \subseteq B(\mathbf{x}, r)$. Define $r_1 \equiv r - |\mathbf{x} - \mathbf{y}|$. Then if $|\mathbf{z} - \mathbf{y}| < r_1$, it follows from the above triangle inequality that

$$\begin{aligned} |\mathbf{z} - \mathbf{x}| &= |\mathbf{z} - \mathbf{y} + \mathbf{y} - \mathbf{x}| \\ &\leq |\mathbf{z} - \mathbf{y}| + |\mathbf{y} - \mathbf{x}| \\ &< r_1 + |\mathbf{y} - \mathbf{x}| = r - |\mathbf{x} - \mathbf{y}| + |\mathbf{y} - \mathbf{x}| = r. \end{aligned}$$

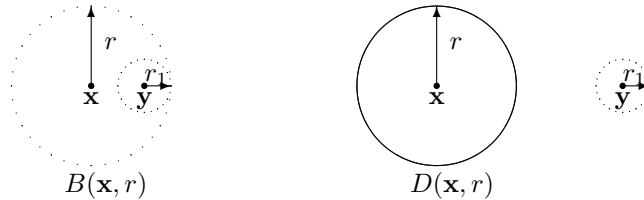
Note that if $r = 0$ then $B(\mathbf{x}, r) = \emptyset$, the empty set. This is because if $\mathbf{y} \in \mathbb{R}^n$, $|\mathbf{x} - \mathbf{y}| \geq 0$ and so $\mathbf{y} \notin B(\mathbf{x}, 0)$. Since $\emptyset$ has no points in it, it must be open because every point in it, (There are none.) satisfies the desired property of being an interior point.

Now suppose $\mathbf{y} \notin D(\mathbf{x}, r)$. Then $|\mathbf{x} - \mathbf{y}| > r$ and defining $\delta \equiv |\mathbf{x} - \mathbf{y}| - r$, it follows that if $\mathbf{z} \in B(\mathbf{y}, \delta)$, then by the triangle inequality,

$$\begin{aligned} |\mathbf{x} - \mathbf{z}| &\geq |\mathbf{x} - \mathbf{y}| - |\mathbf{y} - \mathbf{z}| > |\mathbf{x} - \mathbf{y}| - \delta \\ &= |\mathbf{x} - \mathbf{y}| - (|\mathbf{x} - \mathbf{y}| - r) = r \end{aligned}$$

and this shows that $B(\mathbf{y}, \delta) \subseteq \mathbb{R}^n \setminus D(\mathbf{x}, r)$. Since $\mathbf{y}$ was an arbitrary point in $\mathbb{R}^n \setminus D(\mathbf{x}, r)$, it follows $\mathbb{R}^n \setminus D(\mathbf{x}, r)$ is an open set which shows from the definition that $D(\mathbf{x}, r)$ is a closed set as claimed.

A picture which is descriptive of the conclusion of the above theorem which also implies the manner of proof is the following.



Recall $\mathbb{R}^2$ consists of ordered pairs, (x, y) such that $x \in \mathbb{R}$ and $y \in \mathbb{R}$. $\mathbb{R}^2$ is also written as $\mathbb{R} \times \mathbb{R}$. In general, the following definition holds.

Definition 15.1.4 The *Cartesian product* of two sets, $A \times B$, means

$$\{(a, b) : a \in A, b \in B\}.$$

If you have n sets, $A_1, A_2, \dots, A_n$

$$\prod_{i=1}^n A_i = \{(x_1, x_2, \dots, x_n) : \text{each } x_i \in A_i\}.$$

Now suppose $A \subseteq \mathbb{R}^m$ and $B \subseteq \mathbb{R}^n$. Then if $(\mathbf{x}, \mathbf{y}) \in A \times B$, $\mathbf{x} = (x_1, \dots, x_m)$ and $\mathbf{y} = (y_1, \dots, y_n)$, the following identification will be made.

$$(\mathbf{x}, \mathbf{y}) = (x_1, \dots, x_m, y_1, \dots, y_n) \in \mathbb{R}^{n+m}.$$

Similarly, starting with something in $\mathbb{R}^{n+m}$, you can write it in the form $(\mathbf{x}, \mathbf{y})$ where $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$. The following theorem has to do with the Cartesian product of two closed sets or two open sets. Also here is an important definition.

Definition 15.1.5 A set, $A \subseteq \mathbb{R}^n$ is said to be **bounded** if there exist finite intervals, $[a_i, b_i]$ such that

$$A \subseteq \prod_{i=1}^n [a_i, b_i].$$

Theorem 15.1.6 Let U be an open set in $\mathbb{R}^m$ and let V be an open set in $\mathbb{R}^n$. Then $U \times V$ is an open set in $\mathbb{R}^{n+m}$. If C is a closed set in $\mathbb{R}^m$ and H is a closed set in $\mathbb{R}^n$, then $C \times H$ is a closed set in $\mathbb{R}^{n+m}$. If C and H are bounded, then so is $C \times H$.

Proof: Let $(\mathbf{x}, \mathbf{y}) \in U \times V$. Since U is open, there exists $r_1 > 0$ such that $B(\mathbf{x}, r_1) \subseteq U$. Similarly, there exists $r_2 > 0$ such that $B(\mathbf{y}, r_2) \subseteq V$. Now

$$B((\mathbf{x}, \mathbf{y}), \delta) \equiv$$

$$\left\{ (\mathbf{s}, \mathbf{t}) \in \mathbb{R}^{n+m} : \sum_{k=1}^m |x_k - s_k|^2 + \sum_{j=1}^n |y_j - t_j|^2 < \delta^2 \right\}$$

Therefore, if $\delta \equiv \min(r_1, r_2)$ and $(\mathbf{s}, \mathbf{t}) \in B((\mathbf{x}, \mathbf{y}), \delta)$, then it follows that $\mathbf{s} \in B(\mathbf{x}, r_1) \subseteq U$ and that $\mathbf{t} \in B(\mathbf{y}, r_2) \subseteq V$ which shows that $B((\mathbf{x}, \mathbf{y}), \delta) \subseteq U \times V$. Hence $U \times V$ is open as claimed.

Next suppose $(\mathbf{x}, \mathbf{y}) \notin C \times H$. It is necessary to show there exists $\delta > 0$ such that $B((\mathbf{x}, \mathbf{y}), \delta) \subseteq \mathbb{R}^{n+m} \setminus (C \times H)$. Either $\mathbf{x} \notin C$ or $\mathbf{y} \notin H$ since otherwise $(\mathbf{x}, \mathbf{y})$ would be a point of $C \times H$. Suppose therefore, that $\mathbf{x} \notin C$. Since C is closed, there exists $r > 0$ such that $B(\mathbf{x}, r) \subseteq \mathbb{R}^m \setminus C$. Consider $B((\mathbf{x}, \mathbf{y}), r)$. If $(\mathbf{s}, \mathbf{t}) \in B((\mathbf{x}, \mathbf{y}), r)$, it follows that $\mathbf{s} \in B(\mathbf{x}, r)$ which is contained in $\mathbb{R}^m \setminus C$. Therefore, $B((\mathbf{x}, \mathbf{y}), r) \subseteq \mathbb{R}^{n+m} \setminus (C \times H)$ showing $C \times H$ is closed. A similar argument holds if $\mathbf{y} \notin H$.

If C is bounded, there exist $[a_i, b_i]$ such that $C \subseteq \prod_{i=1}^m [a_i, b_i]$ and if H is bounded, $H \subseteq \prod_{i=m+1}^{m+n} [a_i, b_i]$ for intervals $[a_{m+1}, b_{m+1}], \dots, [a_{m+n}, b_{m+n}]$. Therefore, $C \times H \subseteq \prod_{i=1}^{m+n} [a_i, b_i]$ and this establishes the last part of this theorem.

15.2 Exercises

1. Let $U = \{(x, y, z) \text{ such that } z > 0\}$. Determine whether U is open, closed or neither.
2. Let $U = \{(x, y, z) \text{ such that } z \geq 0\}$. Determine whether U is open, closed or neither.

3. Let $U = \{(x, y, z) \text{ such that } \sqrt{x^2 + y^2 + z^2} < 1\}$. Determine whether U is open, closed or neither.
4. Let $U = \{(x, y, z) \text{ such that } \sqrt{x^2 + y^2 + z^2} \leq 1\}$. Determine whether U is open, closed or neither.
5. Show carefully that $\mathbb{R}^n$ is both open and closed.
6. Show that every open set in $\mathbb{R}^n$ is the union of open balls contained in it.
7. Show the intersection of any two open sets is an open set.
8. If S is a nonempty subset of $\mathbb{R}^p$, a point, $\mathbf{x}$ is said to be a **limit point** of S if $B(\mathbf{x}, r)$ contains infinitely many points of S for each $r > 0$. Show this is equivalent to saying that $B(\mathbf{x}, r)$ contains a point of S different than $\mathbf{x}$ for each $r > 0$.
9. Closed sets were defined to be those sets which are complements of open sets. Show that a set is closed if and only if it contains all its limit points.

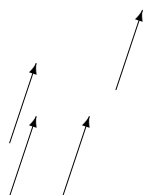
15.3 Physical Vectors

Suppose you push on something. What is important? There are really two things which are important, how hard you push and the direction you push. This illustrates the concept of force.

Definition 15.3.1 *Force is a vector. The magnitude of this vector is a measure of how hard it is pushing. It is measured in units such as Newtons or pounds or tons. Its direction is the direction in which the push is taking place.*

Of course this is a little vague and will be left a little vague until the presentation of Newton's second law later.

Vectors are used to model force and other physical vectors like velocity. What was just described would be called a force vector. It has two essential ingredients, its magnitude and its direction. Geometrically think of vectors as directed line segments or arrows as shown in the following picture in which all the directed line segments are considered to be the same vector because they have the same direction, the direction in which the arrows point, and the same magnitude (length).



Because of this fact that only direction and magnitude are important, it is always possible to put a vector in a certain particularly simple form. Let $\overrightarrow{\mathbf{pq}}$ be a directed line segment or vector. Then from Definition 11.4.4 it follows that $\overrightarrow{\mathbf{pq}}$ consists of the points of the form

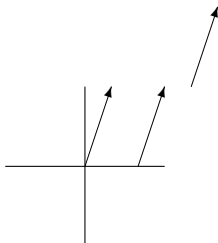
$$\mathbf{p} + t(\mathbf{q} - \mathbf{p})$$

where $t \in [0, 1]$. Subtract $\mathbf{p}$ from all these points to obtain the directed line segment consisting of the points

$$\mathbf{0} + t(\mathbf{q} - \mathbf{p}), \quad t \in [0, 1].$$

The point in $\mathbb{R}^n$, $\mathbf{q} - \mathbf{p}$, will represent the vector.

Geometrically, the arrow, $\overrightarrow{\mathbf{pq}}$, was slid so it points in the same direction and the base is at the origin, $\mathbf{0}$. For example, see the following picture.



In this way vectors can be identified with points of $\mathbb{R}^n$.

Definition 15.3.2 Let $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$. The **position vector** of this point is the vector whose point is at $\mathbf{x}$ and whose tail is at the origin, $(0, \dots, 0)$. If $\mathbf{x} = (x_1, \dots, x_n)$ is called a vector, the vector which is meant is this position vector just described. Another term associated with this is **standard position**. A vector is in standard position if the tail is placed at the origin.

It is customary to identify the point in $\mathbb{R}^n$ with its position vector.

The magnitude of a vector determined by a directed line segment $\overrightarrow{\mathbf{pq}}$ is just the distance between the point $\mathbf{p}$ and the point $\mathbf{q}$. By the distance formula this equals

$$\left(\sum_{k=1}^n (q_k - p_k)^2 \right)^{1/2} = |\mathbf{p} - \mathbf{q}|$$

and for $\mathbf{v}$ any vector in $\mathbb{R}^n$ the magnitude of $\mathbf{v}$ equals $(\sum_{k=1}^n v_k^2)^{1/2} = |\mathbf{v}|$.

Example 15.3.3 Consider the vector, $\mathbf{v} \equiv (1, 2, 3)$ in $\mathbb{R}^n$. Find $|\mathbf{v}|$.

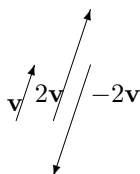
First, the vector is the directed line segment (arrow) which has its base at $\mathbf{0} \equiv (0, 0, 0)$ and its point at $(1, 2, 3)$. Therefore,

$$|\mathbf{v}| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}.$$

What is the geometric significance of scalar multiplication? If $\mathbf{a}$ represents the vector, $\mathbf{v}$ in the sense that when it is slid to place its tail at the origin, the element of $\mathbb{R}^n$ at its point is $\mathbf{a}$, what is $r\mathbf{v}$?

$$\begin{aligned} |r\mathbf{v}| &= \left(\sum_{k=1}^n (ra_k)^2 \right)^{1/2} = \left(\sum_{k=1}^n r^2 (a_k)^2 \right)^{1/2} \\ &= (r^2)^{1/2} \left(\sum_{k=1}^n a_k^2 \right)^{1/2} = |r| |\mathbf{v}|. \end{aligned}$$

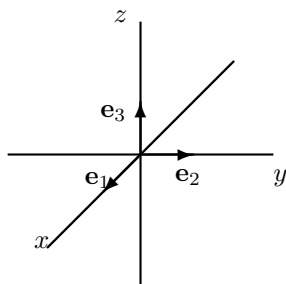
Thus the magnitude of $r\mathbf{v}$ equals $|r|$ times the magnitude of $\mathbf{v}$. If r is positive, then the vector represented by $r\mathbf{v}$ has the same direction as the vector, $\mathbf{v}$ because multiplying by the scalar, r , only has the effect of scaling all the distances. Thus the unit distance along any coordinate axis now has length r and in this rescaled system the vector is represented by $\mathbf{a}$. If $r < 0$ similar considerations apply except in this case all the a_i also change sign. From now on, $\mathbf{a}$ will be referred to as a vector instead of an element of $\mathbb{R}^n$ representing a vector as just described. The following picture illustrates the effect of scalar multiplication.



Note there are n special vectors which point along the coordinate axes. These are

$$\mathbf{e}_i \equiv (0, \dots, 0, 1, 0, \dots, 0)$$

where the 1 is in the i^{th} slot and there are zeros in all the other spaces. See the picture in the case of $\mathbb{R}^3$.



The direction of $\mathbf{e}_i$ is referred to as the i^{th} direction. Given a vector, $\mathbf{v} = (a_1, \dots, a_n)$, $a_i \mathbf{e}_i$ is the i^{th} component of the vector. Thus $a_i \mathbf{e}_i = (0, \dots, 0, a_i, 0, \dots, 0)$ and so this vector gives something possibly nonzero only in the i^{th} direction. Also, knowledge of the i^{th} component of the vector is equivalent to knowledge of the vector because it gives the entry in the i^{th} slot and for $\mathbf{v} = (a_1, \dots, a_n)$,

$$\mathbf{v} = \sum_{k=1}^n a_k \mathbf{e}_k.$$

What does addition of vectors mean physically? Suppose two forces are applied to some object. Each of these would be represented by a force vector and the two forces acting together would yield an overall force acting on the object which would also be a force vector known as the resultant. Suppose the two vectors are $\mathbf{a} = \sum_{k=1}^n a_k \mathbf{e}_k$ and $\mathbf{b} = \sum_{k=1}^n b_k \mathbf{e}_k$. Then the vector, $\mathbf{a}$ involves a component in the i^{th} direction, $a_i \mathbf{e}_i$ while the component in the i^{th} direction of $\mathbf{b}$ is $b_i \mathbf{e}_i$. Then it seems physically reasonable that the resultant vector should have a component in the i^{th} direction equal to $(a_i + b_i) \mathbf{e}_i$. This is exactly what is obtained when the vectors, $\mathbf{a}$ and $\mathbf{b}$ are added.

$$\begin{aligned} \mathbf{a} + \mathbf{b} &= (a_1 + b_1, \dots, a_n + b_n). \\ &= \sum_{i=1}^n (a_i + b_i) \mathbf{e}_i. \end{aligned}$$

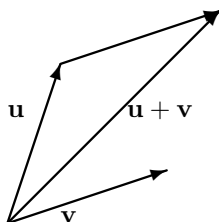
Thus the addition of vectors according to the rules of addition in $\mathbb{R}^n$ which were presented earlier, yields the appropriate vector which duplicates the cumulative effect of all the vectors in the sum.

What is the geometric significance of vector addition? Suppose $\mathbf{u}, \mathbf{v}$ are vectors,

$$\mathbf{u} = (u_1, \dots, u_n), \mathbf{v} = (v_1, \dots, v_n)$$

Then $\mathbf{u} + \mathbf{v} = (u_1 + v_1, \dots, u_n + v_n)$. How can one obtain this geometrically? Consider the directed line segment, $\overrightarrow{\mathbf{0u}}$ and then, starting at the end of this directed line segment,

follow the directed line segment $\overrightarrow{\mathbf{u}(\mathbf{u} + \mathbf{v})}$ to its end, $\mathbf{u} + \mathbf{v}$. In other words, place the vector $\mathbf{u}$ in standard position with its base at the origin and then slide the vector $\mathbf{v}$ till its base coincides with the point of $\mathbf{u}$. The point of this slid vector, determines $\mathbf{u} + \mathbf{v}$. To illustrate, see the following picture



Note the vector $\mathbf{u} + \mathbf{v}$ is the diagonal of a parallelogram determined from the two vectors $\mathbf{u}$ and $\mathbf{v}$ and that identifying $\mathbf{u} + \mathbf{v}$ with the directed diagonal of the parallelogram determined by the vectors $\mathbf{u}$ and $\mathbf{v}$ amounts to the same thing as the above procedure.

An item of notation should be mentioned here. In the case of $\mathbb{R}^n$ where $n \leq 3$, it is standard notation to use $\mathbf{i}$ for $\mathbf{e}_1$, $\mathbf{j}$ for $\mathbf{e}_2$, and $\mathbf{k}$ for $\mathbf{e}_3$. Now here are some applications of vector addition to some problems.

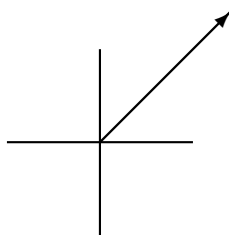
Example 15.3.4 *There are three ropes attached to a car and three people pull on these ropes. The first exerts a force of $2\mathbf{i} + 3\mathbf{j} - 2\mathbf{k}$ Newtons, the second exerts a force of $3\mathbf{i} + 5\mathbf{j} + \mathbf{k}$ Newtons and the third exerts a force of $5\mathbf{i} - \mathbf{j} + 2\mathbf{k}$ Newtons. Find the total force in the direction of $\mathbf{i}$.*

To find the total force add the vectors as described above. This gives $10\mathbf{i} + 7\mathbf{j} + \mathbf{k}$ Newtons. Therefore, the force in the $\mathbf{i}$ direction is 10 Newtons.

As mentioned earlier, the Newton is a unit of force like pounds.

Example 15.3.5 *An airplane flies North East at 100 miles per hour. Write this as a vector.*

A picture of this situation follows.



The vector has length 100. Now using that vector as the hypotenuse of a right triangle having equal sides, the sides should be each of length $100/\sqrt{2}$. Therefore, the vector would be $100/\sqrt{2}\mathbf{i} + 100/\sqrt{2}\mathbf{j}$.

This example also motivates the concept of **velocity**.

Definition 15.3.6 *The **speed** of an object is a measure of how fast it is going. It is measured in units of length per unit time. For example, miles per hour, kilometers per minute, feet per second. The **velocity** is a vector having the speed as the magnitude but also specifying the direction.*

Thus the velocity vector in the above example is $100/\sqrt{2}\mathbf{i} + 100/\sqrt{2}\mathbf{j}$.

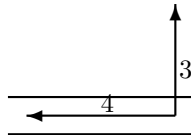
Example 15.3.7 The velocity of an airplane is $100\mathbf{i} + \mathbf{j} + \mathbf{k}$ measured in kilometers per hour and at a certain instant of time its position is $(1, 2, 1)$. Here imagine a Cartesian coordinate system in which the third component is altitude and the first and second components are measured on a line from West to East and a line from South to North. Find the position of this airplane one minute later.

Consider the vector $(1, 2, 1)$, is the initial position vector of the airplane. As it moves, the position vector changes. After one minute the airplane has moved in the $\mathbf{i}$ direction a distance of $100 \times \frac{1}{60} = \frac{5}{3}$ kilometer. In the $\mathbf{j}$ direction it has moved $\frac{1}{60}$ kilometer during this same time, while it moves $\frac{1}{60}$ kilometer in the $\mathbf{k}$ direction. Therefore, the new displacement vector for the airplane is

$$(1, 2, 1) + \left(\frac{5}{3}, \frac{1}{60}, \frac{1}{60}\right) = \left(\frac{8}{3}, \frac{121}{60}, \frac{121}{60}\right)$$

Example 15.3.8 A certain river is one half mile wide with a current flowing at 4 miles per hour from East to West. A man swims directly toward the opposite shore from the South bank of the river at a speed of 3 miles per hour. How far down the river does he find himself when he has swam across? How far does he end up swimming?

Consider the following picture.



You should write these vectors in terms of components. The velocity of the swimmer in still water would be $3\mathbf{j}$ while the velocity of the river would be $-4\mathbf{i}$. Therefore, the velocity of the swimmer is $-4\mathbf{i} + 3\mathbf{j}$. Since the component of velocity in the direction across the river is 3, it follows the trip takes $1/6$ hour or 10 minutes. The speed at which he travels is $\sqrt{4^2 + 3^2} = 5$ miles per hour and so he travels $5 \times \frac{1}{6} = \frac{5}{6}$ miles. Now to find the distance downstream he finds himself, note that if x is this distance, x and $1/2$ are two legs of a right triangle whose hypotenuse equals $5/6$ miles. Therefore, by the Pythagorean theorem the distance downstream is

$$\sqrt{(5/6)^2 - (1/2)^2} = \frac{2}{3} \text{ miles.}$$

15.4 Exercises

1. The wind blows from West to East at a speed of 50 miles per hour and an airplane which travels at 300 miles per hour in still air is heading North West. What is the velocity of the airplane relative to the ground? What is the component of this velocity in the direction North?
2. In the situation of Problem 1 how many degrees to the West of North should the airplane head in order to fly exactly North. What will be the speed of the airplane relative to the ground?
3. In the situation of 2 suppose the airplane uses 34 gallons of fuel every hour at that air speed and that it needs to fly North a distance of 600 miles. Will the airplane have enough fuel to arrive at its destination given that it has 63 gallons of fuel?

4. An airplane is flying due north at 150 miles per hour. A wind is pushing the airplane due east at 40 miles per hour. After 1 hour, the plane starts flying 30° East of North. Assuming the plane starts at $(0,0)$, where is it after 2 hours? Let North be the direction of the positive y axis and let East be the direction of the positive x axis.
5. City A is located at the origin while city B is located at $(300, 500)$ where distances are in miles. An airplane flies at 250 miles per hour in still air. This airplane wants to fly from city A to city B but the wind is blowing in the direction of the positive y axis at a speed of 50 miles per hour. Find a unit vector such that if the plane heads in this direction, it will end up at city B having flown the shortest possible distance. How long will it take to get there?
6. A certain river is one half mile wide with a current flowing at 2 miles per hour from East to West. A man swims directly toward the opposite shore from the South bank of the river at a speed of 3 miles per hour. How far down the river does he find himself when he has swam across? How far does he end up swimming?
7. A certain river is one half mile wide with a current flowing at 2 miles per hour from East to West. A man can swim at 3 miles per hour in still water. In what direction should he swim in order to travel directly across the river? What would the answer to this problem be if the river flowed at 3 miles per hour and the man could swim only at the rate of 2 miles per hour?
8. Three forces are applied to a point which does not move. Two of the forces are $2\mathbf{i} + \mathbf{j} + 3\mathbf{k}$ Newtons and $\mathbf{i} - 3\mathbf{j} + 2\mathbf{k}$ Newtons. Find the third force.
9. The total force acting on an object is to be $2\mathbf{i} + \mathbf{j} + \mathbf{k}$ Newtons. A force of $-\mathbf{i} + \mathbf{j} + \mathbf{k}$ Newtons is being applied. What other force should be applied to achieve the desired total force?
10. A bird flies from its nest 5 km. in the direction 60° north of east where it stops to rest on a tree. It then flies 10 km. in the direction due southeast and lands atop a telephone pole. Place an xy coordinate system so that the origin is the bird's nest, and the positive x axis points east and the positive y axis points north. Find the displacement vector from the nest to the telephone pole.
11. A car is stuck in the mud. There is a cable stretched tightly from this car to a tree which is 20 feet long. A person grasps the cable in the middle and pulls with a force of 100 pounds perpendicular to the stretched cable. The center of the cable moves two feet and remains still. What is the tension in the cable? The tension in the cable is the force exerted on this point by the part of the cable nearer the car as well as the force exerted on this point by the part of the cable nearer the tree.

Vector Products

16.0.1 Outcomes

1. Evaluate a dot product from the angle formula or the coordinate formula.
2. Interpret the dot product geometrically.
3. Evaluate the following using the dot product:
 - (a) the angle between two vectors
 - (b) the magnitude of a vector
 - (c) the work done by a constant force on an object
4. Evaluate a cross product from the angle formula or the coordinate formula.
5. Interpret the cross product geometrically.
6. Evaluate the following using the cross product:
 - (a) the area of a parallelogram
 - (b) the area of a triangle
 - (c) physical quantities such as the torque and angular velocity.
7. Find the volume of a parallelepiped using the box product.
8. Recall, apply and derive the algebraic properties of the dot and cross products.
9. Use the determinant of the Gramian matrix to compute volumes of k dimensional parallelepipeds.
10. Find the angle between two lines.
11. Determine a point of intersection between a line and a surface.
12. Find the equation of a plane in 3 space given a point and a normal vector, three points, a sketch of a plane or a geometric description of the plane.
13. Determine the normal vector and the intercepts of a given plane.
14. Sketch the graph of a plane given its equation.
15. Determine the cosine of the angle between two planes.
16. Find the equation of a plane determined by lines.
17. Identify standard quadric surfaces given their functions or graphs.
18. Sketch the graph of a quadric surface by identifying the intercepts, traces, sections, symmetry and boundedness or unboundedness of the surface.

16.1 The Dot Product

There are two ways of multiplying vectors which are of great importance in applications. The first of these is called the **dot product**, also called the **scalar product** and sometimes the **inner product**.

Definition 16.1.1 Let $\mathbf{a}, \mathbf{b}$ be two vectors in $\mathbb{R}^n$ define $\mathbf{a} \cdot \mathbf{b}$ as

$$\mathbf{a} \cdot \mathbf{b} \equiv \sum_{k=1}^n a_k b_k.$$

With this definition, there are several important properties satisfied by the dot product. In the statement of these properties, α and β will denote scalars and $\mathbf{a}, \mathbf{b}, \mathbf{c}$ will denote vectors.

Proposition 16.1.2 The dot product satisfies the following properties.

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} \quad (16.1)$$

$$\mathbf{a} \cdot \mathbf{a} \geq 0 \text{ and equals zero if and only if } \mathbf{a} = \mathbf{0} \quad (16.2)$$

$$(\alpha \mathbf{a} + \beta \mathbf{b}) \cdot \mathbf{c} = \alpha (\mathbf{a} \cdot \mathbf{c}) + \beta (\mathbf{b} \cdot \mathbf{c}) \quad (16.3)$$

$$\mathbf{c} \cdot (\alpha \mathbf{a} + \beta \mathbf{b}) = \alpha (\mathbf{c} \cdot \mathbf{a}) + \beta (\mathbf{c} \cdot \mathbf{b}) \quad (16.4)$$

$$|\mathbf{a}|^2 = \mathbf{a} \cdot \mathbf{a} \quad (16.5)$$

You should verify these properties. Also be sure you understand that 16.4 follows from the first three and is therefore redundant. It is listed here for the sake of convenience.

Example 16.1.3 Find $(1, 2, 0, -1) \cdot (0, 1, 2, 3)$.

This equals $0 + 2 + 0 + -3 = -1$.

Example 16.1.4 Find the magnitude of $\mathbf{a} = (2, 1, 4, 2)$. That is, find $|\mathbf{a}|$.

This is $\sqrt{(2, 1, 4, 2) \cdot (2, 1, 4, 2)} = 5$.

The dot product satisfies a fundamental inequality known as the **Cauchy Schwarz inequality**.

Theorem 16.1.5 The dot product satisfies the inequality

$$|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}| |\mathbf{b}|. \quad (16.6)$$

Furthermore equality is obtained if and only if one of $\mathbf{a}$ or $\mathbf{b}$ is a scalar multiple of the other.

Proof: First note that if $\mathbf{b} = \mathbf{0}$ both sides of 16.6 equal zero and so the inequality holds in this case. Therefore, it will be assumed in what follows that $\mathbf{b} \neq \mathbf{0}$.

Define a function of $t \in \mathbb{R}$

$$f(t) = (\mathbf{a} + t\mathbf{b}) \cdot (\mathbf{a} + t\mathbf{b}).$$

Then by 16.2, $f(t) \geq 0$ for all $t \in \mathbb{R}$. Also from 16.3, 16.4, 16.1, and 16.5

$$\begin{aligned} f(t) &= \mathbf{a} \cdot (\mathbf{a} + t\mathbf{b}) + t\mathbf{b} \cdot (\mathbf{a} + t\mathbf{b}) \\ &= \mathbf{a} \cdot \mathbf{a} + t(\mathbf{a} \cdot \mathbf{b}) + t\mathbf{b} \cdot \mathbf{a} + t^2 \mathbf{b} \cdot \mathbf{b} \\ &= |\mathbf{a}|^2 + 2t(\mathbf{a} \cdot \mathbf{b}) + |\mathbf{b}|^2 t^2. \end{aligned}$$

Now

$$\begin{aligned}
 f(t) &= |\mathbf{b}|^2 \left(t^2 + 2t \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2} + \frac{|\mathbf{a}|^2}{|\mathbf{b}|^2} \right) \\
 &= |\mathbf{b}|^2 \left(t^2 + 2t \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2} + \left(\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2} \right)^2 - \left(\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2} \right)^2 + \frac{|\mathbf{a}|^2}{|\mathbf{b}|^2} \right) \\
 &= |\mathbf{b}|^2 \left(\left(t + \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2} \right)^2 + \left(\frac{|\mathbf{a}|^2}{|\mathbf{b}|^2} - \left(\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2} \right)^2 \right) \right) \geq 0
 \end{aligned}$$

for all $t \in \mathbb{R}$. In particular $f(t) \geq 0$ when $t = -(\mathbf{a} \cdot \mathbf{b} / |\mathbf{b}|^2)$ which implies

$$\frac{|\mathbf{a}|^2}{|\mathbf{b}|^2} - \left(\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2} \right)^2 \geq 0. \quad (16.7)$$

Multiplying both sides by $|\mathbf{b}|^4$,

$$|\mathbf{a}|^2 |\mathbf{b}|^2 \geq (\mathbf{a} \cdot \mathbf{b})^2$$

which yields 16.6.

From Theorem 16.1.5, equality holds in 16.6 whenever one of the vectors is a scalar multiple of the other. It only remains to verify this is the only way equality can occur. If either vector equals zero, then equality is obtained in 16.6 so it can be assumed both vectors are non zero and that equality is obtained in 16.7. This implies that $f(t) = 0$ when $t = -(\mathbf{a} \cdot \mathbf{b} / |\mathbf{b}|^2)$ and so from 16.2, it follows that for this value of t , $\mathbf{a} + t\mathbf{b} = \mathbf{0}$ showing $\mathbf{a} = -t\mathbf{b}$. This proves the theorem.

You should note that the entire argument was based only on the properties of the dot product listed in 16.1 - 16.5. This means that whenever something satisfies these properties, the Cauchy Schwartz inequality holds. There are many other instances of these properties besides vectors in $\mathbb{R}^n$.

The Cauchy Schwartz inequality allows a proof of the **triangle inequality** for distances in $\mathbb{R}^n$ in much the same way as the triangle inequality for the absolute value.

Theorem 16.1.6 (*Triangle inequality*) For $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$

$$|\mathbf{a} + \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}| \quad (16.8)$$

and equality holds if and only if one of the vectors is a nonnegative scalar multiple of the other. Also

$$||\mathbf{a}| - |\mathbf{b}|| \leq |\mathbf{a} - \mathbf{b}| \quad (16.9)$$

Proof: By properties of the dot product and the Cauchy Schwartz inequality,

$$\begin{aligned}
 |\mathbf{a} + \mathbf{b}|^2 &= (\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) \\
 &= (\mathbf{a} \cdot \mathbf{a}) + (\mathbf{a} \cdot \mathbf{b}) + (\mathbf{b} \cdot \mathbf{a}) + (\mathbf{b} \cdot \mathbf{b}) \\
 &= |\mathbf{a}|^2 + 2(\mathbf{a} \cdot \mathbf{b}) + |\mathbf{b}|^2 \\
 &\leq |\mathbf{a}|^2 + 2|\mathbf{a} \cdot \mathbf{b}| + |\mathbf{b}|^2 \\
 &\leq |\mathbf{a}|^2 + 2|\mathbf{a}||\mathbf{b}| + |\mathbf{b}|^2 \\
 &= (|\mathbf{a}| + |\mathbf{b}|)^2.
 \end{aligned}$$

Taking square roots of both sides you obtain 16.8.

It remains to consider when equality occurs. If either vector equals zero, then that vector equals zero times the other vector and the claim about when equality occurs is verified. Therefore, it can be assumed both vectors are nonzero. To get equality in the second inequality above, Theorem 16.1.5 implies one of the vectors must be a multiple of the other. Say $\mathbf{b} = \alpha \mathbf{a}$. If $\alpha < 0$ then equality cannot occur in the first inequality because in this case

$$(\mathbf{a} \cdot \mathbf{b}) = \alpha |\mathbf{a}|^2 < 0 < |\alpha| |\mathbf{a}|^2 = |\mathbf{a} \cdot \mathbf{b}|$$

Therefore, $\alpha \geq 0$.

To get the other form of the triangle inequality,

$$\mathbf{a} = \mathbf{a} - \mathbf{b} + \mathbf{b}$$

so

$$\begin{aligned} |\mathbf{a}| &= |\mathbf{a} - \mathbf{b} + \mathbf{b}| \\ &\leq |\mathbf{a} - \mathbf{b}| + |\mathbf{b}|. \end{aligned}$$

Therefore,

$$|\mathbf{a}| - |\mathbf{b}| \leq |\mathbf{a} - \mathbf{b}| \quad (16.10)$$

Similarly,

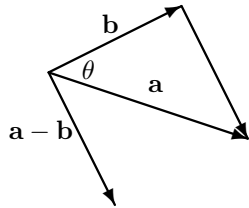
$$|\mathbf{b}| - |\mathbf{a}| \leq |\mathbf{b} - \mathbf{a}| = |\mathbf{a} - \mathbf{b}|. \quad (16.11)$$

It follows from 16.10 and 16.11 that 16.9 holds. This is because $||\mathbf{a}| - |\mathbf{b}||$ equals the left side of either 16.10 or 16.11 and either way, $||\mathbf{a}| - |\mathbf{b}|| \leq |\mathbf{a} - \mathbf{b}|$. This proves the theorem.

16.2 The Geometric Significance Of The Dot Product

16.2.1 The Angle Between Two Vectors

Given two vectors, $\mathbf{a}$ and $\mathbf{b}$, the included angle is the angle between these two vectors which is less than or equal to 180 degrees. The dot product can be used to determine the included angle between two vectors. To see how to do this, consider the following picture.



By the law of cosines,

$$|\mathbf{a} - \mathbf{b}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2|\mathbf{a}||\mathbf{b}|\cos\theta.$$

Also from the properties of the dot product,

$$\begin{aligned} |\mathbf{a} - \mathbf{b}|^2 &= (\mathbf{a} - \mathbf{b}) \cdot (\mathbf{a} - \mathbf{b}) \\ &= |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2\mathbf{a} \cdot \mathbf{b} \end{aligned}$$

and so comparing the above two formulas,

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos \theta. \quad (16.12)$$

In words, the dot product of two vectors equals the product of the magnitude of the two vectors multiplied by the cosine of the included angle. Note this gives a geometric description of the dot product which does not depend explicitly on the coordinates of the vectors.

Example 16.2.1 Find the angle between the vectors $2\mathbf{i} + \mathbf{j} - \mathbf{k}$ and $3\mathbf{i} + 4\mathbf{j} + \mathbf{k}$.

The dot product of these two vectors equals $6 + 4 - 1 = 9$ and the norms are

$$\sqrt{4 + 1 + 1} = \sqrt{6}$$

and $\sqrt{9 + 16 + 1} = \sqrt{26}$. Therefore, from 16.12 the cosine of the included angle equals

$$\cos \theta = \frac{9}{\sqrt{26}\sqrt{6}} = .72058$$

Now the cosine is known, the angle can be determined by solving the equation, $\cos \theta = .72058$. This will involve using a calculator or a table of trigonometric functions. The answer is $\theta = .76616$ radians or in terms of degrees, $\theta = .76616 \times \frac{360}{2\pi} = 43.898^\circ$. Recall how this last computation is done. Set up a proportion, $\frac{x}{.76616} = \frac{360}{2\pi}$ because 360° corresponds to 2π radians. However, in calculus, you should get used to thinking in terms of radians and not degrees. This is because all the important calculus formulas are defined in terms of radians.

Example 16.2.2 Let $\mathbf{u}, \mathbf{v}$ be two vectors whose magnitudes are equal to 3 and 4 respectively and such that if they are placed in standard position with their tails at the origin, the angle between $\mathbf{u}$ and the positive x axis equals 30° and the angle between $\mathbf{v}$ and the positive x axis is -30° . Find $\mathbf{u} \cdot \mathbf{v}$.

From the geometric description of the dot product in 16.12

$$\mathbf{u} \cdot \mathbf{v} = 3 \times 4 \times \cos(60^\circ) = 3 \times 4 \times 1/2 = 6.$$

Observation 16.2.3 Two vectors are said to be **perpendicular** if the included angle is $\pi/2$ radians (90°). You can tell if two nonzero vectors are perpendicular by simply taking their dot product. If the answer is zero, this means they are perpendicular because $\cos \theta = 0$.

Example 16.2.4 Determine whether the two vectors, $2\mathbf{i} + \mathbf{j} - \mathbf{k}$ and $\mathbf{i} + 3\mathbf{j} + 5\mathbf{k}$ are perpendicular.

When you take this dot product you get $2 + 3 - 5 = 0$ and so these two are indeed perpendicular.

Definition 16.2.5 When two lines intersect, the angle between the two lines is the smaller of the two angles determined.

Example 16.2.6 Find the angle between the two lines, $(1, 2, 0) + t(1, 2, 3)$ and $(0, 4, -3) + t(-1, 2, -3)$.

These two lines intersect, when $t = 0$ in the first and $t = -1$ in the second. It is only a matter of finding the angle between the direction vectors. One angle determined is given by

$$\cos \theta = \frac{-6}{14} = \frac{-3}{7}. \quad (16.13)$$

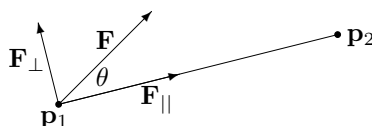
We don't want this angle because it is obtuse. The angle desired is the acute angle given by

$$\cos \theta = \frac{3}{7}.$$

It is obtained by using replacing one of the direction vectors with -1 times it.

16.2.2 Work And Projections

Our first application will be to the concept of work. The physical concept of work does not in any way correspond to the notion of work employed in ordinary conversation. For example, if you were to slide a 150 pound weight off a table which is three feet high and shuffle along the floor for 50 yards, sweating profusely and exerting all your strength to keep the weight from falling on your feet, keeping the height always three feet and then deposit this weight on another three foot high table, the physical concept of work would indicate that the force exerted by your arms did no work during this project even though the muscles in your hands and arms would likely be very tired. The reason for such an unusual definition is that even though your arms exerted considerable force on the weight, enough to keep it from falling, the direction of motion was at right angles to the force they exerted. The only part of a force which does work in the sense of physics is the component of the force in the direction of motion (This is made more precise below.). The work is defined to be the magnitude of the component of this force times the distance over which it acts in the case where this component of force points in the direction of motion and (-1) times the magnitude of this component times the distance in case the force tends to impede the motion. Thus the work done by a force on an object as the object moves from one point to another is a measure of the extent to which the force contributes to the motion. This is illustrated in the following picture in the case where the given force contributes to the motion.



In this picture the force, $\mathbf{F}$ is applied to an object which moves on the straight line from $\mathbf{p}_1$ to $\mathbf{p}_2$. There are two vectors shown, $\mathbf{F}_{||}$ and $\mathbf{F}_{\perp}$ and the picture is intended to indicate that when you add these two vectors you get $\mathbf{F}$ while $\mathbf{F}_{||}$ acts in the direction of motion and $\mathbf{F}_{\perp}$ acts perpendicular to the direction of motion. Only $\mathbf{F}_{||}$ contributes to the work done by $\mathbf{F}$ on the object as it moves from $\mathbf{p}_1$ to $\mathbf{p}_2$. $\mathbf{F}_{||}$ is called the **component of the force** in the direction of motion. From trigonometry, you see the magnitude of $\mathbf{F}_{||}$ should equal $|\mathbf{F}| |\cos \theta|$. Thus, since $\mathbf{F}_{||}$ points in the direction of the vector from $\mathbf{p}_1$ to $\mathbf{p}_2$, the total work done should equal

$$|\mathbf{F}| |\overrightarrow{\mathbf{p}_1 \mathbf{p}_2}| \cos \theta = |\mathbf{F}| |\mathbf{p}_2 - \mathbf{p}_1| \cos \theta$$

If the included angle had been obtuse, then the work done by the force, $\mathbf{F}$ on the object would have been negative because in this case, the force tends to impede the motion from $\mathbf{p}_1$ to $\mathbf{p}_2$ but in this case, $\cos \theta$ would also be negative and so it is still the case

that the work done would be given by the above formula. Thus from the geometric description of the dot product given above, the work equals

$$|\mathbf{F}| |\mathbf{p}_2 - \mathbf{p}_1| \cos \theta = \mathbf{F} \cdot (\mathbf{p}_2 - \mathbf{p}_1).$$

This explains the following definition.

Definition 16.2.7 Let $\mathbf{F}$ be a force acting on an object which moves from the point, $\mathbf{p}_1$ to the point $\mathbf{p}_2$. Then the **work** done on the object by the given force equals $\mathbf{F} \cdot (\mathbf{p}_2 - \mathbf{p}_1)$.

The concept of writing a given vector, $\mathbf{F}$ in terms of two vectors, one which is parallel to a given vector, $\mathbf{D}$ and the other which is perpendicular can also be explained with no reliance on trigonometry, completely in terms of the algebraic properties of the dot product. As before, this is mathematically more significant than any approach involving geometry or trigonometry because it extends to more interesting situations. This is done next.

Theorem 16.2.8 Let $\mathbf{F}$ and $\mathbf{D}$ be nonzero vectors. Then there exist unique vectors $\mathbf{F}_{||}$ and $\mathbf{F}_{\perp}$ such that

$$\mathbf{F} = \mathbf{F}_{||} + \mathbf{F}_{\perp} \quad (16.14)$$

where $\mathbf{F}_{||}$ is a scalar multiple of $\mathbf{D}$, also referred to as

$$\text{proj}_{\mathbf{D}}(\mathbf{F}),$$

and $\mathbf{F}_{\perp} \cdot \mathbf{D} = 0$. The vector $\text{proj}_{\mathbf{D}}(\mathbf{F})$ is called the **projection** of $\mathbf{F}$ onto $\mathbf{D}$.

Proof: Suppose 16.14 and $\mathbf{F}_{||} = \alpha \mathbf{D}$. Taking the dot product of both sides with $\mathbf{D}$ and using $\mathbf{F}_{\perp} \cdot \mathbf{D} = 0$, this yields

$$\mathbf{F} \cdot \mathbf{D} = \alpha |\mathbf{D}|^2$$

which requires $\alpha = \mathbf{F} \cdot \mathbf{D} / |\mathbf{D}|^2$. Thus there can be no more than one vector, $\mathbf{F}_{||}$. It follows $\mathbf{F}_{\perp}$ must equal $\mathbf{F} - \mathbf{F}_{||}$. This verifies there can be no more than one choice for both $\mathbf{F}_{||}$ and $\mathbf{F}_{\perp}$.

Now let

$$\mathbf{F}_{||} \equiv \frac{\mathbf{F} \cdot \mathbf{D}}{|\mathbf{D}|^2} \mathbf{D}$$

and let

$$\mathbf{F}_{\perp} = \mathbf{F} - \mathbf{F}_{||} = \mathbf{F} - \frac{\mathbf{F} \cdot \mathbf{D}}{|\mathbf{D}|^2} \mathbf{D}$$

Then $\mathbf{F}_{||} = \alpha \mathbf{D}$ where $\alpha = \frac{\mathbf{F} \cdot \mathbf{D}}{|\mathbf{D}|^2}$. It only remains to verify $\mathbf{F}_{\perp} \cdot \mathbf{D} = 0$. But

$$\begin{aligned} \mathbf{F}_{\perp} \cdot \mathbf{D} &= \mathbf{F} \cdot \mathbf{D} - \frac{\mathbf{F} \cdot \mathbf{D}}{|\mathbf{D}|^2} \mathbf{D} \cdot \mathbf{D} \\ &= \mathbf{F} \cdot \mathbf{D} - \mathbf{F} \cdot \mathbf{D} = 0. \end{aligned}$$

This proves the theorem.

Example 16.2.9 Let $\mathbf{F} = 2\mathbf{i} + 7\mathbf{j} - 3\mathbf{k}$ Newtons. Find the work done by this force in moving from the point $(1, 2, 3)$ to the point $(-9, -3, 4)$ along the straight line segment joining these points where distances are measured in meters.

According to the definition, this work is

$$\begin{aligned}(2\mathbf{i}+7\mathbf{j}-3\mathbf{k}) \cdot (-10\mathbf{i}-5\mathbf{j}+\mathbf{k}) &= -20 + (-35) + (-3) \\ &= -58 \text{ Newton meters.}\end{aligned}$$

Note that if the force had been given in pounds and the distance had been given in feet, the units on the work would have been foot pounds. In general, work has units equal to units of a force times units of a length. Instead of writing Newton meter, people write joule because a joule is by definition a Newton meter. That word is pronounced “jewel” and it is the unit of work in the metric system of units. Also be sure you observe that the work done by the force can be negative as in the above example. In fact, work can be either positive, negative, or zero. You just have to do the computations to find out.

Example 16.2.10 Find $\text{proj}_{\mathbf{u}}(\mathbf{v})$ if $\mathbf{u} = 2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}$ and $\mathbf{v} = \mathbf{i} - 2\mathbf{j} + \mathbf{k}$.

From the above discussion in Theorem 16.2.8, this is just

$$\begin{aligned}& \frac{1}{4+9+16} (\mathbf{i} - 2\mathbf{j} + \mathbf{k}) \cdot (2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}) (2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}) \\ &= \frac{-8}{29} (2\mathbf{i} + 3\mathbf{j} - 4\mathbf{k}) = -\frac{16}{29}\mathbf{i} - \frac{24}{29}\mathbf{j} + \frac{32}{29}\mathbf{k}.\end{aligned}$$

Example 16.2.11 Suppose $\mathbf{a}$, and $\mathbf{b}$ are vectors and $\mathbf{b}_{\perp} = \mathbf{b} - \text{proj}_{\mathbf{a}}(\mathbf{b})$. What is the magnitude of $\mathbf{b}_{\perp}$ in terms of the included angle?

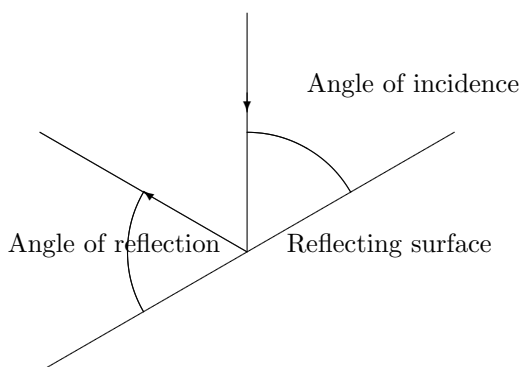
$$\begin{aligned}|\mathbf{b}_{\perp}|^2 &= (\mathbf{b} - \text{proj}_{\mathbf{a}}(\mathbf{b})) \cdot (\mathbf{b} - \text{proj}_{\mathbf{a}}(\mathbf{b})) \\ &= \left(\mathbf{b} - \frac{\mathbf{b} \cdot \mathbf{a}}{|\mathbf{a}|^2} \mathbf{a} \right) \cdot \left(\mathbf{b} - \frac{\mathbf{b} \cdot \mathbf{a}}{|\mathbf{a}|^2} \mathbf{a} \right) \\ &= |\mathbf{b}|^2 - 2 \frac{(\mathbf{b} \cdot \mathbf{a})^2}{|\mathbf{a}|^2} + \left(\frac{\mathbf{b} \cdot \mathbf{a}}{|\mathbf{a}|^2} \right)^2 |\mathbf{a}|^2 \\ &= |\mathbf{b}|^2 \left(1 - \frac{(\mathbf{b} \cdot \mathbf{a})^2}{|\mathbf{a}|^2 |\mathbf{b}|^2} \right) \\ &= |\mathbf{b}|^2 (1 - \cos^2 \theta) = |\mathbf{b}|^2 \sin^2(\theta)\end{aligned}$$

where θ is the included angle between $\mathbf{a}$ and $\mathbf{b}$ which is less than π radians. Therefore, taking square roots,

$$|\mathbf{b}_{\perp}| = |\mathbf{b}| \sin \theta.$$

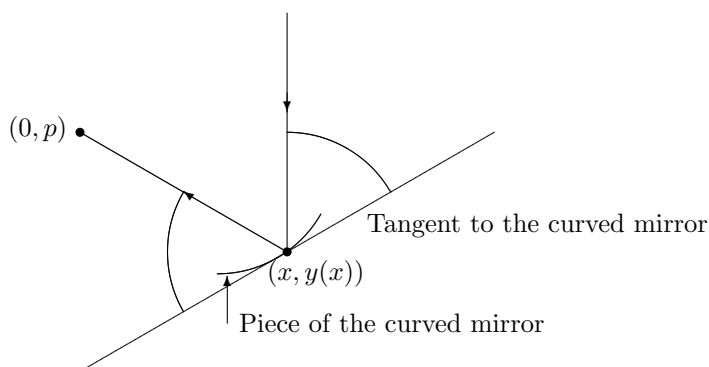
16.2.3 The Parabolic Mirror, An Application

When light is reflected the angle of incidence is always equal to the angle of reflection. This is illustrated in the following picture in which a ray of light reflects off something like a mirror.



An interesting problem is to design a curved mirror which has the property that it will direct all rays of light coming from a long distance away (essentially parallel rays of light) to a single point. You might be interested in a reflecting telescope for example or some sort of scheme for achieving high temperatures by reflecting the rays of the sun to a small area. Turning things around, you could place a source of light at the single point and desire to have the mirror reflect this in a beam of light consisting of parallel rays. How can you design such a mirror?

It turns out this is pretty easy given the above techniques for finding the angle between vectors. Consider the following picture.



It suffices to consider this in a plane for $x > 0$ and then let the mirror be obtained as a surface of revolution. In the above picture, let $(0, p)$ be the special point at which all the parallel rays of light will be directed. This is set up so the rays of light are parallel to the y axis. The two indicated angles will be equal and the equation of the indicated curve will be $y = y(x)$ while the reflection is taking place at the point $(x, y(x))$ as shown. To say the two angles are equal is to say their cosines are equal. Thus from the above,

$$\frac{(0, 1) \cdot (1, y'(x))}{\sqrt{1 + y'(x)^2}} = \frac{(-x, p - y) \cdot (-1, -y'(x))}{\sqrt{x^2 + (y - p)^2} \sqrt{1 + y'(x)^2}}.$$

This follows because the vectors forming the sides of one of the angles are $(0, 1)$ and $(1, y'(x))$ while the vectors forming the other angle are $(-x, p - y)$ and $(-1, -y'(x))$. Therefore, this yields the differential equation,

$$y'(x) = \frac{-y'(x)(p - y) + x}{\sqrt{x^2 + (y - p)^2}}$$

which is written more simply as

$$\left(\sqrt{x^2 + (y-p)^2} + (p-y) \right) y' = x$$

Now let $y-p = xv$ so that $y' = xv' + v$. Then in terms of v the differential equation is

$$xv' = \frac{1}{\sqrt{1+v^2}-v} - v.$$

This reduces to

$$\left(\frac{1}{\sqrt{1+v^2}-v} - v \right) \frac{dv}{dx} = \frac{1}{x}.$$

If $G \in \int \left(\frac{1}{\sqrt{1+v^2}-v} - v \right) dv$, then a solution to the differential equation is of the form

$$G(v) - \ln x = C$$

where C is a constant. This is because if you differentiate both sides with respect to x ,

$$G'(v) \frac{dv}{dx} - \frac{1}{x} = \left(\frac{1}{\sqrt{1+v^2}-v} - v \right) \frac{dv}{dx} - \frac{1}{x} = 0.$$

To find $G \in \int \left(\frac{1}{\sqrt{1+v^2}-v} - v \right) dv$, use a trig. substitution, $v = \tan \theta$. Then in terms of θ , the antiderivative becomes

$$\begin{aligned} \int \left(\frac{1}{\sec \theta - \tan \theta} - \tan \theta \right) \sec^2 \theta d\theta &= \int \sec \theta d\theta \\ &= \ln |\sec \theta + \tan \theta| + C. \end{aligned}$$

Now in terms of v , this is

$$\ln \left(v + \sqrt{1+v^2} \right) = \ln x + c.$$

There is no loss of generality in letting $c = \ln C$ because $\ln$ maps onto $\mathbb{R}$. Therefore, from laws of logarithms,

$$\begin{aligned} \ln \left| v + \sqrt{1+v^2} \right| &= \ln x + c = \ln x + \ln C \\ &= \ln Cx. \end{aligned}$$

Therefore,

$$v + \sqrt{1+v^2} = Cx$$

and so

$$\sqrt{1+v^2} = Cx - v.$$

Now square both sides to get

$$1 + v^2 = C^2 x^2 + v^2 - 2C xv$$

which shows

$$1 = C^2 x^2 - 2Cx \frac{y-p}{x} = C^2 x^2 - 2C(y-p).$$

Solving this for y yields

$$y = \frac{C}{2} x^2 + \left(p - \frac{1}{2C} \right)$$

and for this to correspond to reflection as described above, it must be that $C > 0$. As described in an earlier section, this is just the equation of a parabola. Note it is possible to choose C as desired adjusting the shape of the mirror.

16.2.4 The Dot Product And Distance In $\mathbb{C}^n$

It is necessary to give a generalization of the dot product for vectors in $\mathbb{C}^n$. This definition reduces to the usual one in the case the components of the vector are real.

Definition 16.2.12 Let $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$. Thus $\mathbf{x} = (x_1, \dots, x_n)$ where each $x_k \in \mathbb{C}$ and a similar formula holding for $\mathbf{y}$. Then the dot product of these two vectors is defined to be

$$\mathbf{x} \cdot \mathbf{y} \equiv \sum_j x_j \overline{y_j} \equiv x_1 \overline{y_1} + \dots + x_n \overline{y_n}.$$

Notice how you put the conjugate on the entries of the vector, $\mathbf{y}$. It makes no difference if the vectors happen to be real vectors but with complex vectors you must do it this way. The reason for this is that when you take the dot product of a vector with itself, you want to get the square of the length of the vector, a positive number. Placing the conjugate on the components of $\mathbf{y}$ in the above definition assures this will take place. Thus

$$\mathbf{x} \cdot \mathbf{x} = \sum_j x_j \overline{x_j} = \sum_j |x_j|^2 \geq 0.$$

If you didn't place a conjugate as in the above definition, things wouldn't work out correctly. For example,

$$(1 + i)^2 + 2^2 = 4 + 2i$$

and this is not a positive number.

The following properties of the dot product follow immediately from the definition and you should verify each of them.

Properties of the dot product:

1. $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$.
2. If a, b are numbers and $\mathbf{u}, \mathbf{v}, \mathbf{z}$ are vectors then $(a\mathbf{u} + b\mathbf{v}) \cdot \mathbf{z} = a(\mathbf{u} \cdot \mathbf{z}) + b(\mathbf{v} \cdot \mathbf{z})$.
3. $\mathbf{u} \cdot \mathbf{u} \geq 0$ and it equals 0 if and only if $\mathbf{u} = \mathbf{0}$.

The norm is defined in the usual way.

Definition 16.2.13 For $\mathbf{x} \in \mathbb{C}^n$,

$$|\mathbf{x}| \equiv \left(\sum_{k=1}^n |x_k|^2 \right)^{1/2} = (\mathbf{x} \cdot \mathbf{x})^{1/2}$$

Here is a fundamental inequality called the **Cauchy Schwarz inequality** which is stated here in $\mathbb{C}^n$. First here is a simple lemma.

Lemma 16.2.14 If $z \in \mathbb{C}$ there exists $\theta \in \mathbb{C}$ such that $\theta z = |z|$ and $|\theta| = 1$.

Proof: Let $\theta = 1$ if $z = 0$ and otherwise, let $\theta = \frac{\overline{z}}{|z|}$. Recall that for $z = x + iy$, $\overline{z} = x - iy$ and $\overline{z}z = |z|^2$.

Theorem 16.2.15 (Cauchy Schwarz) The following inequality holds for x_i and $y_i \in \mathbb{C}$.

$$|(\mathbf{x} \cdot \mathbf{y})| = \left| \sum_{i=1}^n x_i \overline{y_i} \right| \leq \left(\sum_{i=1}^n |x_i|^2 \right)^{1/2} \left(\sum_{i=1}^n |y_i|^2 \right)^{1/2} = |\mathbf{x}| |\mathbf{y}| \quad (16.15)$$

Proof: Let $\theta \in \mathbb{C}$ such that $|\theta| = 1$ and

$$\theta \sum_{i=1}^n x_i \bar{y}_i = \left| \sum_{i=1}^n x_i \bar{y}_i \right|$$

Thus

$$\theta \sum_{i=1}^n x_i \bar{y}_i = \sum_{i=1}^n x_i \overline{(\theta y_i)} = \left| \sum_{i=1}^n x_i \bar{y}_i \right|.$$

Consider $p(t) \equiv \sum_{i=1}^n (x_i + t\bar{\theta}y_i) \overline{(x_i + t\bar{\theta}y_i)}$ where $t \in \mathbb{R}$.

$$\begin{aligned} 0 &\leq p(t) = \sum_{i=1}^n |x_i|^2 + 2t \operatorname{Re} \left(\theta \sum_{i=1}^n x_i \bar{y}_i \right) + t^2 \sum_{i=1}^n |y_i|^2 \\ &= |\mathbf{x}|^2 + 2t \left| \sum_{i=1}^n x_i \bar{y}_i \right| + t^2 |\mathbf{y}|^2 \end{aligned}$$

If $|\mathbf{y}| = 0$ then 16.15 is obviously true because both sides equal zero. Therefore, assume $|\mathbf{y}| \neq 0$ and then $p(t)$ is a polynomial of degree two whose graph opens up. Therefore, it either has no zeroes, two zeros or one repeated zero. If it has two zeros, the above inequality must be violated because in this case the graph must dip below the x axis. Therefore, it either has no zeros or exactly one. From the quadratic formula this happens exactly when

$$4 \left| \sum_{i=1}^n x_i \bar{y}_i \right|^2 - 4 |\mathbf{x}|^2 |\mathbf{y}|^2 \leq 0$$

and so

$$\left| \sum_{i=1}^n x_i \bar{y}_i \right| \leq |\mathbf{x}| |\mathbf{y}|$$

as claimed. This proves the inequality.

By analogy to the case of $\mathbb{R}^n$, length or magnitude of vectors in $\mathbb{C}^n$ can be defined.

Definition 16.2.16 Let $\mathbf{z} \in \mathbb{C}^n$. Then $|\mathbf{z}| \equiv (\mathbf{z} \cdot \mathbf{z})^{1/2}$.

Theorem 16.2.17 For length defined in Definition 16.2.16, the following hold.

$$|\mathbf{z}| \geq 0 \text{ and } |\mathbf{z}| = 0 \text{ if and only if } \mathbf{z} = \mathbf{0} \quad (16.16)$$

$$\text{If } \alpha \text{ is a scalar, } |\alpha \mathbf{z}| = |\alpha| |\mathbf{z}| \quad (16.17)$$

$$|\mathbf{z} + \mathbf{w}| \leq |\mathbf{z}| + |\mathbf{w}|. \quad (16.18)$$

Proof: The first two claims are left as exercises. To establish the third, you use the same argument which was used in $\mathbb{R}^n$.

$$\begin{aligned} |\mathbf{z} + \mathbf{w}|^2 &= (\mathbf{z} + \mathbf{w}, \mathbf{z} + \mathbf{w}) \\ &= \mathbf{z} \cdot \mathbf{z} + \mathbf{w} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{z} + \mathbf{z} \cdot \mathbf{w} \\ &= |\mathbf{z}|^2 + |\mathbf{w}|^2 + 2 \operatorname{Re} \mathbf{w} \cdot \mathbf{z} \\ &\leq |\mathbf{z}|^2 + |\mathbf{w}|^2 + 2 |\mathbf{w}| |\mathbf{z}| \\ &\leq |\mathbf{z}|^2 + |\mathbf{w}|^2 + 2 |\mathbf{w}| |\mathbf{z}| = (|\mathbf{z}| + |\mathbf{w}|)^2. \end{aligned}$$

All other considerations such as open and closed sets and the like are identical in this more general context with the corresponding definition in $\mathbb{R}^n$. The main difference is that here the scalars are complex numbers.

Definition 16.2.18 Suppose you have a vector space, V and for $\mathbf{z}, \mathbf{w} \in V$ and α a scalar a norm is a way of measuring distance or magnitude which satisfies the properties 16.16 - 16.18. Thus a norm is something which does the following.

$$\|\mathbf{z}\| \geq 0 \text{ and } \|\mathbf{z}\| = 0 \text{ if and only if } \mathbf{z} = \mathbf{0} \quad (16.19)$$

$$\text{If } \alpha \text{ is a scalar, } \|\alpha\mathbf{z}\| = |\alpha| \|\mathbf{z}\| \quad (16.20)$$

$$\|\mathbf{z} + \mathbf{w}\| \leq \|\mathbf{z}\| + \|\mathbf{w}\|. \quad (16.21)$$

Here is understood that for all $\mathbf{z} \in V$, $\|\mathbf{z}\| \in [0, \infty)$.

16.3 Exercises

1. Find $(1, 2, 3, 4) \cdot (2, 0, 1, 3)$.
2. Use formula 16.12 to verify the Cauchy Schwartz inequality and to show that equality occurs if and only if one of the vectors is a scalar multiple of the other.
3. For $\mathbf{u}, \mathbf{v}$ vectors in $\mathbb{R}^3$, define the product, $\mathbf{u} * \mathbf{v} \equiv u_1v_1 + 2u_2v_2 + 3u_3v_3$. Show the axioms for a dot product all hold for this funny product. Prove

$$|\mathbf{u} * \mathbf{v}| \leq (\mathbf{u} * \mathbf{u})^{1/2} (\mathbf{v} * \mathbf{v})^{1/2}.$$

Hint: Do not try to do this with methods from trigonometry.

4. Find the angle between the vectors $3\mathbf{i} - \mathbf{j} - \mathbf{k}$ and $\mathbf{i} + 4\mathbf{j} + 2\mathbf{k}$.
5. Find the angle between the vectors $\mathbf{i} - 2\mathbf{j} + \mathbf{k}$ and $\mathbf{i} + 2\mathbf{j} - 7\mathbf{k}$.
6. Find $\text{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{v} = (1, 0, -2)$ and $\mathbf{u} = (1, 2, 3)$.
7. Find $\text{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{v} = (1, 2, -2)$ and $\mathbf{u} = (1, 0, 3)$.
8. Find $\text{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{v} = (1, 2, -2, 1)$ and $\mathbf{u} = (1, 2, 3, 0)$.
9. Does it make sense to speak of $\text{proj}_{\mathbf{0}}(\mathbf{v})$?
10. Prove that $T\mathbf{v} \equiv \text{proj}_{\mathbf{u}}(\mathbf{v})$ is a linear transformation and find the matrix of $\text{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{u} = (1, 2, 3)$.
11. If $\mathbf{F}$ is a force and $\mathbf{D}$ is a vector, show $\text{proj}_{\mathbf{D}}(\mathbf{F}) = (|\mathbf{F}| \cos \theta) \mathbf{u}$ where $\mathbf{u}$ is the unit vector in the direction of $\mathbf{D}$, $\mathbf{u} = \mathbf{D}/|\mathbf{D}|$ and θ is the included angle between the two vectors, $\mathbf{F}$ and $\mathbf{D}$. $|\mathbf{F}| \cos \theta$ is sometimes called the component of the force, $\mathbf{F}$ in the direction, $\mathbf{D}$.
12. A boy drags a sled for 100 feet along the ground by pulling on a rope which is 20 degrees from the horizontal with a force of 40 pounds. How much work does this force do?
13. A girl drags a sled for 200 feet along the ground by pulling on a rope which is 30 degrees from the horizontal with a force of 20 pounds. How much work does this force do?
14. A large dog drags a sled for 300 feet along the ground by pulling on a rope which is 45 degrees from the horizontal with a force of 20 pounds. How much work does this force do?

15. How much work in Newton meters does it take to slide a crate 20 meters along a loading dock by pulling on it with a 200 Newton force at an angle of 30° from the horizontal?
16. An object moves 10 meters in the direction of $\mathbf{j}$. There are two forces acting on this object, $\mathbf{F}_1 = \mathbf{i} + \mathbf{j} + 2\mathbf{k}$, and $\mathbf{F}_2 = -5\mathbf{i} + 2\mathbf{j} - 6\mathbf{k}$. Find the total work done on the object by the two forces. **Hint:** You can take the work done by the resultant of the two forces or you can add the work done by each force. Why?
17. An object moves 10 meters in the direction of $\mathbf{j} + \mathbf{i}$. There are two forces acting on this object, $\mathbf{F}_1 = \mathbf{i} + 2\mathbf{j} + 2\mathbf{k}$, and $\mathbf{F}_2 = 5\mathbf{i} + 2\mathbf{j} - 6\mathbf{k}$. Find the total work done on the object by the two forces. **Hint:** You can take the work done by the resultant of the two forces or you can add the work done by each force. Why?
18. An object moves 20 meters in the direction of $\mathbf{k} + \mathbf{j}$. There are two forces acting on this object, $\mathbf{F}_1 = \mathbf{i} + \mathbf{j} + 2\mathbf{k}$, and $\mathbf{F}_2 = \mathbf{i} + 2\mathbf{j} - 6\mathbf{k}$. Find the total work done on the object by the two forces. **Hint:** You can take the work done by the resultant of the two forces or you can add the work done by each force.
19. If $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are vectors. Show that $(\mathbf{b} + \mathbf{c})_\perp = \mathbf{b}_\perp + \mathbf{c}_\perp$ where $\mathbf{b}_\perp = \mathbf{b} - \text{proj}_{\mathbf{a}}(\mathbf{b})$.
20. In the discussion of the reflecting mirror which directs all rays to a particular point, $(0, p)$. Show that for any choice of positive C this point is the focus of the parabola and the directrix is $y = p - \frac{1}{C}$.
21. Suppose you wanted to make a solar powered oven to cook food. Are there reasons for using a mirror which is not parabolic? Also describe how you would design a good flash light with a beam which does not spread out too quickly.
22. Show that $(\mathbf{a} \cdot \mathbf{b}) = \frac{1}{4} [|\mathbf{a} + \mathbf{b}|^2 - |\mathbf{a} - \mathbf{b}|^2]$.
23. Prove from the axioms of the dot product the parallelogram identity, $|\mathbf{a} + \mathbf{b}|^2 + |\mathbf{a} - \mathbf{b}|^2 = 2|\mathbf{a}|^2 + 2|\mathbf{b}|^2$.
24. Let A and be a real $m \times n$ matrix and let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. Show $(A\mathbf{x}, \mathbf{y})_{\mathbb{R}^m} = (\mathbf{x}, A^T \mathbf{y})_{\mathbb{R}^n}$ where $(\cdot, \cdot)_{\mathbb{R}^k}$ denotes the dot product in $\mathbb{R}^k$. In the notation above, $A\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A^T \mathbf{y}$. Use the definition of matrix multiplication to do this.
25. Use the result of Problem 24 to verify directly that $(AB)^T = B^T A^T$ without making any reference to subscripts.
26. Suppose f, g are two continuous functions defined on $[0, 1]$. Define

$$(f \cdot g) = \int_0^1 f(x) g(x) dx.$$

Show this dot product satisfies conditons 16.1 - 16.5. Explain why the Cauchy Schwarz inequality continues to hold in this context and state the Cauchy Schwarz inequality in terms of integrals.

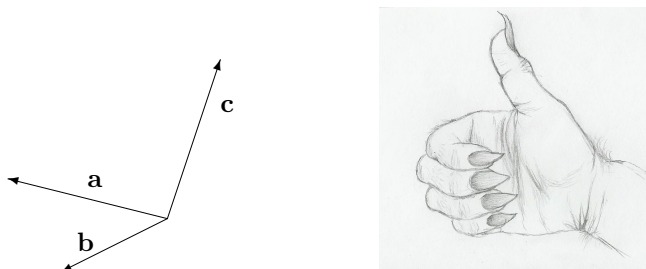
16.4 The Cross Product

The cross product is the other way of multiplying two vectors in $\mathbb{R}^3$. It is very different from the dot product in many ways. First the geometric meaning is discussed and then a description in terms of coordinates is given. Both descriptions of the cross product are

important. The geometric description is essential in order to understand the applications to physics and geometry while the coordinate description is the only way to practically compute the cross product.

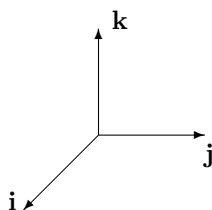
Definition 16.4.1 *Three vectors, $\mathbf{a}, \mathbf{b}, \mathbf{c}$ form a right handed system if when you extend the fingers of your right hand along the vector, $\mathbf{a}$ and close them in the direction of $\mathbf{b}$, the thumb points roughly in the direction of $\mathbf{c}$.*

For an example of a right handed system of vectors, see the following picture.



In this picture the vector $\mathbf{c}$ points upwards from the plane determined by the other two vectors. You should consider how a right hand system would differ from a left hand system. Try using your left hand and you will see that the vector, $\mathbf{c}$ would need to point in the opposite direction as it would for a right hand system.

From now on, the vectors, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ will always form a right handed system. To repeat, if you extend the fingers of our right hand along $\mathbf{i}$ and close them in the direction $\mathbf{j}$, the thumb points in the direction of $\mathbf{k}$.

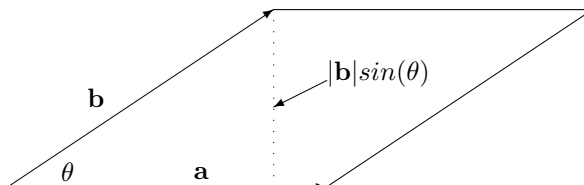


The following is the geometric description of the cross product. It gives both the direction and the magnitude and therefore specifies the vector.

Definition 16.4.2 *Let $\mathbf{a}$ and $\mathbf{b}$ be two vectors in $\mathbb{R}^3$. Then $\mathbf{a} \times \mathbf{b}$ is defined by the following two rules.*

1. $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}| |\mathbf{b}| \sin \theta$ where θ is the included angle.
2. $\mathbf{a} \times \mathbf{b} \cdot \mathbf{a} = 0$, $\mathbf{a} \times \mathbf{b} \cdot \mathbf{b} = 0$, and $\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}$ forms a right hand system.

Note that $|\mathbf{a} \times \mathbf{b}|$ is the area of the parallelogram spanned by $\mathbf{a}$ and $\mathbf{b}$.



The cross product satisfies the following properties.

$$\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a}), \quad \mathbf{a} \times \mathbf{a} = \mathbf{0}, \quad (16.22)$$

For α a scalar,

$$(\alpha \mathbf{a}) \times \mathbf{b} = \alpha (\mathbf{a} \times \mathbf{b}) = \mathbf{a} \times (\alpha \mathbf{b}), \quad (16.23)$$

For $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ vectors, one obtains the distributive laws,

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}, \quad (16.24)$$

$$(\mathbf{b} + \mathbf{c}) \times \mathbf{a} = \mathbf{b} \times \mathbf{a} + \mathbf{c} \times \mathbf{a}. \quad (16.25)$$

Formula 16.22 follows immediately from the definition. The vectors $\mathbf{a} \times \mathbf{b}$ and $\mathbf{b} \times \mathbf{a}$ have the same magnitude, $|\mathbf{a}||\mathbf{b}|\sin\theta$, and an application of the right hand rule shows they have opposite direction. Formula 16.23 is also fairly clear. If α is a nonnegative scalar, the direction of $(\alpha \mathbf{a}) \times \mathbf{b}$ is the same as the direction of $\mathbf{a} \times \mathbf{b}$, $\alpha (\mathbf{a} \times \mathbf{b})$ and $\mathbf{a} \times (\alpha \mathbf{b})$ while the magnitude is just α times the magnitude of $\mathbf{a} \times \mathbf{b}$ which is the same as the magnitude of $\alpha (\mathbf{a} \times \mathbf{b})$ and $\mathbf{a} \times (\alpha \mathbf{b})$. Using this yields equality in 16.23. In the case where $\alpha < 0$, everything works the same way except the vectors are all pointing in the opposite direction and you must multiply by $|\alpha|$ when comparing their magnitudes. The distributive laws are much harder to establish but the second follows from the first quite easily. Thus, assuming the first, and using 16.22,

$$\begin{aligned} (\mathbf{b} + \mathbf{c}) \times \mathbf{a} &= -\mathbf{a} \times (\mathbf{b} + \mathbf{c}) \\ &= -(\mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}) \\ &= \mathbf{b} \times \mathbf{a} + \mathbf{c} \times \mathbf{a}. \end{aligned}$$

A proof of the distributive law is given in a later section for those who are interested. Now from the definition of the cross product,

$$\begin{array}{ll} \mathbf{i} \times \mathbf{j} = \mathbf{k} & \mathbf{j} \times \mathbf{i} = -\mathbf{k} \\ \mathbf{k} \times \mathbf{i} = \mathbf{j} & \mathbf{i} \times \mathbf{k} = -\mathbf{j} \\ \mathbf{j} \times \mathbf{k} = \mathbf{i} & \mathbf{k} \times \mathbf{j} = -\mathbf{i} \end{array}$$

With this information, the following gives the coordinate description of the cross product.

Proposition 16.4.3 *Let $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$ be two vectors. Then*

$$\begin{aligned} \mathbf{a} \times \mathbf{b} &= (a_2b_3 - a_3b_2)\mathbf{i} + (a_3b_1 - a_1b_3)\mathbf{j} + \\ &\quad + (a_1b_2 - a_2b_1)\mathbf{k}. \end{aligned} \quad (16.26)$$

Proof: From the above table and the properties of the cross product listed,

$$\begin{aligned} & (a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}) \times (b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}) = \\ & a_1b_2\mathbf{i} \times \mathbf{j} + a_1b_3\mathbf{i} \times \mathbf{k} + a_2b_1\mathbf{j} \times \mathbf{i} + a_2b_3\mathbf{j} \times \mathbf{k} + \\ & \quad + a_3b_1\mathbf{k} \times \mathbf{i} + a_3b_2\mathbf{k} \times \mathbf{j} \\ & = a_1b_2\mathbf{k} - a_1b_3\mathbf{j} - a_2b_1\mathbf{k} + a_2b_3\mathbf{i} + a_3b_1\mathbf{j} - a_3b_2\mathbf{i} \\ & = (a_2b_3 - a_3b_2)\mathbf{i} + (a_3b_1 - a_1b_3)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k} \end{aligned} \quad (16.27)$$

This proves the proposition.

It is probably impossible for most people to remember 16.26. Fortunately, there is a somewhat easier way to remember it.

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \quad (16.28)$$

where you expand the determinant along the top row. This yields

$$(a_2b_3 - a_3b_2)\mathbf{i} - (a_1b_3 - a_3b_1)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k} \quad (16.29)$$

which is the same as 16.27.

Example 16.4.4 Find $(\mathbf{i} - \mathbf{j} + 2\mathbf{k}) \times (3\mathbf{i} - 2\mathbf{j} + \mathbf{k})$.

Use 16.28 to compute this.

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & -1 & 2 \\ 3 & -2 & 1 \end{vmatrix} = \begin{vmatrix} -1 & 2 \\ -2 & 1 \end{vmatrix} \mathbf{i} - \begin{vmatrix} 1 & 2 \\ 3 & 1 \end{vmatrix} \mathbf{j} + \begin{vmatrix} 1 & -1 \\ 3 & -2 \end{vmatrix} \mathbf{k} \\ = 3\mathbf{i} + 5\mathbf{j} + \mathbf{k}.$$

Example 16.4.5 Find the area of the parallelogram determined by the vectors,

$$(\mathbf{i} - \mathbf{j} + 2\mathbf{k}), (3\mathbf{i} - 2\mathbf{j} + \mathbf{k}).$$

These are the same two vectors in Example 16.4.4.

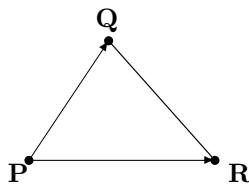
From Example 16.4.4 and the geometric description of the cross product, the area is just the norm of the vector obtained in Example 16.4.4. Thus the area is $\sqrt{9 + 25 + 1} = \sqrt{35}$.

Example 16.4.6 Find the area of the triangle determined by $(1, 2, 3)$, $(0, 2, 5)$, and $(5, 1, 2)$.

This triangle is obtained by connecting the three points with lines. Picking $(1, 2, 3)$ as a starting point, there are two displacement vectors, $(-1, 0, 2)$ and $(4, -1, -1)$ such that the given vector added to these displacement vectors gives the other two vectors. The area of the triangle is half the area of the parallelogram determined by $(-1, 0, 2)$ and $(4, -1, -1)$. Thus $(-1, 0, 2) \times (4, -1, -1) = (2, 7, 1)$ and so the area of the triangle is $\frac{1}{2}\sqrt{4 + 49 + 1} = \frac{3}{2}\sqrt{6}$.

Observation 16.4.7 In general, if you have three points (vectors) in $\mathbb{R}^3$, $\mathbf{P}, \mathbf{Q}, \mathbf{R}$ the area of the triangle is given by

$$\frac{1}{2} |(\mathbf{Q} - \mathbf{P}) \times (\mathbf{R} - \mathbf{P})|.$$

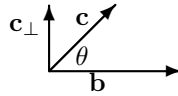


16.4.1 The Distributive Law For The Cross Product

This section gives a proof for 16.24, a fairly difficult topic. It is included here for the interested student. If you are satisfied with taking the distributive law on faith, it is not necessary to read this section. The proof given here is quite clever and follows the one given in [7]. Another approach, based on volumes of parallelepipeds is found in [25] and is discussed a little later.

Lemma 16.4.8 *Let $\mathbf{b}$ and $\mathbf{c}$ be two vectors. Then $\mathbf{b} \times \mathbf{c} = \mathbf{b} \times \mathbf{c}_\perp$ where $\mathbf{c}_\parallel + \mathbf{c}_\perp = \mathbf{c}$ and $\mathbf{c}_\perp \cdot \mathbf{b} = 0$.*

Proof: Consider the following picture.



Now $\mathbf{c}_\perp = \mathbf{c} - \mathbf{c} \cdot \frac{\mathbf{b}}{|\mathbf{b}|} \frac{\mathbf{b}}{|\mathbf{b}|}$ and so $\mathbf{c}_\perp$ is in the plane determined by $\mathbf{c}$ and $\mathbf{b}$. Therefore, from the geometric definition of the cross product, $\mathbf{b} \times \mathbf{c}$ and $\mathbf{b} \times \mathbf{c}_\perp$ have the same direction. Now, referring to the picture,

$$\begin{aligned} |\mathbf{b} \times \mathbf{c}_\perp| &= |\mathbf{b}| |\mathbf{c}_\perp| \\ &= |\mathbf{b}| |\mathbf{c}| \sin \theta \\ &= |\mathbf{b} \times \mathbf{c}|. \end{aligned}$$

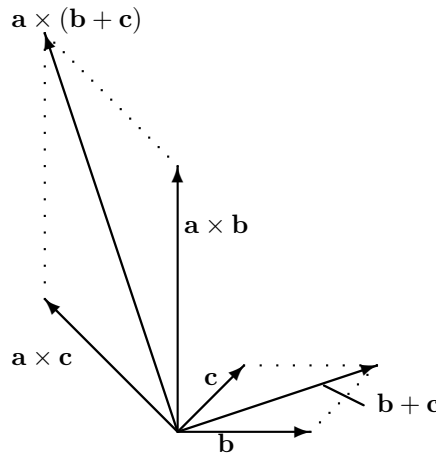
Therefore, $\mathbf{b} \times \mathbf{c}$ and $\mathbf{b} \times \mathbf{c}_\perp$ also have the same magnitude and so they are the same vector.

With this, the proof of the distributive law is in the following theorem.

Theorem 16.4.9 *Let $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ be vectors in $\mathbb{R}^3$. Then*

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c} \quad (16.30)$$

Proof: Suppose first that $\mathbf{a} \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{c} = 0$. Now imagine $\mathbf{a}$ is a vector coming out of the page and let $\mathbf{b}, \mathbf{c}$ and $\mathbf{b} + \mathbf{c}$ be as shown in the following picture.



Then $\mathbf{a} \times \mathbf{b}$, $\mathbf{a} \times (\mathbf{b} + \mathbf{c})$, and $\mathbf{a} \times \mathbf{c}$ are each vectors in the same plane, perpendicular to $\mathbf{a}$ as shown. Thus $\mathbf{a} \times \mathbf{c} \cdot \mathbf{c} = 0$, $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) \cdot (\mathbf{b} + \mathbf{c}) = 0$, and $\mathbf{a} \times \mathbf{b} \cdot \mathbf{b} = 0$. This implies that to get $\mathbf{a} \times \mathbf{b}$ you move counterclockwise through an angle of $\pi/2$ radians from the vector, $\mathbf{b}$. Similar relationships exist between the vectors $\mathbf{a} \times (\mathbf{b} + \mathbf{c})$ and $\mathbf{b} + \mathbf{c}$ and the vectors $\mathbf{a} \times \mathbf{c}$ and $\mathbf{c}$. Thus the angle between $\mathbf{a} \times \mathbf{b}$ and $\mathbf{a} \times (\mathbf{b} + \mathbf{c})$ is the same

as the angle between $\mathbf{b} + \mathbf{c}$ and $\mathbf{b}$ and the angle between $\mathbf{a} \times \mathbf{c}$ and $\mathbf{a} \times (\mathbf{b} + \mathbf{c})$ is the same as the angle between $\mathbf{c}$ and $\mathbf{b} + \mathbf{c}$. In addition to this, since $\mathbf{a}$ is perpendicular to these vectors,

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}| |\mathbf{b}|, |\mathbf{a} \times (\mathbf{b} + \mathbf{c})| = |\mathbf{a}| |\mathbf{b} + \mathbf{c}|, \text{ and}$$

$$|\mathbf{a} \times \mathbf{c}| = |\mathbf{a}| |\mathbf{c}|.$$

Therefore,

$$\frac{|\mathbf{a} \times (\mathbf{b} + \mathbf{c})|}{|\mathbf{b} + \mathbf{c}|} = \frac{|\mathbf{a} \times \mathbf{c}|}{|\mathbf{c}|} = \frac{|\mathbf{a} \times \mathbf{b}|}{|\mathbf{b}|} = |\mathbf{a}|$$

and so

$$\frac{|\mathbf{a} \times (\mathbf{b} + \mathbf{c})|}{|\mathbf{a} \times \mathbf{c}|} = \frac{|\mathbf{b} + \mathbf{c}|}{|\mathbf{c}|}, \frac{|\mathbf{a} \times (\mathbf{b} + \mathbf{c})|}{|\mathbf{a} \times \mathbf{b}|} = \frac{|\mathbf{b} + \mathbf{c}|}{|\mathbf{b}|}$$

showing the triangles making up the parallelogram on the right and the four sided figure on the left in the above picture are similar. It follows the four sided figure on the left is in fact a parallelogram and this implies the diagonal is the vector sum of the vectors on the sides, yielding 16.30.

Now suppose it is not necessarily the case that $\mathbf{a} \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{c} = 0$. Then write $\mathbf{b} = \mathbf{b}_{\parallel} + \mathbf{b}_{\perp}$ where $\mathbf{b}_{\perp} \cdot \mathbf{a} = 0$. Similarly $\mathbf{c} = \mathbf{c}_{\parallel} + \mathbf{c}_{\perp}$. By the above lemma and what was just shown,

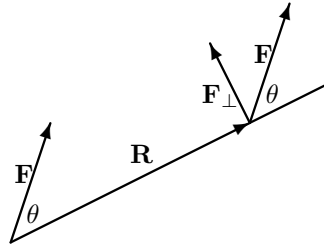
$$\begin{aligned} \mathbf{a} \times (\mathbf{b} + \mathbf{c}) &= \mathbf{a} \times (\mathbf{b} + \mathbf{c})_{\perp} \\ &= \mathbf{a} \times (\mathbf{b}_{\perp} + \mathbf{c}_{\perp}) \\ &= \mathbf{a} \times \mathbf{b}_{\perp} + \mathbf{a} \times \mathbf{c}_{\perp} \\ &= \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}. \end{aligned}$$

This proves the theorem.

The result of Problem 19 of the exercises 16.3 is used to go from the first to the second line.

16.4.2 Torque

Imagine you are using a wrench to loosen a nut. The idea is to turn the nut by applying a force to the end of the wrench. If you push or pull the wrench directly toward or away from the nut, it should be obvious from experience that no progress will be made in turning the nut. The important thing is the component of force perpendicular to the wrench. It is this component of force which will cause the nut to turn. For example see the following picture.



In the picture a force, $\mathbf{F}$ is applied at the end of a wrench represented by the position vector, $\mathbf{R}$ and the angle between these two is θ . Then the tendency to turn will be $|\mathbf{R}| |\mathbf{F}_{\perp}| = |\mathbf{R}| |\mathbf{F}| \sin \theta$, which you recognize as the magnitude of the cross product of $\mathbf{R}$ and $\mathbf{F}$. If there were just one force acting at one point whose position vector is $\mathbf{R}$, perhaps this would be sufficient, but what if there are numerous forces acting at many

different points with neither the position vectors nor the force vectors in the same plane; what then? To keep track of this sort of thing, define for each $\mathbf{R}$ and $\mathbf{F}$, the torque vector,

$$\tau \equiv \mathbf{R} \times \mathbf{F}.$$

This is also called the moment of the force, $\mathbf{F}$. That way, if there are several forces acting at several points, the total torque can be obtained by simply adding up the torques associated with the different forces and positions.

Example 16.4.10 Suppose $\mathbf{R}_1 = 2\mathbf{i} - \mathbf{j} + 3\mathbf{k}$, $\mathbf{R}_2 = \mathbf{i} + 2\mathbf{j} - 6\mathbf{k}$ meters and at the points determined by these vectors there are forces, $\mathbf{F}_1 = \mathbf{i} - \mathbf{j} + 2\mathbf{k}$ and $\mathbf{F}_2 = \mathbf{i} - 5\mathbf{j} + \mathbf{k}$ Newtons respectively. Find the total torque about the origin produced by these forces acting at the given points.

It is necessary to take $\mathbf{R}_1 \times \mathbf{F}_1 + \mathbf{R}_2 \times \mathbf{F}_2$. Thus the total torque equals

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & -1 & 3 \\ 1 & -1 & 2 \end{vmatrix} + \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & -6 \\ 1 & -5 & 1 \end{vmatrix} = -27\mathbf{i} - 8\mathbf{j} - 8\mathbf{k} \text{ Newton meters}$$

Example 16.4.11 Find if possible a single force vector, $\mathbf{F}$ which if applied at the point $\mathbf{i} + \mathbf{j} + \mathbf{k}$ will produce the same torque as the above two forces acting at the given points.

This is fairly routine. The problem is to find $\mathbf{F} = F_1\mathbf{i} + F_2\mathbf{j} + F_3\mathbf{k}$ which produces the above torque vector. Therefore,

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 1 & 1 \\ F_1 & F_2 & F_3 \end{vmatrix} = -27\mathbf{i} - 8\mathbf{j} - 8\mathbf{k}$$

which reduces to $(F_3 - F_2)\mathbf{i} + (F_1 - F_3)\mathbf{j} + (F_2 - F_1)\mathbf{k} = -27\mathbf{i} - 8\mathbf{j} - 8\mathbf{k}$. This amounts to solving the system of three equations in three unknowns, F_1 , F_2 , and F_3 ,

$$\begin{aligned} F_3 - F_2 &= -27 \\ F_1 - F_3 &= -8 \\ F_2 - F_1 &= -8 \end{aligned}$$

However, there is no solution to these three equations. (Why?) Therefore no single force acting at the point $\mathbf{i} + \mathbf{j} + \mathbf{k}$ will produce the given torque.

16.4.3 Center Of Mass

The mass of an object is a measure of how much stuff there is in the object. An object has mass equal to one kilogram, a unit of mass in the metric system, if it would exactly balance a known one kilogram object when placed on a balance. The known object is one kilogram by definition. The mass of an object does not depend on where the balance is used. It would be one kilogram on the moon as well as on the earth. The weight of an object is something else. It is the force exerted on the object by gravity and has magnitude gm where g is a constant called the acceleration of gravity. Thus the weight of a one kilogram object would be different on the moon which has much less gravity, smaller g , than on the earth. An important idea is that of the center of mass. This is the point at which an object will balance no matter how it is turned.

Definition 16.4.12 Let an object consist of p point masses, $m_1, \dots, m_p$ with the position of the k^{th} of these at $\mathbf{R}_k$. The center of mass of this object, $\mathbf{R}_0$ is the point satisfying

$$\sum_{k=1}^p (\mathbf{R}_k - \mathbf{R}_0) \times gm_k \mathbf{u} = \mathbf{0}$$

for all unit vectors, $\mathbf{u}$.

The above definition indicates that no matter how the object is suspended, the total torque on it due to gravity is such that no rotation occurs. Using the properties of the cross product,

$$\left(\sum_{k=1}^p \mathbf{R}_k g m_k - \mathbf{R}_0 \sum_{k=1}^p g m_k \right) \times \mathbf{u} = \mathbf{0} \quad (16.31)$$

for any choice of unit vector, $\mathbf{u}$. You should verify that if $\mathbf{a} \times \mathbf{u} = \mathbf{0}$ for all $\mathbf{u}$, then it must be the case that $\mathbf{a} = \mathbf{0}$. Then the above formula requires that

$$\sum_{k=1}^p \mathbf{R}_k g m_k - \mathbf{R}_0 \sum_{k=1}^p g m_k = \mathbf{0}.$$

dividing by g , and then by $\sum_{k=1}^p m_k$,

$$\mathbf{R}_0 = \frac{\sum_{k=1}^p \mathbf{R}_k m_k}{\sum_{k=1}^p m_k}. \quad (16.32)$$

This is the formula for the center of mass of a collection of point masses. To consider the center of mass of a solid consisting of continuously distributed masses, you need the methods of calculus.

Example 16.4.13 Let $m_1 = 5$, $m_2 = 6$, and $m_3 = 3$ where the masses are in kilograms. Suppose m_1 is located at $2\mathbf{i} + 3\mathbf{j} + \mathbf{k}$, m_2 is located at $\mathbf{i} - 3\mathbf{j} + 2\mathbf{k}$ and m_3 is located at $2\mathbf{i} - \mathbf{j} + 3\mathbf{k}$. Find the center of mass of these three masses.

Using 16.32

$$\begin{aligned} \mathbf{R}_0 &= \frac{5(2\mathbf{i} + 3\mathbf{j} + \mathbf{k}) + 6(\mathbf{i} - 3\mathbf{j} + 2\mathbf{k}) + 3(2\mathbf{i} - \mathbf{j} + 3\mathbf{k})}{5 + 6 + 3} \\ &= \frac{11}{7}\mathbf{i} - \frac{3}{7}\mathbf{j} + \frac{13}{7}\mathbf{k} \end{aligned}$$

16.4.4 Angular Velocity

Definition 16.4.14 In a rotating body, a vector, $\boldsymbol{\Omega}$ is called an **angular velocity vector** if the velocity of a point having position vector, $\mathbf{u}$ relative to the body is given by $\boldsymbol{\Omega} \times \mathbf{u}$.

The existence of an angular velocity vector is the key to understanding motion in a moving system of coordinates. It is used to explain the motion on the surface of the rotating earth. For example, have you ever wondered why low pressure areas rotate counter clockwise in the upper hemisphere but clockwise in the lower hemisphere? To quantify these things, you will need the concept of an angular velocity vector. Details are presented later for interesting examples. Here is a simple example. In the above example, think of a coordinate system fixed in the rotating body. Thus if you were riding on the rotating body, you would observe this coordinate system as fixed but it is not fixed.

Example 16.4.15 A wheel rotates counter clockwise about the vector $\mathbf{i} + \mathbf{j} + \mathbf{k}$ at 60 revolutions per minute. This means that if the thumb of your right hand were to point in the direction of $\mathbf{i} + \mathbf{j} + \mathbf{k}$ your fingers of this hand would wrap in the direction of rotation. Find the angular velocity vector for this wheel. Assume the unit of distance is meters and the unit of time is minutes.

Let $\omega = 60 \times 2\pi = 120\pi$. This is the number of radians per minute corresponding to 60 revolutions per minute. Then the angular velocity vector is $\frac{120\pi}{\sqrt{3}}(\mathbf{i} + \mathbf{j} + \mathbf{k})$. Note this gives what you would expect in the case the position vector to the point is perpendicular to $\mathbf{i} + \mathbf{j} + \mathbf{k}$ and at a distance of r . This is because of the geometric description of the cross product. The magnitude of the vector is $r120\pi$ meters per minute and corresponds to the speed and an exercise with the right hand shows the direction is correct also. However, if this body is rigid, this will work for every other point in it, even those for which the position vector is not perpendicular to the given vector. A complete analysis of this is given later.

Example 16.4.16 *A wheel rotates counter clockwise about the vector $\mathbf{i} + \mathbf{j} + \mathbf{k}$ at 60 revolutions per minute exactly as in Example 16.4.15. Let $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ denote an orthogonal right handed system attached to the rotating wheel in which $\mathbf{u}_3 = \frac{1}{\sqrt{3}}(\mathbf{i} + \mathbf{j} + \mathbf{k})$. Thus $\mathbf{u}_1$ and $\mathbf{u}_2$ depend on time. Find the velocity of the point of the wheel located at the point $2\mathbf{u}_1 + 3\mathbf{u}_2 - \mathbf{u}_3$. Note this point is not fixed in space. It is moving.*

Since $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a right handed system like $\mathbf{i}, \mathbf{j}, \mathbf{k}$, everything applies to this system in the same way as with $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Thus the cross product is given by

$$\begin{aligned} & (a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{u}_3) \times (d\mathbf{u}_1 + e\mathbf{u}_2 + f\mathbf{u}_3) \\ &= \begin{vmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \\ a & b & c \\ d & e & f \end{vmatrix} \end{aligned}$$

Therefore, in terms of the given vectors $\mathbf{u}_i$, the angular velocity vector is

$$120\pi\mathbf{u}_3$$

the velocity of the given point is

$$\begin{aligned} & \begin{vmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \\ 0 & 0 & 120\pi \\ 2 & 3 & -1 \end{vmatrix} \\ &= -360\pi\mathbf{u}_1 + 240\pi\mathbf{u}_2 \end{aligned}$$

in meters per minute. Note how this gives the answer in terms of these vectors which are fixed in the body, not in space. Since $\mathbf{u}_i$ depends on t , this shows the answer in this case does also. Of course this is right. Just think of what is going on with the wheel rotating. Those vectors which are fixed in the wheel are moving in space. The velocity of a point in the wheel should be constantly changing. However, its speed will not change. The speed will be the magnitude of the velocity and this is

$$\sqrt{(-360\pi\mathbf{u}_1 + 240\pi\mathbf{u}_2) \cdot (-360\pi\mathbf{u}_1 + 240\pi\mathbf{u}_2)}$$

which from the properties of the dot product equals

$$\sqrt{(-360\pi)^2 + (240\pi)^2} = 120\sqrt{13}\pi$$

because the $\mathbf{u}_i$ are given to be orthogonal.

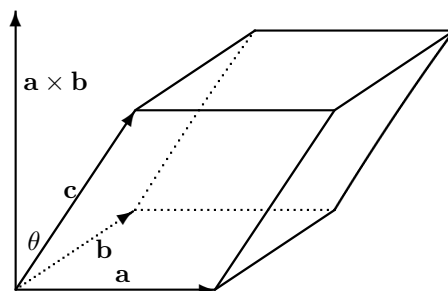
16.4.5 The Box Product

Definition 16.4.17 *A parallelepiped determined by the three vectors, $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ consists of*

$$\{\mathbf{r}\mathbf{a} + \mathbf{s}\mathbf{b} + \mathbf{t}\mathbf{c} : r, s, t \in [0, 1]\}.$$

That is, if you pick three numbers, r, s , and t each in $[0, 1]$ and form $r\mathbf{a} + s\mathbf{b} + t\mathbf{c}$, then the collection of all such points is what is meant by the parallelepiped determined by these three vectors.

The following is a picture of such a thing.



You notice the area of the base of the parallelepiped, the parallelogram determined by the vectors, $\mathbf{a}$ and $\mathbf{b}$ has area equal to $|\mathbf{a} \times \mathbf{b}|$ while the altitude of the parallelepiped is $|\mathbf{c}| \cos \theta$ where θ is the angle shown in the picture between $\mathbf{c}$ and $\mathbf{a} \times \mathbf{b}$. Therefore, the volume of this parallelepiped is the area of the base times the altitude which is just

$$|\mathbf{a} \times \mathbf{b}| |\mathbf{c}| \cos \theta = \mathbf{a} \times \mathbf{b} \cdot \mathbf{c}.$$

This expression is known as the box product and is sometimes written as $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$. You should consider what happens if you interchange the $\mathbf{b}$ with the $\mathbf{c}$ or the $\mathbf{a}$ with the $\mathbf{c}$. You can see geometrically from drawing pictures that this merely introduces a minus sign. In any case the box product of three vectors always equals either the volume of the parallelepiped determined by the three vectors or else minus this volume.

Example 16.4.18 Find the volume of the parallelepiped determined by the vectors, $\mathbf{i} + 2\mathbf{j} - 5\mathbf{k}$, $\mathbf{i} + 3\mathbf{j} - 6\mathbf{k}$, $3\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$.

According to the above discussion, pick any two of these, take the cross product and then take the dot product of this with the third of these vectors. The result will be either the desired volume or minus the desired volume.

$$\begin{aligned} (\mathbf{i} + 2\mathbf{j} - 5\mathbf{k}) \times (\mathbf{i} + 3\mathbf{j} - 6\mathbf{k}) &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & -5 \\ 1 & 3 & -6 \end{vmatrix} \\ &= 3\mathbf{i} + \mathbf{j} + \mathbf{k} \end{aligned}$$

Now take the dot product of this vector with the third which yields

$$(3\mathbf{i} + \mathbf{j} + \mathbf{k}) \cdot (3\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}) = 9 + 2 + 3 = 14.$$

This shows the volume of this parallelepiped is 14 cubic units.

There is a fundamental observation which comes directly from the geometric definitions of the cross product and the dot product.

Lemma 16.4.19 Let $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ be vectors. Then $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$.

Proof: This follows from observing that either $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ and $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ both give the volume of the parallelepiped or they both give -1 times the volume.

An Alternate Proof Of The Distributive Law

Here is another proof of the distributive law for the cross product. Let $\mathbf{x}$ be a vector. From the above observation,

$$\begin{aligned}\mathbf{x} \cdot \mathbf{a} \times (\mathbf{b} + \mathbf{c}) &= (\mathbf{x} \times \mathbf{a}) \cdot (\mathbf{b} + \mathbf{c}) \\ &= (\mathbf{x} \times \mathbf{a}) \cdot \mathbf{b} + (\mathbf{x} \times \mathbf{a}) \cdot \mathbf{c} \\ &= \mathbf{x} \cdot \mathbf{a} \times \mathbf{b} + \mathbf{x} \cdot \mathbf{a} \times \mathbf{c} \\ &= \mathbf{x} \cdot (\mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}).\end{aligned}$$

Therefore,

$$\mathbf{x} \cdot [\mathbf{a} \times (\mathbf{b} + \mathbf{c}) - (\mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c})] = 0$$

for all $\mathbf{x}$. In particular, this holds for $\mathbf{x} = \mathbf{a} \times (\mathbf{b} + \mathbf{c}) - (\mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c})$ showing that $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$ and this proves the distributive law for the cross product another way.

Observation 16.4.20 Suppose you have three vectors, $\mathbf{u} = (a, b, c)$, $\mathbf{v} = (d, e, f)$, and $\mathbf{w} = (g, h, i)$. Then $\mathbf{u} \cdot \mathbf{v} \times \mathbf{w}$ is given by the following.

$$\begin{aligned}\mathbf{u} \cdot \mathbf{v} \times \mathbf{w} &= (a, b, c) \cdot \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ d & e & f \\ g & h & i \end{vmatrix} \\ &= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= \det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}.\end{aligned}$$

The message is that to take the box product, you can simply take the determinant of the matrix which results by letting the rows be the rectangular components of the given vectors in the order in which they occur in the box product.

16.5 Vector Identities And Notation

To begin with consider $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ and it is desired to simplify this quantity. It turns out this is an important quantity which comes up in many different contexts. Let $\mathbf{u} = (u_1, u_2, u_3)$ and let $\mathbf{v}$ and $\mathbf{w}$ be defined similarly.

$$\begin{aligned}\mathbf{v} \times \mathbf{w} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \\ &= (v_2 w_3 - v_3 w_2) \mathbf{i} + (w_1 v_3 - v_1 w_3) \mathbf{j} + (v_1 w_2 - v_2 w_1) \mathbf{k}\end{aligned}$$

Next consider $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$ which is given by

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ (v_2 w_3 - v_3 w_2) & (w_1 v_3 - v_1 w_3) & (v_1 w_2 - v_2 w_1) \end{vmatrix}.$$

When you multiply this out, you get

$$\begin{aligned}&\mathbf{i}(u_1 u_2 w_2 + u_3 v_1 w_3 - w_1 u_2 v_2 - u_3 w_1 v_3) + \mathbf{j}(v_2 u_1 w_1 + v_2 w_3 u_3 - w_2 u_1 v_1 - u_3 w_2 v_3) \\ &+ \mathbf{k}(u_1 w_1 v_3 + v_3 w_2 u_2 - u_1 v_1 w_3 - v_2 w_3 u_2)\end{aligned}$$

and if you are clever, you see right away that

$$(\mathbf{i}v_1 + \mathbf{j}v_2 + \mathbf{k}v_3)(u_1w_1 + u_2w_2 + u_3w_3) - (\mathbf{i}w_1 + \mathbf{j}w_2 + \mathbf{k}w_3)(u_1v_1 + u_2v_2 + u_3v_3).$$

Thus

$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = \mathbf{v}(\mathbf{u} \cdot \mathbf{w}) - \mathbf{w}(\mathbf{u} \cdot \mathbf{v}). \quad (16.33)$$

A related formula is

$$\begin{aligned} (\mathbf{u} \times \mathbf{v}) \times \mathbf{w} &= -[\mathbf{w} \times (\mathbf{u} \times \mathbf{v})] \\ &= -[\mathbf{u}(\mathbf{w} \cdot \mathbf{v}) - \mathbf{v}(\mathbf{w} \cdot \mathbf{u})] \\ &= \mathbf{v}(\mathbf{w} \cdot \mathbf{u}) - \mathbf{u}(\mathbf{w} \cdot \mathbf{v}). \end{aligned} \quad (16.34)$$

This derivation is simply wretched and it does nothing for other identities which may arise in applications. Actually, the above two formulas, 16.33 and 16.34 are sufficient for most applications if you are creative in using them, but there is another way. This other way allows you to discover such vector identities as the above without any creativity or any cleverness. Therefore, it is far superior to the above nasty computation. It is a vector identity discovering machine and it is this which is the main topic in what follows.

There are two special symbols, δ_{ij} and ε_{ijk} which are very useful in dealing with vector identities. To begin with, here is the definition of these symbols.

Definition 16.5.1 *The symbol, δ_{ij} , called the Kronecker delta symbol is defined as follows.*

$$\delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

With the Kronecker symbol, i and j can equal any integer in $\{1, 2, \dots, n\}$ for any $n \in \mathbb{N}$.

Definition 16.5.2 *For i, j , and k integers in the set, $\{1, 2, 3\}$, ε_{ijk} is defined as follows.*

$$\varepsilon_{ijk} \equiv \begin{cases} 1 & \text{if } (i, j, k) = (1, 2, 3), (2, 3, 1), \text{ or } (3, 1, 2) \\ -1 & \text{if } (i, j, k) = (2, 1, 3), (1, 3, 2), \text{ or } (3, 2, 1) \\ 0 & \text{if there are any repeated integers} \end{cases}.$$

The subscripts ijk and ij in the above are called indices. A single one is called an index. This symbol, ε_{ijk} is also called the permutation symbol.

The way to think of ε_{ijk} is that $\varepsilon_{123} = 1$ and if you switch any two of the numbers in the list i, j, k , it changes the sign. Thus $\varepsilon_{ijk} = -\varepsilon_{jik}$ and $\varepsilon_{ijk} = -\varepsilon_{kji}$ etc. You should check that this rule reduces to the above definition. For example, it immediately implies that if there is a repeated index, the answer is zero. This follows because $\varepsilon_{iij} = -\varepsilon_{iij}$ and so $\varepsilon_{iij} = 0$.

It is useful to use the Einstein summation convention when dealing with these symbols. Simply stated, the convention is that you sum over the repeated index. Thus $a_i b_i$ means $\sum_i a_i b_i$. Also, $\delta_{ij} x_j$ means $\sum_j \delta_{ij} x_j = x_i$. When you use this convention, there is one very important thing to never forget. It is this: Never have an index be repeated more than once. Thus $a_i b_i$ is all right but $a_{ii} b_i$ is not. The reason for this is that you end up getting confused about what is meant. If you want to write $\sum_i a_i b_i c_i$ it is best to simply use the summation notation. There is a very important reduction identity connecting these two symbols.

Lemma 16.5.3 *The following holds.*

$$\varepsilon_{ijk} \varepsilon_{irs} = (\delta_{jr} \delta_{ks} - \delta_{kr} \delta_{js}).$$

Proof: If $\{j, k\} \neq \{r, s\}$ then every term in the sum on the left must have either ε_{ijk} or ε_{irs} contains a repeated index. Therefore, the left side equals zero. The right side also equals zero in this case. To see this, note that if the two sets are not equal, then there is one of the indices in one of the sets which is not in the other set. For example, it could be that j is not equal to either r or s . Then the right side equals zero.

Therefore, it can be assumed $\{j, k\} = \{r, s\}$. If $i = r$ and $j = s$ for $s \neq r$, then there is exactly one term in the sum on the left and it equals 1. The right also reduces to 1 in this case. If $i = s$ and $j = r$, there is exactly one term in the sum on the left which is nonzero and it must equal -1. The right side also reduces to -1 in this case. If there is a repeated index in $\{j, k\}$, then every term in the sum on the left equals zero. The right also reduces to zero in this case because then $j = k = r = s$ and so the right side becomes $(1)(1) - (-1)(-1) = 0$.

Proposition 16.5.4 *Let $\mathbf{u}, \mathbf{v}$ be vectors in $\mathbb{R}^n$ where the Cartesian coordinates of $\mathbf{u}$ are $(u_1, \dots, u_n)$ and the Cartesian coordinates of $\mathbf{v}$ are $(v_1, \dots, v_n)$. Then $\mathbf{u} \cdot \mathbf{v} = u_i v_i$. If $\mathbf{u}, \mathbf{v}$ are vectors in $\mathbb{R}^3$, then*

$$(\mathbf{u} \times \mathbf{v})_i = \varepsilon_{ijk} u_j v_k.$$

Also, $\delta_{ik} a_k = a_i$.

Proof: The first claim is obvious from the definition of the dot product. The second is verified by simply checking it works. For example,

$$\mathbf{u} \times \mathbf{v} \equiv \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$$

and so

$$(\mathbf{u} \times \mathbf{v})_1 = (u_2 v_3 - u_3 v_2).$$

From the above formula in the proposition,

$$\varepsilon_{1jk} u_j v_k \equiv u_2 v_3 - u_3 v_2,$$

the same thing. The cases for $(\mathbf{u} \times \mathbf{v})_2$ and $(\mathbf{u} \times \mathbf{v})_3$ are verified similarly. The last claim follows directly from the definition.

With this notation, you can easily discover vector identities and simplify expressions which involve the cross product.

Example 16.5.5 *Discover a formula which simplifies $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w}$.*

From the above reduction formula,

$$\begin{aligned} ((\mathbf{u} \times \mathbf{v}) \times \mathbf{w})_i &= \varepsilon_{ijk} (\mathbf{u} \times \mathbf{v})_j w_k \\ &= \varepsilon_{ijk} \varepsilon_{jrs} u_r v_s w_k \\ &= -\varepsilon_{jik} \varepsilon_{jrs} u_r v_s w_k \\ &= -(\delta_{ir} \delta_{ks} - \delta_{is} \delta_{kr}) u_r v_s w_k \\ &= -(u_i v_k w_k - u_k v_i w_k) \\ &= \mathbf{u} \cdot \mathbf{w} v_i - \mathbf{v} \cdot \mathbf{w} u_i \\ &= ((\mathbf{u} \cdot \mathbf{w}) \mathbf{v} - (\mathbf{v} \cdot \mathbf{w}) \mathbf{u})_i. \end{aligned}$$

Since this holds for all i , it follows that

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w}) \mathbf{v} - (\mathbf{v} \cdot \mathbf{w}) \mathbf{u}.$$

This is good notation and it will be used in the rest of the book whenever convenient.

16.6 Volumes

The determinant and the concept of eigenvalues and eigenvectors provide a way to give a unified treatment of the concept of volumes in various dimensions. First here is a useful theorem which is of considerable interest for its own sake.

Theorem 16.6.1 *Let A be an $n \times n$ matrix. Then*

$$\det(A) = \prod_{i=1}^n \lambda_i$$

where λ_i are the eigenvalues of A . In words, the determinant of a matrix equals the product of its eigenvalues.

Proof: The characteristic polynomial is $\det(\lambda I - A) = \prod_{j=1}^n (\lambda - \lambda_j)$ where λ_i are the eigenvalues. This follows from the fundamental theorem of algebra which says every polynomial can be factored. Then, letting $\lambda = 0$ it follows $\det(-A) = (-1)^n \det(A) = \prod_{j=1}^n (0 - \lambda_j) = (-1)^n \prod_{j=1}^n \lambda_j$ and this proves the theorem.

Recall the geometric definition of the cross product of two vectors found on Page 381. As explained there, the magnitude of the cross product of two vectors was the area of the parallelogram determined by the two vectors. There was also a coordinate description of the cross product. In terms of the notation of Proposition 16.5.4 on Page 392 the i^{th} coordinate of the cross product is given by

$$\varepsilon_{ijk} u_j v_k$$

where the two vectors are (u_1, u_2, u_3) and (v_1, v_2, v_3) . Therefore, using the reduction identity of Lemma 16.5.3 on Page 391

$$\begin{aligned} |\mathbf{u} \times \mathbf{v}|^2 &= \varepsilon_{ijk} u_j v_k \varepsilon_{irs} u_r v_s \\ &= (\delta_{jr} \delta_{ks} - \delta_{kr} \delta_{js}) u_j v_k u_r v_s \\ &= u_j v_k u_j v_k - u_j v_k u_k v_j \\ &= (\mathbf{u} \cdot \mathbf{u})(\mathbf{v} \cdot \mathbf{v}) - (\mathbf{u} \cdot \mathbf{v})^2 \end{aligned}$$

which equals

$$\det \begin{pmatrix} \mathbf{u} \cdot \mathbf{u} & \mathbf{u} \cdot \mathbf{v} \\ \mathbf{u} \cdot \mathbf{v} & \mathbf{v} \cdot \mathbf{v} \end{pmatrix}.$$

Now recall the box product and how the box product was $\pm$ the volume of the parallelepiped spanned by the three vectors. From the definition of the box product

$$\begin{aligned} \mathbf{u} \times \mathbf{v} \cdot \mathbf{w} &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \cdot (w_1 \mathbf{i} + w_2 \mathbf{j} + w_3 \mathbf{k}) \\ &= \det \begin{pmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{pmatrix}. \end{aligned}$$

Therefore,

$$|\mathbf{u} \times \mathbf{v}|^2 = \det \begin{pmatrix} w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{pmatrix}^2$$

which from the theory of determinants equals

$$\begin{aligned} & \det \begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix} \det \begin{pmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{pmatrix} = \\ & \det \left(\begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix} \begin{pmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{pmatrix} \right) = \\ & \det \begin{pmatrix} u_1^2 + u_2^2 + u_3^2 & u_1v_1 + u_2v_2 + u_3v_3 & u_1w_1 + u_2w_2 + u_3w_3 \\ u_1v_1 + u_2v_2 + u_3v_3 & v_1^2 + v_2^2 + v_3^2 & v_1w_1 + v_2w_2 + v_3w_3 \\ u_1w_1 + u_2w_2 + u_3w_3 & v_1w_1 + v_2w_2 + v_3w_3 & w_1^2 + w_2^2 + w_3^2 \end{pmatrix} \\ & = \det \begin{pmatrix} \mathbf{u} \cdot \mathbf{u} & \mathbf{u} \cdot \mathbf{v} & \mathbf{u} \cdot \mathbf{w} \\ \mathbf{u} \cdot \mathbf{v} & \mathbf{v} \cdot \mathbf{v} & \mathbf{v} \cdot \mathbf{w} \\ \mathbf{u} \cdot \mathbf{w} & \mathbf{v} \cdot \mathbf{w} & \mathbf{w} \cdot \mathbf{w} \end{pmatrix} \end{aligned}$$

You see there is a definite pattern emerging here. These earlier cases were for a parallelepiped determined by either two or three vectors in $\mathbb{R}^3$. It makes sense to speak of a parallelepiped in any number of dimensions.

Definition 16.6.2 Let $\mathbf{u}_1, \dots, \mathbf{u}_p$ be vectors in $\mathbb{R}^k$. The parallelepiped determined by these vectors will be denoted by $P(\mathbf{u}_1, \dots, \mathbf{u}_p)$ and it is defined as

$$P(\mathbf{u}_1, \dots, \mathbf{u}_p) \equiv \left\{ \sum_{j=1}^p s_j \mathbf{u}_j : s_j \in [0, 1] \right\}.$$

The volume of this parallelepiped is defined as

$$\text{volume of } P(\mathbf{u}_1, \dots, \mathbf{u}_p) \equiv (\det(\mathbf{u}_i \cdot \mathbf{u}_j))^{1/2}.$$

In this definition, $\mathbf{u}_i \cdot \mathbf{u}_j$ is the ij^{th} entry of a $p \times p$ matrix. Note this definition agrees with all earlier notions of area and volume for parallelepipeds and it makes sense in any number of dimensions. However, it is important to verify the above determinant is nonnegative. After all, the above definition requires a square root of this determinant.

Lemma 16.6.3 Let $\mathbf{u}_1, \dots, \mathbf{u}_p$ be vectors in $\mathbb{R}^k$ for some k . Then $\det(\mathbf{u}_i \cdot \mathbf{u}_j) \geq 0$.

Proof: Recall $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w}$. Therefore, in terms of matrix multiplication, the matrix $(\mathbf{u}_i \cdot \mathbf{u}_j)$ is just the following

$$\overbrace{\begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_p^T \end{pmatrix}}^{p \times k} \overbrace{\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_p \end{pmatrix}}^{k \times p}$$

which is of the form

$$U^T U.$$

Now the eigenvalues of the matrix $U^T U$ are all nonnegative. Here is why. Suppose $U^T U \mathbf{x} = \lambda \mathbf{x}$. Then

$$0 \leq \bar{\mathbf{x}}^T U^T U \mathbf{x} = \lambda \bar{\mathbf{x}}^T \mathbf{x} = \lambda \sum_k |x_k|^2.$$

Therefore, from Theorem 16.6.1, $\det(U^T U) \geq 0$ because it is the product of nonnegative numbers. This proves the lemma and shows the definition of volume is well defined.

Note it gives the right answer in the case where all the vectors are perpendicular. Here is why. Suppose $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ are vectors which have the property that $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ if $i \neq j$. Thus $P(\mathbf{u}_1, \dots, \mathbf{u}_p)$ is a box which has all p sides perpendicular. What should its p dimensional volume be? Shouldn't it equal the product of the lengths of the sides? What does $\det(\mathbf{u}_i \cdot \mathbf{u}_j)$ give? The matrix $(\mathbf{u}_i \cdot \mathbf{u}_j)$ is a diagonal matrix having the squares of the magnitudes of the sides down the diagonal. Therefore, $\det(\mathbf{u}_i \cdot \mathbf{u}_j)^{1/2}$ equals the product of the lengths of the sides as it should.

The matrix, $(\mathbf{u}_i \cdot \mathbf{u}_j)$ whose determinant gives the square of the volume of the parallelepiped spanned by the vectors, $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ is called the Gramian matrix and sometimes the metric tensor.

These considerations are of great significance because they allow the computation in a systematic manner of k dimensional volumes of parallelepipeds which happen to be in $\mathbb{R}^n$ for $n \neq k$. Think for example of a plane in $\mathbb{R}^3$ and the problem of finding the area of something on this plane.

Example 16.6.4 Find the equation of the plane containing the three points,

$$(1, 2, 3), (0, 2, 1), (3, 1, 0).$$

These three points determine two vectors, the one from $(0, 2, 1)$ to $(1, 2, 3)$, $\mathbf{i} + 0\mathbf{j} + 2\mathbf{k}$, and the one from $(0, 2, 1)$ to $(3, 1, 0)$, $3\mathbf{i} + (-1)\mathbf{j} + (-1)\mathbf{k}$. If (x, y, z) denotes a point in the plane, then the volume of the parallelepiped spanned by the vector from $(0, 2, 1)$ to (x, y, z) and these other two vectors must be zero. Thus

$$\det \begin{pmatrix} x & y-2 & z-1 \\ 3 & -1 & -1 \\ 1 & 0 & 2 \end{pmatrix} = 0$$

Therefore, $-2x - 7y + 13 + z = 0$ is the equation of the plane. You should check it contains all three points.

Example 16.6.5 In the above example find the two dimensional volume of the parallelogram determined by the two vectors, $\mathbf{i} + 0\mathbf{j} + 2\mathbf{k}$ and $3\mathbf{i} + (-1)\mathbf{j} + (-1)\mathbf{k}$.

This is easy. Just take the square root of the determinant of the Gramian matrix. Thus the area is

$$\sqrt{\det \begin{pmatrix} 5 & 1 \\ 1 & 11 \end{pmatrix}} = 3\sqrt{6}.$$

16.7 Exercises

1. Show that if $\mathbf{a} \times \mathbf{u} = \mathbf{0}$ for all unit vectors, $\mathbf{u}$, then $\mathbf{a} = \mathbf{0}$.
2. If you only assume 16.31 holds for $\mathbf{u} = \mathbf{i}, \mathbf{j}, \mathbf{k}$, show that this implies 16.31 holds for all unit vectors, $\mathbf{u}$.
3. Let $m_1 = 5, m_2 = 1$, and $m_3 = 4$ where the masses are in kilograms and the distance is in meters. Suppose m_1 is located at $2\mathbf{i} - 3\mathbf{j} + \mathbf{k}$, m_2 is located at $\mathbf{i} - 3\mathbf{j} + 6\mathbf{k}$ and m_3 is located at $2\mathbf{i} + \mathbf{j} + 3\mathbf{k}$. Find the center of mass of these three masses.
4. Let $m_1 = 2, m_2 = 3$, and $m_3 = 1$ where the masses are in kilograms and the distance is in meters. Suppose m_1 is located at $2\mathbf{i} - \mathbf{j} + \mathbf{k}$, m_2 is located at $\mathbf{i} - 2\mathbf{j} + \mathbf{k}$ and m_3 is located at $4\mathbf{i} + \mathbf{j} + 3\mathbf{k}$. Find the center of mass of these three masses.
5. Find the angular velocity vector of a rigid body which rotates counter clockwise about the vector $\mathbf{i} - 2\mathbf{j} + \mathbf{k}$

- at 40 revolutions per minute. Assume distance is measured in meters.
6. Let $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ be a right handed system with $\mathbf{u}_3$ pointing in the direction of $\mathbf{i} - 2\mathbf{j} + \mathbf{k}$ and $\mathbf{u}_1$ and $\mathbf{u}_2$ being fixed with the body which is rotating at 40 revolutions per minute. Assuming all distances are in meters, find the constant speed of the point of the body located at $3\mathbf{u}_1 + \mathbf{u}_2 - \mathbf{u}_3$ in meters per minute.
 7. Find the area of the triangle determined by the three points, $(1, 2, 3)$, $(4, 2, 0)$ and $(-3, 2, 1)$.
 8. Find the area of the triangle determined by the three points, $(1, 0, 3)$, $(4, 1, 0)$ and $(-3, 1, 1)$.
 9. Find the area of the triangle determined by the three points, $(1, 2, 3)$, $(2, 3, 4)$ and $(0, 1, 2)$. Did something interesting happen here? What does it mean geometrically?
 10. Find the area of the parallelogram determined by the vectors, $(1, 2, 3)$ and $(3, -2, 1)$.
 11. Find the area of the parallelogram determined by the vectors, $(1, 0, 3)$ and $(4, -2, 1)$.
 12. Find the area of the parallelogram determined by the vectors, $(1, -2, 2)$ and $(3, 1, 1)$.
 13. Find the volume of the parallelepiped determined by the vectors, $\mathbf{i} - 7\mathbf{j} - 5\mathbf{k}$, $\mathbf{i} - 2\mathbf{j} - 6\mathbf{k}$, $3\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$.
 14. Find the volume of the parallelepiped determined by the vectors, $\mathbf{i} + \mathbf{j} - 5\mathbf{k}$, $\mathbf{i} + 5\mathbf{j} - 6\mathbf{k}$, $3\mathbf{i} + \mathbf{j} + 3\mathbf{k}$.
 15. Find the volume of the parallelepiped determined by the vectors, $\mathbf{i} + 6\mathbf{j} + 5\mathbf{k}$, $\mathbf{i} + 5\mathbf{j} - 6\mathbf{k}$, $3\mathbf{i} + \mathbf{j} + \mathbf{k}$.
 16. Suppose $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are three vectors whose components are all integers. Can you conclude the volume of the parallelepiped determined from these three vectors will always be an integer?
 17. What does it mean geometrically if the box product of three vectors gives zero?
 18. Here are three vectors in $\mathbb{R}^4$: $(1, 2, 0, 3)^T$, $(2, 1, -3, 2)^T$, $(0, 0, 1, 2)^T$. Find the volume of the parallelepiped determined by these three vectors.
 19. Here are two vectors in $\mathbb{R}^4$: $(1, 2, 0, 3)^T$, $(2, 1, -3, 2)^T$. Find the volume of the parallelepiped determined by these two vectors.
 20. Here are three vectors in $\mathbb{R}^2$: $(1, 2)^T$, $(2, 1)^T$, $(0, 1)^T$. Find the volume of the parallelepiped determined by these three vectors. Why should this volume equal zero?
 21. If there are $n + 1$ or more vectors in $\mathbb{R}^n$, Lemma 16.6.3 implies the parallelepiped determined by these $n + 1$ vectors must have zero volume. What is the geometric significance of this assertion?
 22. Find the equation of the plane through the three points $(1, 2, 3)$, $(2, -3, 1)$, $(1, 1, 7)$.
 23. It is desired to find an equation of a plane containing the two vectors, $\mathbf{a}$ and $\mathbf{b}$ and the point $\mathbf{0}$. Using Problem 17, show an equation for this plane is

$$\begin{vmatrix} x & y & z \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = 0$$

That is, the set of all (x, y, z) such that the above expression equals zero.
 24. Using the notion of the box product yielding either plus or minus the volume of the parallelepiped determined by the given three vectors, show that

$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$$

In other words, the dot and the cross can be switched as long as the order of the vectors remains the same.
Hint: There are two ways to do this,

by the coordinate description of the dot and cross product and by geometric reasoning.

25. Is $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$? What is the meaning of $\mathbf{a} \times \mathbf{b} \times \mathbf{c}$? Explain. **Hint:** Try $(\mathbf{i} \times \mathbf{j}) \times \mathbf{j}$.

26. Verify directly that the coordinate description of the cross product, $\mathbf{a} \times \mathbf{b}$ has the property that it is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$. Then show by direct computation that this coordinate description satisfies

$$\begin{aligned} |\mathbf{a} \times \mathbf{b}|^2 &= |\mathbf{a}|^2 |\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2 \\ &= |\mathbf{a}|^2 |\mathbf{b}|^2 (1 - \cos^2(\theta)) \end{aligned}$$

where θ is the angle included between the two vectors. Explain why $|\mathbf{a} \times \mathbf{b}|$ has the correct magnitude. All that is missing is the material about the right hand rule. Verify directly from the coordinate description of the cross product that the right thing happens with regards to the vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Next verify that the distributive law holds for the coordinate description of the cross product. This gives another way to approach the cross product. First define it in terms of coordinates and then get the geometric properties from this.

27. Discover a vector identity for $\mathbf{u} \times (\mathbf{v} \times \mathbf{w})$.
28. Discover a vector identity for $(\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{z} \times \mathbf{w})$.

29. Discover a vector identity for $(\mathbf{u} \times \mathbf{v}) \times (\mathbf{z} \times \mathbf{w})$ in terms of box products.

30. Simplify $(\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{v} \times \mathbf{w}) \times (\mathbf{w} \times \mathbf{z})$.

31. Simplify $|\mathbf{u} \times \mathbf{v}|^2 + (\mathbf{u} \cdot \mathbf{v})^2 - |\mathbf{u}|^2 |\mathbf{v}|^2$.

32. Prove that $\varepsilon_{ijk} \varepsilon_{ijr} = 2\delta_{kr}$.

33. If A is a 3×3 matrix such that $A = \begin{pmatrix} \mathbf{u} & \mathbf{v} & \mathbf{w} \end{pmatrix}$ where these are the columns of the matrix, A . Show that $\det(A) = \varepsilon_{ijk} u_i v_j w_k$.

34. If A is a 3×3 matrix, show $\varepsilon_{rps} \det(A) = \varepsilon_{ijk} A_{ri} A_{pj} A_{sk}$.

35. Suppose A is a 3×3 matrix and $\det(A) \neq 0$. Show using 34 and 32 that

$$(A^{-1})_{ks} = \frac{1}{2 \det(A)} \varepsilon_{rps} \varepsilon_{ijk} A_{pj} A_{ri}.$$

36. When you have a rotating rigid body with angular velocity vector, $\boldsymbol{\Omega}$ then the velocity, $\mathbf{u}'$ is given by $\mathbf{u}' = \boldsymbol{\Omega} \times \mathbf{u}$. It turns out that all the usual calculus rules such as the product rule hold. Also, $\mathbf{u}''$ is the acceleration. Show using the product rule that for $\boldsymbol{\Omega}$ a constant vector,

$$\mathbf{u}'' = \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{u}).$$

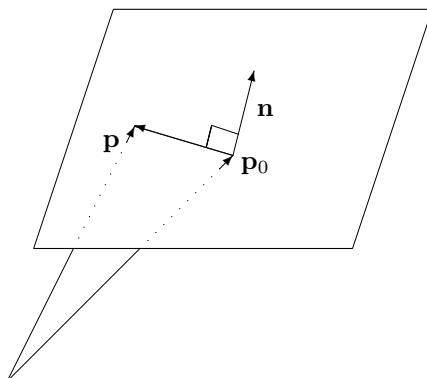
It turns out this is the centripetal acceleration. Note how it involves cross products. Things get really interesting when you move about on the rotating body. weird forces are felt. This is in the section in the appendix on moving coordinate systems.

16.8 Planes

You have an idea of what a plane is already. It is the span of some vectors. However, it can also be considered geometrically in terms of a dot product. To find the equation of a plane, you need two things, a point contained in the plane and a vector normal to the plane. Let $\mathbf{p}_0 = (x_0, y_0, z_0)$ denote the position vector of a point in the plane, let $\mathbf{p} = (x, y, z)$ be the position vector of an arbitrary point in the plane, and let $\mathbf{n}$ denote a vector normal to the plane. This means that

$$\mathbf{n} \cdot (\mathbf{p} - \mathbf{p}_0) = 0$$

whenever $\mathbf{p}$ is the position vector of a point in the plane. The following picture illustrates the geometry of this idea.



Expressed equivalently, the plane is just the set of all points $\mathbf{p}$ such that the vector, $\mathbf{p} - \mathbf{p}_0$ is perpendicular to the given normal vector, $\mathbf{n}$.

Example 16.8.1 Find the equation of the plane with normal vector, $\mathbf{n} = (1, 2, 3)$ containing the point $(2, -1, 5)$.

From the above, the equation of this plane is just

$$(1, 2, 3) \cdot (x - 2, y + 1, z - 3) = x - 9 + 2y + 3z = 0$$

Example 16.8.2 $2x + 4y - 5z = 11$ is the equation of a plane. Find the normal vector and a point on this plane.

You can write this in the form $2(x - \frac{11}{2}) + 4(y - 0) + (-5)(z - 0) = 0$. Therefore, a normal vector to the plane is $2\mathbf{i} + 4\mathbf{j} - 5\mathbf{k}$ and a point in this plane is $(\frac{11}{2}, 0, 0)$. Of course there are many other points in the plane.

Definition 16.8.3 Suppose two planes intersect. The angle between the planes is defined to be the angle between their normal vectors.

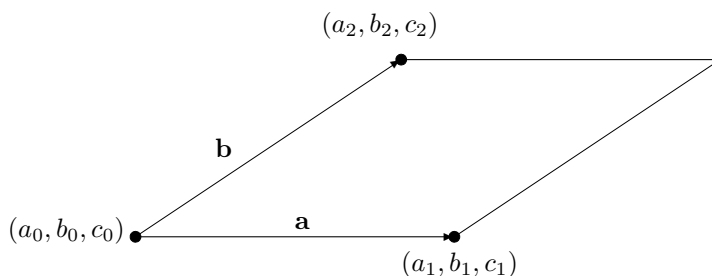
Example 16.8.4 Find the angle between the two planes, $x + 2y - z = 6$ and $3x + 2y - z = 7$.

The two normal vectors are $(1, 2, -1)$ and $(3, 2, -1)$. Therefore, the cosine of the angle desired is

$$\cos \theta = \frac{(1, 2, -1) \cdot (3, 2, -1)}{\sqrt{1^2 + 2^2 + (-1)^2} \sqrt{3^2 + 2^2 + (-1)^2}} = .87287$$

Now use a calculator or table to find what the angle is. $\cos \theta = .87287$, Solution is : $\{\theta = .50974\}$. This value is in radians.

Sometimes you need to find the equation of a plane which contains three points. Consider the following picture.



You have plenty of points but you need a normal. This can be obtained by taking $\mathbf{a} \times \mathbf{b}$ where $\mathbf{a} = (a_1 - a_0, b_1 - b_0, c_1 - c_0)$ and $\mathbf{b} = (a_2 - a_0, b_2 - b_0, c_2 - c_0)$.

Example 16.8.5 Find the equation of the plane which contains the three points,

$$(1, 2, 1), (3, -1, 2), \text{ and } (4, 2, 1).$$

You just need to get a normal vector to this plane. This can be done by taking the cross products of the two vectors,

$$(3, -1, 2) - (1, 2, 1) \text{ and } (4, 2, 1) - (1, 2, 1)$$

Thus a normal vector is $(2, -3, 1) \times (3, 0, 0) = (0, 3, 9)$. Therefore, the equation of the plane is

$$0(x - 1) + 3(y - 2) + 9(z - 1) = 0$$

or $3y + 9z = 15$ which is the same as $y + 3z = 5$. When you have what you think is the plane containing the three points, you ought to check it by seeing if it really does contain the three points.

Example 16.8.6 Find the equation of the plane which contains the three points,

$$(1, 2, 1), (3, -1, 2), \text{ and } (4, 2, 1).$$

You just need to get a normal vector to this plane. This can be done by taking the cross products of the two vectors,

$$(3, -1, 2) - (1, 2, 1) \text{ and } (4, 2, 1) - (1, 2, 1)$$

Thus a normal vector is $(2, -3, 1) \times (3, 0, 0) = (0, 3, 9)$. Therefore, the equation of the plane is

$$0(x - 1) + 3(y - 2) + 9(z - 1) = 0$$

or $3y + 9z = 15$ which is the same as $y + 3z = 5$.

Example 16.8.7 Find the equation of the plane which contains the three points,

$$(1, 2, 1), (3, -1, 2), \text{ and } (4, 2, 1)$$

another way.

Letting (x, y, z) be a point on the plane, the volume of the parallelepiped spanned by $(x, y, z) - (1, 2, 1)$ and the two vectors, $(2, -3, 1)$ and $(3, 0, 0)$ must be equal to zero. Thus the equation of the plane is

$$\det \begin{pmatrix} 3 & 0 & 0 \\ 2 & -3 & 1 \\ x - 1 & y - 2 & z - 1 \end{pmatrix} = 0.$$

Hence $-9z + 15 - 3y = 0$ and dividing by 3 yields the same answer as the above.

Proposition 16.8.8 If $(a, b, c) \neq (0, 0, 0)$, then $ax + by + cz = d$ is the equation of a plane with normal vector $a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$. Conversely, any plane can be written in this form.

Proof: One of a, b, c is nonzero. Suppose for example that $c \neq 0$. Then the equation can be written as

$$a(x - 0) + b(y - 0) + c\left(z - \frac{d}{c}\right) = 0$$

Therefore, $(0, 0, \frac{d}{c})$ is a point on the plane and a normal vector is $a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$. The converse follows from the above discussion involving the point and a normal vector. This proves the proposition.

Example 16.8.9 Find the equation of the plane which contains the three points,

$$(1, 2, 1), (3, -1, 2), \text{ and } (4, 2, 1)$$

another way.

You need to find numbers, a, b, c, d not all zero such that each of the given three points satisfies the equation,

$$ax + by + cz = d.$$

Then you must have for (x, y, z) a point on this plane,

$$\begin{aligned} a + 2b + c - d &= 0, \\ 3a - b + 2c - d &= 0, \\ 4a + 2b + c - d &= 0, \\ xa + yb + zc - d &= 0. \end{aligned}$$

You need a nonzero solution to the above system of four equations for the unknowns, a, b, c , and d . Therefore,

$$\det \begin{pmatrix} 1 & 2 & 1 & -1 \\ 3 & -1 & 2 & -1 \\ 4 & 2 & 1 & -1 \\ x & y & z & -1 \end{pmatrix} = 0$$

because the matrix sends a nonzero vector, $(a, b, c, -d)$ to zero and is therefore, not one to one. Consequently from Theorem 13.2.1 on Page 333, its determinant equals zero. Hence upon evaluating the determinant,

$$-15 + 9z + 3y = 0$$

which reduces to $3z + y = 5$.

Example 16.8.10 Find the equation of the plane containing the points $(1, 2, 3)$ and the line $(0, 1, 1) + t(2, 1, 2) = (x, y, z)$.

There are several ways to do this. One is to find three points and use any of the above procedures. Let $t = 0$ and then let $t = 1$ to get two points on the line. This yields $(1, 2, 3)$, $(0, 1, 1)$, and $(2, 2, 3)$. Then the equation of the plane is

$$\det \begin{pmatrix} x & y & z & -1 \\ 1 & 2 & 3 & -1 \\ 0 & 1 & 1 & -1 \\ 2 & 2 & 3 & -1 \end{pmatrix} = 2y - z - 1 = 0.$$

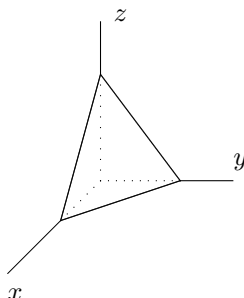
Example 16.8.11 Find the equation of the plane which contains the two lines, given by the following parametric expressions in which $t \in \mathbb{R}$.

$$(2t, 1 + t, 1 + 2t) = (x, y, z), \quad (2t + 2, 1, 3 + 2t) = (x, y, z)$$

Note first that you don't know there even is such a plane. However, if there is, you could find it by obtaining three points, two on one line and one on another and then using any of the above procedures for finding the plane. From the first line, two points are $(0, 1, 1)$ and $(2, 2, 3)$ while a third point can be obtained from second line, $(2, 1, 3)$. You need a normal vector and then use any of these points. To get a normal vector, form $(2, 0, 2) \times (2, 1, 2) = (-2, 0, 2)$. Therefore, the plane is $-2x + 0(y - 1) + 2(z - 1) = 0$. This reduces to $z - x = 1$. If there is a plane, this is it. Now you can simply verify that both of the lines are really in this plane. From the first, $(1 + 2t) - 2t = 1$ and the second, $(3 + 2t) - (2t + 2) = 1$ so both lines lie in the plane.

One way to understand how a plane looks is to connect the points where it intercepts the x , y , and z axes. This allows you to visualize the plane somewhat and is a good way to sketch the plane. Not surprisingly these points are called intercepts.

Example 16.8.12 Sketch the plane which has intercepts $(2, 0, 0)$, $(0, 3, 0)$, and $(0, 0, 4)$.



You see how connecting the intercepts gives a fairly good geometric description of the plane. These lines which connect the intercepts are also called the traces of the plane. Thus the line which joins $(0, 3, 0)$ to $(0, 0, 4)$ is the intersection of the plane with the yz plane. It is the trace on the yz plane.

Example 16.8.13 Identify the intercepts of the plane, $3x - 4y + 5z = 11$.

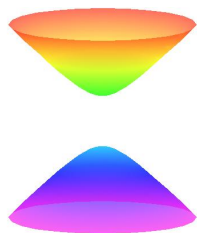
The easy way to do this is to divide both sides by 11.

$$\frac{x}{(11/3)} + \frac{y}{(-11/4)} + \frac{z}{(11/5)} = 1$$

The intercepts are $(11/3, 0, 0)$, $(0, -11/4, 0)$ and $(0, 0, 11/5)$. You can see this by letting both y and z equal to zero to find the point on the x axis which is intersected by the plane. The other axes are handled similarly.

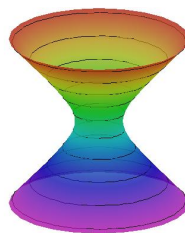
16.9 Quadric Surfaces

In the above it was shown that the equation of an arbitrary plane is an equation of the form $ax + by + cz = d$. Such equations are called level surfaces. There are some standard level surfaces which involve certain variables being raised to a power of 2 which are sufficiently important that they are given names, usually involving the portentous semi-word "oid". These are graphed below using Maple, a computer algebra system.



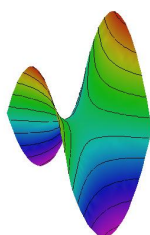
$$z^2/a^2 - x^2/b^2 - y^2/c^2 = 1$$

hyperboloid of two sheets



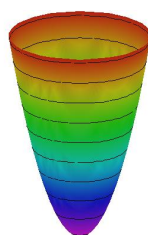
$$x^2/b^2 + y^2/c^2 - z^2/a^2 = 1$$

hyperboloid of one sheet



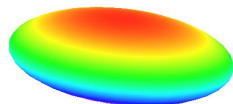
$$z = x^2/a^2 - y^2/b^2$$

hyperbolic paraboloid



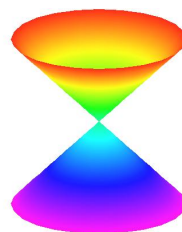
$$z = x^2/a^2 + y^2/b^2$$

elliptic paraboloid



$$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$$

ellipsoid



$$z^2/a^2 = x^2/b^2 + y^2/c^2$$

elliptic cone

Why do the graphs of these level surfaces look the way they do? Consider first the hyperboloid of two sheets. The equation defining this surface can be written in the form

$$\frac{z^2}{a^2} - 1 = \frac{x^2}{b^2} + \frac{y^2}{c^2}.$$

Suppose you fix a value for z . What ordered pairs, (x, y) will satisfy the equation? If $\frac{z^2}{a^2} < 1$, there is no such ordered pair because the above equation would require a negative number to equal a nonnegative one. This is why there is a gap and there are two sheets. If $\frac{z^2}{a^2} > 1$, then the above equation is the equation for an ellipse. That is why if you slice the graph by letting $z = z_0$ the result is an ellipse in the plane $z = z_0$.

Consider the hyperboloid of one sheet.

$$\frac{x^2}{b^2} + \frac{y^2}{c^2} = 1 + \frac{z^2}{a^2}.$$

This time, it doesn't matter what value z takes. The resulting equation for (x, y) is an ellipse.

Similar considerations apply to the elliptic paraboloid as long as $z > 0$ and the ellipsoid. The elliptic cone is like the hyperboloid of two sheets without the 1. Therefore, z can have any value. In case $z = 0$, $(x, y) = (0, 0)$. Viewed from the side, it appears straight, not curved like the hyperboloid of two sheets. This is because if (x, y, z) is a point on the surface, then if t is a scalar, it follows (tx, ty, tz) is also on this surface.

The most interesting of these graphs is the hyperbolic paraboloid¹, $z = \frac{x^2}{a^2} - \frac{y^2}{b^2}$. If $z > 0$ this is the equation of a hyperbola which opens to the right and left while if $z < 0$ it is a hyperbola which opens up and down. As z passes from positive to negative, the hyperbola changes type and this is what yields the shape shown in the picture.

Not surprisingly, you can find intercepts and traces of quadric surfaces just as with planes.

Example 16.9.1 Find the trace on the xy plane of the hyperbolic paraboloid, $z = x^2 - y^2$.

This occurs when $z = 0$ and so this reduces to $y^2 = x^2$. In other words, this trace is just the two straight lines, $y = x$ and $y = -x$.

Example 16.9.2 Find the intercepts of the ellipsoid, $x^2 + 2y^2 + 4z^2 = 9$.

To find the intercept on the x axis, let $y = z = 0$ and this yields $x = \pm 3$. Thus there are two intercepts, $(3, 0, 0)$ and $(-3, 0, 0)$. The other intercepts are left for you to find. You can see this is an aid in graphing the quadric surface. The surface is said to be bounded if there is some number, C such that whenever, (x, y, z) is a point on the surface, $\sqrt{x^2 + y^2 + z^2} < C$. The surface is called unbounded if no such constant, C exists. Ellipsoids are bounded but the other quadric surfaces are not bounded.

Example 16.9.3 Why is the hyperboloid of one sheet, $x^2 + 2y^2 - z^2 = 1$ unbounded?

Let z be very large. Does there correspond (x, y) such that (x, y, z) is a point on the hyperboloid of one sheet? Certainly. Simply pick any (x, y) on the ellipse $x^2 + 2y^2 = 1 + z^2$. Then $\sqrt{x^2 + y^2 + z^2}$ is large, at least as large as z . Thus it is unbounded.

You can also find intersections between lines and surfaces.

Example 16.9.4 Find the points of intersection of the line

$$(x, y, z) = (1 + t, 1 + 2t, 1 + t)$$

with the surface, $z = x^2 + y^2$.

First of all, there is no guarantee there is any intersection at all. But if it exists, you have only to solve the equation for t

$$1 + t = (1 + t)^2 + (1 + 2t)^2$$

This occurs at the two values of $t = -\frac{1}{2} + \frac{1}{10}\sqrt{5}$, $t = -\frac{1}{2} - \frac{1}{10}\sqrt{5}$. Therefore, the two points are

$$(1, 1, 1) + \left(-\frac{1}{2} + \frac{1}{10}\sqrt{5}\right)(1, 2, 1), \text{ and } (1, 1, 1) + \left(-\frac{1}{2} - \frac{1}{10}\sqrt{5}\right)(1, 2, 1)$$

¹It is traditional to refer to this as a hyperbolic paraboloid. Not a parabolic hyperboloid.

That is

$$\left(\frac{1}{2} + \frac{1}{10}\sqrt{5}, \frac{1}{5}\sqrt{5}, \frac{1}{2} + \frac{1}{10}\sqrt{5}\right), \left(\frac{1}{2} - \frac{1}{10}\sqrt{5}, -\frac{1}{5}\sqrt{5}, \frac{1}{2} - \frac{1}{10}\sqrt{5}\right).$$

16.10 Exercises

1. Determine whether the lines $(1, 1, 2) + t(1, 0, 3)$ and $(4, 1, 3) + t(3, 0, 1)$ have a point of intersection. If they do not intersect, explain why they do not.
2. Determine whether the lines $(1, 1, 2) + t(1, 0, 3)$ and $(4, 2, 3) + t(3, 0, 1)$ have a point of intersection. . If they do not intersect, explain why they do not.
3. Find where the line $(1, 0, 1) + t(1, 2, 1)$ intersects the surface $x^2 + y^2 + z^2 = 9$ if possible. If there is no intersection, explain why.
4. Find a parametric equation for the line through the points $(2, 3, 4, 5)$, $(-2, 3, 0, 1)$.
5. Find the equation of a line through $(1, 2, 3, 0)$ which has direction vector, $(2, 1, 3, 1)$.
6. Let $(x, y) = (2 \cos(t), 2 \sin(t))$ where $t \in [0, 2\pi]$. Describe the set of points encountered as t changes.
7. Let $(x, y, z) = (2 \cos(t), 2 \sin(t), t)$ where $t \in \mathbb{R}$. Describe the set of points encountered as t changes.
8. If there is a plane which contains the two lines,

$$(2t + 2, 1 + t, 3 + 2t) = (x, y, z)$$

and $(4 + t, 3 + 2t, 4 + t) = (x, y, z)$ find it. If there is no such plane tell why.

9. If there is a plane which contains the two lines,

$$(2t + 4, 1 + t, 3 + 2t) = (x, y, z)$$

and $(4 + t, t + 1, 3 + t) = (x, y, z)$ find it. If there is no such plane tell why.

10. Find the equation of the plane which contains the three points $(1, -2, 3)$, $(2, 3, 4)$, and $(3, 1, 2)$.
11. Find the equation of the plane which contains the three points $(1, 2, 3)$, $(2, 0, 4)$, and $(3, 1, 2)$.
12. Find the equation of the plane which contains the three points $(0, 2, 3)$, $(2, 3, 4)$, and $(3, 5, 2)$.
13. Find the equation of the plane which contains the three points $(1, 2, 3)$, $(0, 3, 4)$, and $(3, 6, 2)$.
14. Find the equation of the plane having a normal vector, $5\mathbf{i} + 2\mathbf{j} - 6\mathbf{k}$ which contains the point $(2, 1, 3)$.
15. Find the equation of the plane having a normal vector, $\mathbf{i} + 2\mathbf{j} - 4\mathbf{k}$ which contains the point $(2, 0, 1)$.
16. Find the equation of the plane having a normal vector, $2\mathbf{i} + \mathbf{j} - 6\mathbf{k}$ which contains the point $(1, 1, 2)$.

17. Find the equation of the plane having a normal vector, $\mathbf{i} + 2\mathbf{j} - 3\mathbf{k}$ which contains the point $(1, 0, 3)$.
18. Determine the intercepts and sketch the plane $3x - 2y + z = 4$.
19. Determine the intercepts and sketch the plane $x - 2y + z = 2$.
20. Determine the intercepts and sketch the plane $x + y + z = 3$.
21. Based on an analogy with the above pictures, sketch or otherwise describe the graph of $y = \frac{x^2}{a^2} - \frac{z^2}{b^2}$.
22. Based on an analogy with the above pictures, sketch or otherwise describe the graph of $\frac{z^2}{b^2} + \frac{y^2}{c^2} = 1 + \frac{x^2}{a^2}$.
23. The equation of a cone is $z^2 = x^2 + y^2$. Suppose this cone is intersected with the plane, $z = ay + 1, a > 0$. Consider the projection of the intersection of the cone with this plane. This means $\{(x, y) : (ay + 1)^2 = x^2 + y^2\}$. Show this sometimes results in a parabola, sometimes a hyperbola, and sometimes an ellipse depending on a .
24. Find the intercepts of the quadric surface, $x^2 + 4y^2 - z^2 = 4$ and sketch the surface. It does not intercept the z axis.
25. Find the intercepts of the quadric surface, $x^2 - (4y^2 + z^2) = 4$ and sketch the surface.
The intercepts are 2, -2 for x .
26. Find the intersection of the line $(x, y, z) = (1 + t, t, 3t)$ with the surface, $x^2/9 + y^2/4 + z^2/16 = 1$ if possible.

Part IV

Vector Calculus

Vector Valued Functions

17.0.1 Outcomes

1. Identify the domain of a vector function.
2. Represent combinations of multivariable functions algebraically.
3. Evaluate the limit of a function of several variables or show that it does not exist.
4. Determine whether a function is continuous at a given point. Give examples of continuous functions.
5. Recall and apply the extreme value theorem.

17.1 Vector Valued Functions

Vector valued functions have values in $\mathbb{R}^p$ where p is an integer at least as large as 1. Here is a simple example which is obviously of interest.

Example 17.1.1 *A rocket is launched from the rotating earth. You could define a function having values in $\mathbb{R}^3$ as $(r(t), \theta(t), \phi(t))$ where $r(t)$ is the distance of the center of mass of the rocket from the center of the earth, $\theta(t)$ is the longitude, and $\phi(t)$ is the latitude of the rocket.*

Example 17.1.2 *Let $\mathbf{f}(x, y) = (\sin xy, y^3 + x, x^4)$. Then $\mathbf{f}$ is a function defined on $\mathbb{R}^2$ which has values in $\mathbb{R}^3$. For example, $\mathbf{f}(1, 2) = (\sin 2, 9, 16)$.*

As usual, $D(\mathbf{f})$ denotes the domain of the function, $\mathbf{f}$ which is written in bold face because it will possibly have values in $\mathbb{R}^p$. When $D(\mathbf{f})$ is not specified, it will be understood that the domain of $\mathbf{f}$ consists of those things for which $\mathbf{f}$ makes sense.

Example 17.1.3 *Let $\mathbf{f}(x, y, z) = (\frac{x+y}{z}, \sqrt{1-x^2}, y)$. Then $D(\mathbf{f})$ would consist of the set of all (x, y, z) such that $|x| \leq 1$ and $z \neq 0$.*

There are many ways to make new functions from old ones.

Definition 17.1.4 *Let $\mathbf{f}, \mathbf{g}$ be functions with values in $\mathbb{R}^p$. Let a, b be points of $\mathbb{R}$ (scalars). Then $a\mathbf{f} + b\mathbf{g}$ is the name of a function whose domain is $D(\mathbf{f}) \cap D(\mathbf{g})$ which is defined as*

$$(a\mathbf{f} + b\mathbf{g})(\mathbf{x}) = a\mathbf{f}(\mathbf{x}) + b\mathbf{g}(\mathbf{x}).$$

$\mathbf{f} \cdot \mathbf{g}$ or $(\mathbf{f}, \mathbf{g})$ is the name of a function whose domain is $D(\mathbf{f}) \cap D(\mathbf{g})$ which is defined as

$$(\mathbf{f}, \mathbf{g})(\mathbf{x}) \equiv \mathbf{f} \cdot \mathbf{g}(\mathbf{x}) \equiv \mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x}).$$

If $\mathbf{f}$ and $\mathbf{g}$ have values in $\mathbb{R}^3$, define a new function, $\mathbf{f} \times \mathbf{g}$ by

$$\mathbf{f} \times \mathbf{g}(t) \equiv \mathbf{f}(t) \times \mathbf{g}(t).$$

If $\mathbf{f} : D(\mathbf{f}) \rightarrow X$ and $\mathbf{g} : X \rightarrow Y$, then $\mathbf{g} \circ \mathbf{f}$ is the name of a function whose domain is

$$\{\mathbf{x} \in D(\mathbf{f}) : \mathbf{f}(\mathbf{x}) \in D(\mathbf{g})\}$$

which is defined as

$$\mathbf{g} \circ \mathbf{f}(\mathbf{x}) \equiv \mathbf{g}(\mathbf{f}(\mathbf{x})).$$

This is called the composition of the two functions.

You should note that $\mathbf{f}(\mathbf{x})$ is not a function. It is the value of the function at the point, $\mathbf{x}$. The name of the function is $\mathbf{f}$. Nevertheless, people often write $\mathbf{f}(\mathbf{x})$ to denote a function and it doesn't cause too many problems in beginning courses. When this is done, the variable, $\mathbf{x}$ should be considered as a generic variable free to be anything in $D(\mathbf{f})$. I will use this slightly sloppy abuse of notation whenever convenient.

Example 17.1.5 Let $\mathbf{f}(t) \equiv (t, 1+t, 2)$ and $\mathbf{g}(t) \equiv (t^2, t, t)$. Then $\mathbf{f} \cdot \mathbf{g}$ is the name of the function satisfying

$$\mathbf{f} \cdot \mathbf{g}(t) = \mathbf{f}(t) \cdot \mathbf{g}(t) = t^3 + t + t^2 + 2t = t^3 + t^2 + 3t$$

Note that in this case it was assumed the domains of the functions consisted of all of $\mathbb{R}$ because this was the set on which the two both made sense. Also note that $\mathbf{f}$ and $\mathbf{g}$ map $\mathbb{R}$ into $\mathbb{R}^3$ but $\mathbf{f} \cdot \mathbf{g}$ maps $\mathbb{R}$ into $\mathbb{R}$.

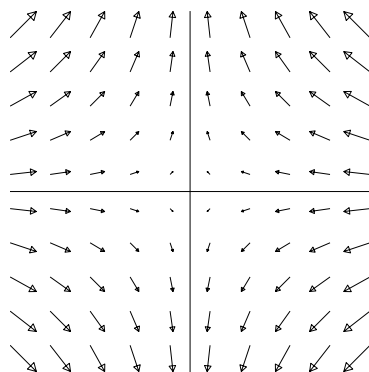
Example 17.1.6 Suppose $\mathbf{f}(t) = (2t, 1+t^2)$ and $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by $g(x, y) \equiv x+y$. Then $g \circ \mathbf{f} : \mathbb{R} \rightarrow \mathbb{R}$ and

$$g \circ \mathbf{f}(t) = g(\mathbf{f}(t)) = g(2t, 1+t^2) = 1 + 2t + t^2.$$

17.2 Vector Fields

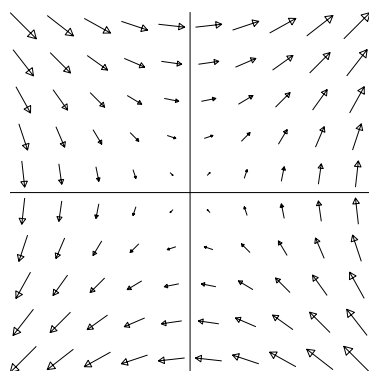
Some people find it useful to try and draw pictures to illustrate a vector valued function. This can be a very useful idea in the case where the function takes points in $D \subseteq \mathbb{R}^2$ and delivers a vector in $\mathbb{R}^2$. For many points, $(x, y) \in D$, you draw an arrow of the appropriate length and direction with its tail at (x, y) . The picture of all these arrows can give you an understanding of what is happening. For example if the vector valued function gives the velocity of a fluid at the point, (x, y) , the picture of these arrows can give an idea of the motion of the fluid. When they are long the fluid is moving fast, when they are short, the fluid is moving slowly the direction of these arrows is an indication of the direction of motion. The only sensible way to produce such a picture is with a computer. Otherwise, it becomes a worthless exercise in busy work. Furthermore, it is of limited usefulness in three dimensions because in three dimensions such pictures are too cluttered to convey much insight.

Example 17.2.1 Draw a picture of the vector field, $(-x, y)$ which gives the velocity of a fluid flowing in two dimensions.



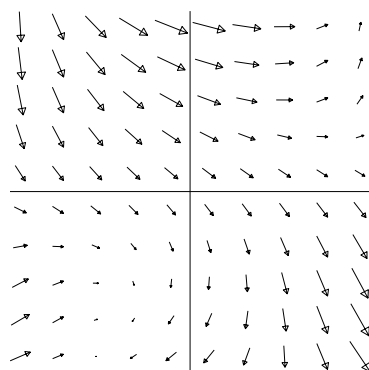
In this example, drawn by Maple, you can see how the arrows indicate the motion of this fluid.

Example 17.2.2 Draw a picture of the vector field (y, x) for the velocity of a fluid flowing in two dimensions.



So much for art. Get the computer to do it and it can be useful. If you try to do it, you will mainly waste time.

Example 17.2.3 Draw a picture of the vector field $(y \cos(x) + 1, x \sin(y) - 1)$ for the velocity of a fluid flowing in two dimensions.

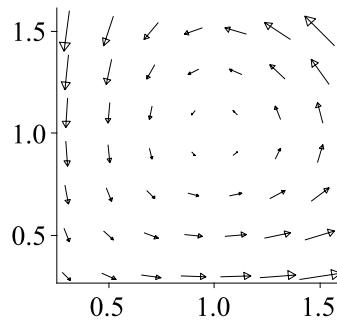


These pictures were drawn by maple. Note how they reveal both the direction and the magnitude of the vectors.

Written in terms of vectors,

$$(x', y') = (x(a - by), -y(c - dx))$$

The parameters a, b, c, d depend on the problem. The differential equations are saying that at a point (x, y) , the population vector, (x, y) moves in the direction of $(x(a - by), -y(c - dx))$. Here is the graph of the vector field which determines the Lotka Volterra system in the case where all the parameters equal 1 which is graphed near the point $(1, 1)$. What conclusions seem to be true based on the graph of this vector field? What happens if you start with a population vector near the point $(1, 1)$? Remember these vectors in the plane determine the directions of motion of the population vector.



How did I know to graph the vector field near $(1, 1)$?

17.4 Continuous Functions

What was done in one variable calculus for scalar functions is generalized here to include the case of a vector valued function of possibly many variables.

Definition 17.4.1 A function $\mathbf{f} : D(\mathbf{f}) \subseteq \mathbb{R}^p \rightarrow \mathbb{R}^q$ is continuous at $\mathbf{x} \in D(\mathbf{f})$ if for each $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $\mathbf{y} \in D(\mathbf{f})$ and

$$|\mathbf{y} - \mathbf{x}| < \delta$$

it follows that

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon.$$

$\mathbf{f}$ is continuous if it is continuous at every point of $D(\mathbf{f})$.

Note the total similarity to the scalar valued case.

17.4.1 Sufficient Conditions For Continuity

The next theorem is a fundamental result which allows less worry about the $\varepsilon \delta$ definition of continuity.

Theorem 17.4.2 The following assertions are valid.

1. The function, $a\mathbf{f} + b\mathbf{g}$ is continuous at $\mathbf{x}$ whenever $\mathbf{f}, \mathbf{g}$ are continuous at $\mathbf{x} \in D(\mathbf{f}) \cap D(\mathbf{g})$ and $a, b \in \mathbb{R}$.
2. If $\mathbf{f}$ is continuous at $\mathbf{x}$, $\mathbf{f}(\mathbf{x}) \in D(\mathbf{g}) \subseteq \mathbb{R}^p$, and $\mathbf{g}$ is continuous at $\mathbf{f}(\mathbf{x})$, then $\mathbf{g} \circ \mathbf{f}$ is continuous at $\mathbf{x}$.

3. If $\mathbf{f} = (f_1, \dots, f_q) : D(\mathbf{f}) \rightarrow \mathbb{R}^q$, then $\mathbf{f}$ is continuous if and only if each f_k is a continuous real valued function.
4. The function $f : \mathbb{R}^p \rightarrow \mathbb{R}$, given by $f(\mathbf{x}) = |\mathbf{x}|$ is continuous.

The proof of this theorem is in the last section of this chapter. Its conclusions are not surprising. For example the first claim says that $(a\mathbf{f} + b\mathbf{g})(\mathbf{y})$ is close to $(a\mathbf{f} + b\mathbf{g})(\mathbf{x})$ when $\mathbf{y}$ is close to $\mathbf{x}$ provided the same can be said about $\mathbf{f}$ and $\mathbf{g}$. For the second claim, if $\mathbf{y}$ is close to $\mathbf{x}$, $\mathbf{f}(\mathbf{x})$ is close to $\mathbf{f}(\mathbf{y})$ and so by continuity of $\mathbf{g}$ at $\mathbf{f}(\mathbf{x})$, $\mathbf{g}(\mathbf{f}(\mathbf{y}))$ is close to $\mathbf{g}(\mathbf{f}(\mathbf{x}))$. To see the third claim is likely, note that closeness in $\mathbb{R}^p$ is the same as closeness in each coordinate. The fourth claim is immediate from the triangle inequality.

For functions defined on $\mathbb{R}^n$, there is a notion of polynomial just as there is for functions defined on $\mathbb{R}$.

Definition 17.4.3 Let α be an n dimensional multi-index. This means

$$\alpha = (\alpha_1, \dots, \alpha_n)$$

where each α_i is a natural number or zero. Also, let

$$|\alpha| \equiv \sum_{i=1}^n |\alpha_i|$$

The symbol, $\mathbf{x}^\alpha$ means

$$\mathbf{x}^\alpha \equiv x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}.$$

An n dimensional polynomial of degree m is a function of the form

$$p(\mathbf{x}) = \sum_{|\alpha| \leq m} d_\alpha \mathbf{x}^\alpha.$$

where the d_α are real numbers.

The above theorem implies that polynomials are all continuous.

17.5 Limits Of A Function

As in the case of scalar valued functions of one variable, a concept closely related to continuity is that of the **limit of a function**. The notion of limit of a function makes sense at points, $\mathbf{x}$, which are limit points of $D(\mathbf{f})$ and this concept is defined next.

Definition 17.5.1 Let $A \subseteq \mathbb{R}^m$ be a set. A point, $\mathbf{x}$, is a limit point of A if $B(\mathbf{x}, r)$ contains infinitely many points of A for every $r > 0$.

Definition 17.5.2 Let $\mathbf{f} : D(\mathbf{f}) \subseteq \mathbb{R}^p \rightarrow \mathbb{R}^q$ be a function and let $\mathbf{x}$ be a **limit point** of $D(\mathbf{f})$. Then

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$$

if and only if the following condition holds. For all $\varepsilon > 0$ there exists $\delta > 0$ such that if

$$0 < |\mathbf{y} - \mathbf{x}| < \delta, \text{ and } \mathbf{y} \in D(\mathbf{f})$$

then,

$$|\mathbf{L} - \mathbf{f}(\mathbf{y})| < \varepsilon.$$

Theorem 17.5.3 *If $\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$ and $\lim_{y \rightarrow x} \mathbf{f}(\mathbf{y}) = \mathbf{L}_1$, then $\mathbf{L} = \mathbf{L}_1$.*

Proof: Let $\varepsilon > 0$ be given. There exists $\delta > 0$ such that if $0 < |\mathbf{y} - \mathbf{x}| < \delta$ and $\mathbf{y} \in D(\mathbf{f})$, then

$$|\mathbf{f}(\mathbf{y}) - \mathbf{L}| < \varepsilon, \quad |\mathbf{f}(\mathbf{y}) - \mathbf{L}_1| < \varepsilon.$$

Pick such a $\mathbf{y}$. There exists one because $\mathbf{x}$ is a limit point of $D(\mathbf{f})$. Then

$$|\mathbf{L} - \mathbf{L}_1| \leq |\mathbf{L} - \mathbf{f}(\mathbf{y})| + |\mathbf{f}(\mathbf{y}) - \mathbf{L}_1| < \varepsilon + \varepsilon = 2\varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this shows $\mathbf{L} = \mathbf{L}_1$.

As in the case of functions of one variable, one can define what it means for $\lim_{\mathbf{y} \rightarrow \mathbf{x}} f(\mathbf{x}) = \pm\infty$.

Definition 17.5.4 *If $f(\mathbf{x}) \in \mathbb{R}$, $\lim_{\mathbf{y} \rightarrow \mathbf{x}} f(\mathbf{x}) = \infty$ if for every number l , there exists $\delta > 0$ such that whenever $|\mathbf{y} - \mathbf{x}| < \delta$ and $\mathbf{y} \in D(\mathbf{f})$, then $f(\mathbf{x}) > l$.*

The following theorem is just like the one variable version of calculus.

Theorem 17.5.5 *Suppose*

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}, \quad \lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{g}(\mathbf{y}) = \mathbf{K}$$

where $\mathbf{K}, \mathbf{L} \in \mathbb{R}^q$. Then if $a, b \in \mathbb{R}$,

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} (a\mathbf{f}(\mathbf{y}) + b\mathbf{g}(\mathbf{y})) = a\mathbf{L} + b\mathbf{K}, \quad (17.1)$$

$$\lim_{y \rightarrow x} \mathbf{f} \cdot \mathbf{g}(y) = \mathbf{L} \cdot \mathbf{K} \quad (17.2)$$

and if g is scalar valued with $\lim_{\mathbf{y} \rightarrow \mathbf{x}} g(\mathbf{y}) = K \neq 0$,

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) g(\mathbf{y}) = \mathbf{L}K. \quad (17.3)$$

Also, if $\mathbf{h}$ is a continuous function defined near $\mathbf{L}$, then

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{h} \circ \mathbf{f}(\mathbf{y}) = \mathbf{h}(\mathbf{L}). \quad (17.4)$$

Suppose $\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$. If $|\mathbf{f}(\mathbf{y}) - \mathbf{b}| \leq r$ for all $\mathbf{y}$ sufficiently close to $\mathbf{x}$, then $|\mathbf{L} - \mathbf{b}| \leq r$ also.

Proof: The proof of 17.1 is left for you. It is like a corresponding theorem for continuous functions. Now 17.2 is to be verified. Let $\varepsilon > 0$ be given. Then by the triangle inequality,

$$\begin{aligned} |\mathbf{f} \cdot \mathbf{g}(\mathbf{y}) - \mathbf{L} \cdot \mathbf{K}| &\leq |\mathbf{f} \cdot \mathbf{g}(\mathbf{y}) - \mathbf{f}(\mathbf{y}) \cdot \mathbf{K}| + |\mathbf{f}(\mathbf{y}) \cdot \mathbf{K} - \mathbf{L} \cdot \mathbf{K}| \\ &\leq |\mathbf{f}(\mathbf{y})| |\mathbf{g}(\mathbf{y}) - \mathbf{K}| + |\mathbf{K}| |\mathbf{f}(\mathbf{y}) - \mathbf{L}|. \end{aligned}$$

There exists δ_1 such that if $0 < |\mathbf{y} - \mathbf{x}| < \delta_1$ and $\mathbf{y} \in D(\mathbf{f})$, then

$$|\mathbf{f}(\mathbf{y}) - \mathbf{L}| < 1,$$

and so for such $\mathbf{y}$, the triangle inequality implies, $|\mathbf{f}(\mathbf{y})| < 1 + |\mathbf{L}|$. Therefore, for $0 < |\mathbf{y} - \mathbf{x}| < \delta_1$,

$$|\mathbf{f} \cdot \mathbf{g}(\mathbf{y}) - \mathbf{L} \cdot \mathbf{K}| \leq (1 + |\mathbf{K}| + |\mathbf{L}|) [|\mathbf{g}(\mathbf{y}) - \mathbf{K}| + |\mathbf{f}(\mathbf{y}) - \mathbf{L}|]. \quad (17.5)$$

Now let $0 < \delta_2$ be such that if $\mathbf{y} \in D(\mathbf{f})$ and $0 < |\mathbf{x} - \mathbf{y}| < \delta_2$,

$$|\mathbf{f}(\mathbf{y}) - \mathbf{L}| < \frac{\varepsilon}{2(1 + |\mathbf{K}| + |\mathbf{L}|)}, \quad |\mathbf{g}(\mathbf{y}) - \mathbf{K}| < \frac{\varepsilon}{2(1 + |\mathbf{K}| + |\mathbf{L}|)}.$$

Then letting $0 < \delta \leq \min(\delta_1, \delta_2)$, it follows from 17.5 that

$$|\mathbf{f} \cdot \mathbf{g}(\mathbf{y}) - \mathbf{L} \cdot \mathbf{K}| < \varepsilon$$

and this proves 17.2.

The proof of 17.3 is left to you.

Consider 17.4. Since $\mathbf{h}$ is continuous near $\mathbf{L}$, it follows that for $\varepsilon > 0$ given, there exists $\eta > 0$ such that if $|\mathbf{y} - \mathbf{L}| < \eta$, then

$$|\mathbf{h}(\mathbf{y}) - \mathbf{h}(\mathbf{L})| < \varepsilon$$

Now since $\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$, there exists $\delta > 0$ such that if $0 < |\mathbf{y} - \mathbf{x}| < \delta$, then

$$|\mathbf{f}(\mathbf{y}) - \mathbf{L}| < \eta.$$

Therefore, if $0 < |\mathbf{y} - \mathbf{x}| < \delta$,

$$|\mathbf{h}(\mathbf{f}(\mathbf{y})) - \mathbf{h}(\mathbf{L})| < \varepsilon.$$

It only remains to verify the last assertion. Assume $|\mathbf{f}(\mathbf{y}) - \mathbf{b}| \leq r$. It is required to show that $|\mathbf{L} - \mathbf{b}| \leq r$. If this is not true, then $|\mathbf{L} - \mathbf{b}| > r$. Consider $B(\mathbf{L}, |\mathbf{L} - \mathbf{b}| - r)$. Since $\mathbf{L}$ is the limit of $\mathbf{f}$, it follows $\mathbf{f}(\mathbf{y}) \in B(\mathbf{L}, |\mathbf{L} - \mathbf{b}| - r)$ whenever $\mathbf{y} \in D(\mathbf{f})$ is close enough to $\mathbf{x}$. Thus, by the triangle inequality,

$$|\mathbf{f}(\mathbf{y}) - \mathbf{L}| < |\mathbf{L} - \mathbf{b}| - r$$

and so

$$\begin{aligned} r &< |\mathbf{L} - \mathbf{b}| - |\mathbf{f}(\mathbf{y}) - \mathbf{L}| \leq ||\mathbf{b} - \mathbf{L}| - |\mathbf{f}(\mathbf{y}) - \mathbf{L}|| \\ &\leq |\mathbf{b} - \mathbf{f}(\mathbf{y})|, \end{aligned}$$

a contradiction to the assumption that $|\mathbf{b} - \mathbf{f}(\mathbf{y})| \leq r$.

Theorem 17.5.6 For $\mathbf{f} : D(\mathbf{f}) \rightarrow \mathbb{R}^q$ and $\mathbf{x} \in D(\mathbf{f})$ a limit point of $D(\mathbf{f})$, $\mathbf{f}$ is continuous at $\mathbf{x}$ if and only if

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{f}(\mathbf{x}).$$

Proof: First suppose $\mathbf{f}$ is continuous at $\mathbf{x}$ a limit point of $D(\mathbf{f})$. Then for every $\varepsilon > 0$ there exists $\delta > 0$ such that if $|\mathbf{y} - \mathbf{x}| < \delta$ and $\mathbf{y} \in D(\mathbf{f})$, then $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$. In particular, this holds if $0 < |\mathbf{x} - \mathbf{y}| < \delta$ and this is just the definition of the limit. Hence $\mathbf{f}(\mathbf{x}) = \lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y})$.

Next suppose $\mathbf{x}$ is a limit point of $D(\mathbf{f})$ and $\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{f}(\mathbf{x})$. This means that if $\varepsilon > 0$ there exists $\delta > 0$ such that for $0 < |\mathbf{x} - \mathbf{y}| < \delta$ and $\mathbf{y} \in D(\mathbf{f})$, it follows $|\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})| < \varepsilon$. However, if $\mathbf{y} = \mathbf{x}$, then $|\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x})| = |\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x})| = 0$ and so whenever $\mathbf{y} \in D(\mathbf{f})$ and $|\mathbf{x} - \mathbf{y}| < \delta$, it follows $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$, showing $\mathbf{f}$ is continuous at $\mathbf{x}$.

The following theorem is important.

Theorem 17.5.7 Suppose $\mathbf{f} : D(\mathbf{f}) \rightarrow \mathbb{R}^q$. Then for $\mathbf{x}$ a limit point of $D(\mathbf{f})$,

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L} \tag{17.6}$$

if and only if

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} f_k(\mathbf{y}) = L_k \quad (17.7)$$

where $\mathbf{f}(\mathbf{y}) \equiv (f_1(\mathbf{y}), \dots, f_p(\mathbf{y}))$ and $\mathbf{L} \equiv (L_1, \dots, L_p)$.

In the case where $q = 3$ and $\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$ and $\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{g}(\mathbf{y}) = \mathbf{K}$, then

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) \times \mathbf{g}(\mathbf{y}) = \mathbf{L} \times \mathbf{K}. \quad (17.8)$$

Proof: Suppose 17.6. Then letting $\varepsilon > 0$ be given there exists $\delta > 0$ such that if $0 < |\mathbf{y} - \mathbf{x}| < \delta$, it follows

$$|f_k(\mathbf{y}) - L_k| \leq |\mathbf{f}(\mathbf{y}) - \mathbf{L}| < \varepsilon$$

which verifies 17.7.

Now suppose 17.7 holds. Then letting $\varepsilon > 0$ be given, there exists δ_k such that if $0 < |\mathbf{y} - \mathbf{x}| < \delta_k$, then

$$|f_k(\mathbf{y}) - L_k| < \frac{\varepsilon}{\sqrt{p}}.$$

Let $0 < \delta < \min(\delta_1, \dots, \delta_p)$. Then if $0 < |\mathbf{y} - \mathbf{x}| < \delta$, it follows

$$\begin{aligned} |\mathbf{f}(\mathbf{y}) - \mathbf{L}| &= \left(\sum_{k=1}^p |f_k(\mathbf{y}) - L_k|^2 \right)^{1/2} \\ &< \left(\sum_{k=1}^p \frac{\varepsilon^2}{p} \right)^{1/2} = \varepsilon. \end{aligned}$$

It remains to verify 17.8. But from the first part of this theorem and the description of the cross product presented earlier in terms of the permutation symbol,

$$\begin{aligned} \lim_{\mathbf{y} \rightarrow \mathbf{x}} (\mathbf{f}(\mathbf{y}) \times \mathbf{g}(\mathbf{y}))_i &= \lim_{\mathbf{y} \rightarrow \mathbf{x}} \varepsilon_{ijk} f_j(\mathbf{y}) g_k(\mathbf{y}) \\ &= \varepsilon_{ijk} L_j K_k = (\mathbf{L} \times \mathbf{K})_i. \end{aligned}$$

Therefore, from the first part of this theorem, this establishes 21.5. This completes the proof.

Example 17.5.8 Find $\lim_{(x,y) \rightarrow (3,1)} \left(\frac{x^2-9}{x-3}, y \right)$.

It is clear that $\lim_{(x,y) \rightarrow (3,1)} \frac{x^2-9}{x-3} = 6$ and $\lim_{(x,y) \rightarrow (3,1)} y = 1$. Therefore, this limit equals $(6, 1)$.

Example 17.5.9 Find $\lim_{(x,y) \rightarrow (0,0)} \frac{xy}{x^2+y^2}$.

First of all observe the domain of the function is $\mathbb{R}^2 \setminus \{(0,0)\}$, every point in $\mathbb{R}^2$ except the origin. Therefore, $(0,0)$ is a limit point of the domain of the function so it might make sense to take a limit. However, just as in the case of a function of one variable, the limit may not exist. In fact, this is the case here. To see this, take points on the line $y = 0$. At these points, the value of the function equals 0. Now consider points on the line $y = x$ where the value of the function equals $1/2$. Since arbitrarily close to $(0,0)$ there are points where the function equals $1/2$ and points where the function has the value 0, it follows there can be no limit. Just take $\varepsilon = 1/10$ for example. You can't be within $1/10$ of $1/2$ and also within $1/10$ of 0 at the same time.

Note it is necessary to rely on the definition of the limit much more than in the case of a function of one variable and there are no easy ways to do limit problems for functions of more than one variable. It is what it is and you will not deal with these concepts without suffering and anguish.

17.6 Properties Of Continuous Functions

Functions of p variables have many of the same properties as functions of one variable. First there is a version of the extreme value theorem generalizing the one dimensional case.

Theorem 17.6.1 *Let C be closed and bounded and let $f : C \rightarrow \mathbb{R}$ be continuous. Then f achieves its maximum and its minimum on C . This means there exist, $\mathbf{x}_1, \mathbf{x}_2 \in C$ such that for all $\mathbf{x} \in C$,*

$$f(\mathbf{x}_1) \leq f(\mathbf{x}) \leq f(\mathbf{x}_2).$$

There is also the long technical theorem about sums and products of continuous functions. These theorems are proved in the next section.

Theorem 17.6.2 *The following assertions are valid*

1. *The function, $a\mathbf{f} + b\mathbf{g}$ is continuous at $\mathbf{x}$ when $\mathbf{f}, \mathbf{g}$ are continuous at $\mathbf{x} \in D(\mathbf{f}) \cap D(\mathbf{g})$ and $a, b \in \mathbb{R}$.*
2. *If f and g are each real valued functions continuous at $\mathbf{x}$, then fg is continuous at $\mathbf{x}$. If, in addition to this, $g(\mathbf{x}) \neq 0$, then f/g is continuous at $\mathbf{x}$.*
3. *If $\mathbf{f}$ is continuous at $\mathbf{x}$, $\mathbf{f}(\mathbf{x}) \in D(\mathbf{g}) \subseteq \mathbb{R}^p$, and $\mathbf{g}$ is continuous at $\mathbf{f}(\mathbf{x})$, then $\mathbf{g} \circ \mathbf{f}$ is continuous at $\mathbf{x}$.*
4. *If $\mathbf{f} = (f_1, \dots, f_q) : D(\mathbf{f}) \rightarrow \mathbb{R}^q$, then $\mathbf{f}$ is continuous if and only if each f_k is a continuous real valued function.*
5. *The function $f : \mathbb{R}^p \rightarrow \mathbb{R}$, given by $f(\mathbf{x}) = |\mathbf{x}|$ is continuous.*

17.7 Exercises

1. Let $\mathbf{f}(t) = \left(t, t^2 + 1, \frac{t}{t+1}\right)$ and let $\mathbf{g}(t) = \left(t + 1, 1, \frac{t}{t^2+1}\right)$. Find $\mathbf{f} \cdot \mathbf{g}$.
2. Let $\mathbf{f}, \mathbf{g}$ be given in the previous problem. Find $\mathbf{f} \times \mathbf{g}$.
3. Let $\mathbf{f}(t) = (t, t^2, t^3)$, $\mathbf{g}(t) = (1, t^2, t^2)$, and $\mathbf{h}(t) = (\sin t, t, 1)$. Find the time rate of change of the box product of the vectors $\mathbf{f}, \mathbf{g}$, and $\mathbf{h}$.
4. Let $\mathbf{f}(t) = (t, \sin t)$. Show $\mathbf{f}$ is continuous at every point t .
5. Suppose $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq K|\mathbf{x} - \mathbf{y}|$ where K is a constant. Show that $\mathbf{f}$ is everywhere continuous. Functions satisfying such an inequality are called Lipschitz functions.
6. Suppose $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq K|\mathbf{x} - \mathbf{y}|^\alpha$ where K is a constant and $\alpha \in (0, 1)$. Show that $\mathbf{f}$ is everywhere continuous.
7. Suppose $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ is given by $f(\mathbf{x}) = 3x_1x_2 + 2x_3^2$. Use Theorem 17.4.2 to verify that f is continuous. **Hint:** You should first verify that the function, $\pi_k : \mathbb{R}^3 \rightarrow \mathbb{R}$ given by $\pi_k(\mathbf{x}) = x_k$ is a continuous function.
8. Show that if $f : \mathbb{R}^q \rightarrow \mathbb{R}$ is a polynomial then it is continuous.
9. State and prove a theorem which involves quotients of functions encountered in the previous problem.

10. Let

$$f(x, y) \equiv \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}.$$

Find $\lim_{(x,y) \rightarrow (0,0)} f(x, y)$ if it exists. If it does not exist, tell why it does not exist. **Hint:** Consider along the line $y = x$ and along the line $y = 0$.

11. Find the following limits if possible

$$(a) \lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - y^2}{x^2 + y^2}$$

$$(b) \lim_{(x,y) \rightarrow (0,0)} \frac{x(x^2 - y^2)}{(x^2 + y^2)} = 0$$

$$(c) \lim_{(x,y) \rightarrow (0,0)} \frac{(x^2 - y^4)^2}{(x^2 + y^4)^2} \quad \textbf{Hint:} \text{ Consider along } y = 0 \text{ and along } x = y^2.$$

$$(d) \lim_{(x,y) \rightarrow (0,0)} x \sin\left(\frac{1}{x^2 + y^2}\right)$$

$$(e) \lim_{(x,y) \rightarrow (1,2)} \frac{-2yx^2 + 8yx + 34y^3 - 18y^2 + 6x^2 - 13x - 20 - xy^2 - x^3}{-y^2 + 4y - 5 - x^2 + 2x}. \quad \textbf{Hint:} \text{ It might help to write this in terms of the variables } (s, t) = (x - 1, y - 2).$$

12. Suppose
- $\lim_{x \rightarrow 0} f(x, 0) = 0 = \lim_{y \rightarrow 0} f(0, y)$
- . Does it follow that

$$\lim_{(x,y) \rightarrow (0,0)} f(x, y) = 0?$$

Prove or give counter example.

- 13.
- $\mathbf{f} : D \subseteq \mathbb{R}^p \rightarrow \mathbb{R}^q$
- is Lipschitz continuous or just Lipschitz for short if there exists a constant,
- K
- such that

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq K |\mathbf{x} - \mathbf{y}|$$

for all $\mathbf{x}, \mathbf{y} \in D$. Show every Lipschitz function is uniformly continuous which means that given $\varepsilon > 0$ there exists $\delta > 0$ independent of $\mathbf{x}$ such that if $|\mathbf{x} - \mathbf{y}| < \delta$, then $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$.

14. If $\mathbf{f}$ is uniformly continuous, does it follow that $|\mathbf{f}|$ is also uniformly continuous? If $|\mathbf{f}|$ is uniformly continuous does it follow that $\mathbf{f}$ is uniformly continuous? Answer the same questions with “uniformly continuous” replaced with “continuous”. Explain why.
15. Let f be defined on the positive integers. Thus $D(f) = \mathbb{N}$. Show that f is automatically continuous at every point of $D(f)$. Is it also uniformly continuous? What does this mean about the concept of continuous functions being those which can be graphed without taking the pencil off the paper?

16. Let

$$f(x, y) = \frac{(x^2 - y^4)^2}{(x^2 + y^4)^2} \text{ if } (x, y) \neq (0, 0)$$

Show $\lim_{t \rightarrow 0} f(tx, ty) = 1$ for any choice of (x, y) . Using Problem 11c, what does this tell you about limits existing just because the limit along any line exists.

17. Let
- $f(x, y, z) = x^2y + \sin(xyz)$
- . Does
- f
- achieve a maximum on the set

$$\{(x, y, z) : x^2 + y^2 + 2z^2 \leq 8\}?$$

Explain why.

18. Suppose $\mathbf{x}$ is defined to be a limit point of a set, A if and only if for all $r > 0$, $B(\mathbf{x}, r)$ contains a point of A different than $\mathbf{x}$. Show this is equivalent to the above definition of limit point.
19. Give an example of a set of points in $\mathbb{R}^3$ which has no limit points. Show that if $D(\mathbf{f})$ equals this set, then $\mathbf{f}$ is continuous. Show that more generally, if $\mathbf{f}$ is any function for which $D(\mathbf{f})$ has no limit points, then $\mathbf{f}$ is continuous.
20. Let $\{\mathbf{x}_k\}_{k=1}^n$ be any finite set of points in $\mathbb{R}^p$. Show this set has no limit points.
21. Suppose S is any set of points such that every pair of points is at least as far apart as 1. Show S has no limit points.
22. Find $\lim_{\mathbf{x} \rightarrow \mathbf{0}} \frac{\sin(|\mathbf{x}|)}{|\mathbf{x}|}$ and prove your answer from the definition of limit.
23. Suppose $\mathbf{g}$ is a continuous vector valued function of one variable defined on $[0, \infty)$. Prove

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{g}(|\mathbf{x}|) = \mathbf{g}(|\mathbf{x}_0|).$$

17.8 Some Fundamentals*



This section contains the proofs of the theorems which were stated without proof along with some other significant topics which will be useful later. These topics are of fundamental significance but are difficult.

Theorem 17.8.1 *The following assertions are valid*

1. The function, $a\mathbf{f} + b\mathbf{g}$ is continuous at $\mathbf{x}$ when $\mathbf{f}, \mathbf{g}$ are continuous at $\mathbf{x} \in D(\mathbf{f}) \cap D(\mathbf{g})$ and $a, b \in \mathbb{R}$.
2. If f and g are each real valued functions continuous at $\mathbf{x}$, then fg is continuous at $\mathbf{x}$. If, in addition to this, $g(\mathbf{x}) \neq 0$, then f/g is continuous at $\mathbf{x}$.
3. If $\mathbf{f}$ is continuous at $\mathbf{x}$, $\mathbf{f}(\mathbf{x}) \in D(\mathbf{g}) \subseteq \mathbb{R}^p$, and $\mathbf{g}$ is continuous at $\mathbf{f}(\mathbf{x})$, then $\mathbf{g} \circ \mathbf{f}$ is continuous at $\mathbf{x}$.
4. If $\mathbf{f} = (f_1, \dots, f_q) : D(\mathbf{f}) \rightarrow \mathbb{R}^q$, then $\mathbf{f}$ is continuous if and only if each f_k is a continuous real valued function.
5. The function $f : \mathbb{R}^p \rightarrow \mathbb{R}$, given by $f(\mathbf{x}) = |\mathbf{x}|$ is continuous.

Proof: Begin with 1.) Let $\varepsilon > 0$ be given. By assumption, there exist $\delta_1 > 0$ such that whenever $|\mathbf{x} - \mathbf{y}| < \delta_1$, it follows $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \frac{\varepsilon}{2(|a|+|b|+1)}$ and there exists $\delta_2 > 0$ such that whenever $|\mathbf{x} - \mathbf{y}| < \delta_2$, it follows that $|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y})| < \frac{\varepsilon}{2(|a|+|b|+1)}$. Then let

$0 < \delta \leq \min(\delta_1, \delta_2)$. If $|\mathbf{x} - \mathbf{y}| < \delta$, then everything happens at once. Therefore, using the triangle inequality

$$\begin{aligned} & |a\mathbf{f}(\mathbf{x}) + b\mathbf{f}(\mathbf{x}) - (a\mathbf{g}(\mathbf{y}) + b\mathbf{g}(\mathbf{y}))| \\ & \leq |a| |\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| + |b| |\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{y})| \\ & < |a| \left(\frac{\varepsilon}{2(|a| + |b| + 1)} \right) + |b| \left(\frac{\varepsilon}{2(|a| + |b| + 1)} \right) < \varepsilon. \end{aligned}$$

Now begin on 2.) There exists $\delta_1 > 0$ such that if $|\mathbf{y} - \mathbf{x}| < \delta_1$, then $|f(\mathbf{x}) - f(\mathbf{y})| < 1$. Therefore, for such $\mathbf{y}$,

$$|f(\mathbf{y})| < 1 + |f(\mathbf{x})|.$$

It follows that for such $\mathbf{y}$,

$$\begin{aligned} |fg(\mathbf{x}) - fg(\mathbf{y})| & \leq |f(\mathbf{x})g(\mathbf{x}) - g(\mathbf{x})f(\mathbf{y})| + |g(\mathbf{x})f(\mathbf{y}) - f(\mathbf{y})g(\mathbf{y})| \\ & \leq |g(\mathbf{x})| |f(\mathbf{x}) - f(\mathbf{y})| + |f(\mathbf{y})| |g(\mathbf{x}) - g(\mathbf{y})| \\ & \leq (1 + |g(\mathbf{x})| + |f(\mathbf{y})|) [|g(\mathbf{x}) - g(\mathbf{y})| + |f(\mathbf{x}) - f(\mathbf{y})|]. \end{aligned}$$

Now let $\varepsilon > 0$ be given. There exists δ_2 such that if $|\mathbf{x} - \mathbf{y}| < \delta_2$, then

$$|g(\mathbf{x}) - g(\mathbf{y})| < \frac{\varepsilon}{2(1 + |g(\mathbf{x})| + |f(\mathbf{y})|)},$$

and there exists δ_3 such that if $|\mathbf{x} - \mathbf{y}| < \delta_3$, then

$$|f(\mathbf{x}) - f(\mathbf{y})| < \frac{\varepsilon}{2(1 + |g(\mathbf{x})| + |f(\mathbf{y})|)}$$

Now let $0 < \delta \leq \min(\delta_1, \delta_2, \delta_3)$. Then if $|\mathbf{x} - \mathbf{y}| < \delta$, all the above hold at once and

$$\begin{aligned} & |fg(\mathbf{x}) - fg(\mathbf{y})| \leq \\ & (1 + |g(\mathbf{x})| + |f(\mathbf{y})|) [|g(\mathbf{x}) - g(\mathbf{y})| + |f(\mathbf{x}) - f(\mathbf{y})|] \\ & < (1 + |g(\mathbf{x})| + |f(\mathbf{y})|) \left(\frac{\varepsilon}{2(1 + |g(\mathbf{x})| + |f(\mathbf{y})|)} + \frac{\varepsilon}{2(1 + |g(\mathbf{x})| + |f(\mathbf{y})|)} \right) = \varepsilon. \end{aligned}$$

This proves the first part of 2.) To obtain the second part, let δ_1 be as described above and let $\delta_0 > 0$ be such that for $|\mathbf{x} - \mathbf{y}| < \delta_0$,

$$|g(\mathbf{x}) - g(\mathbf{y})| < |g(\mathbf{x})|/2$$

and so by the triangle inequality,

$$-|g(\mathbf{x})|/2 \leq |g(\mathbf{y})| - |g(\mathbf{x})| \leq |g(\mathbf{x})|/2$$

which implies $|g(\mathbf{y})| \geq |g(\mathbf{x})|/2$, and $|g(\mathbf{y})| < 3|g(\mathbf{x})|/2$.

Then if $|\mathbf{x} - \mathbf{y}| < \min(\delta_0, \delta_1)$,

$$\begin{aligned} \left| \frac{f(\mathbf{x})}{g(\mathbf{x})} - \frac{f(\mathbf{y})}{g(\mathbf{y})} \right| & = \left| \frac{f(\mathbf{x})g(\mathbf{y}) - f(\mathbf{y})g(\mathbf{x})}{g(\mathbf{x})g(\mathbf{y})} \right| \\ & \leq \frac{|f(\mathbf{x})g(\mathbf{y}) - f(\mathbf{y})g(\mathbf{x})|}{\left(\frac{|g(\mathbf{x})|^2}{2} \right)} \\ & = \frac{2|f(\mathbf{x})g(\mathbf{y}) - f(\mathbf{y})g(\mathbf{x})|}{|g(\mathbf{x})|^2} \end{aligned}$$

$$\begin{aligned}
&\leq \frac{2}{|g(\mathbf{x})|^2} [|f(\mathbf{x})g(\mathbf{y}) - f(\mathbf{y})g(\mathbf{y}) + f(\mathbf{y})g(\mathbf{y}) - f(\mathbf{y})g(\mathbf{x})|] \\
&\leq \frac{2}{|g(\mathbf{x})|^2} [|g(\mathbf{y})||f(\mathbf{x}) - f(\mathbf{y})| + |f(\mathbf{y})||g(\mathbf{y}) - g(\mathbf{x})|] \\
&\leq \frac{2}{|g(\mathbf{x})|^2} \left[\frac{3}{2} |g(\mathbf{x})||f(\mathbf{x}) - f(\mathbf{y})| + (1 + |f(\mathbf{x})|)|g(\mathbf{y}) - g(\mathbf{x})| \right] \\
&\leq \frac{2}{|g(\mathbf{x})|^2} (1 + 2|f(\mathbf{x})| + 2|g(\mathbf{x})|) [|f(\mathbf{x}) - f(\mathbf{y})| + |g(\mathbf{y}) - g(\mathbf{x})|] \\
&\equiv M [|f(\mathbf{x}) - f(\mathbf{y})| + |g(\mathbf{y}) - g(\mathbf{x})|]
\end{aligned}$$

where

$$M \equiv \frac{2}{|g(\mathbf{x})|^2} (1 + 2|f(\mathbf{x})| + 2|g(\mathbf{x})|)$$

Now let δ_2 be such that if $|\mathbf{x} - \mathbf{y}| < \delta_2$, then

$$|f(\mathbf{x}) - f(\mathbf{y})| < \frac{\varepsilon}{2} M^{-1}$$

and let δ_3 be such that if $|\mathbf{x} - \mathbf{y}| < \delta_3$, then

$$|g(\mathbf{y}) - g(\mathbf{x})| < \frac{\varepsilon}{2} M^{-1}.$$

Then if $0 < \delta \leq \min(\delta_0, \delta_1, \delta_2, \delta_3)$, and $|\mathbf{x} - \mathbf{y}| < \delta$, everything holds and

$$\begin{aligned}
\left| \frac{f(\mathbf{x})}{g(\mathbf{x})} - \frac{f(\mathbf{y})}{g(\mathbf{y})} \right| &\leq M [|f(\mathbf{x}) - f(\mathbf{y})| + |g(\mathbf{y}) - g(\mathbf{x})|] \\
&< M \left[\frac{\varepsilon}{2} M^{-1} + \frac{\varepsilon}{2} M^{-1} \right] = \varepsilon.
\end{aligned}$$

This completes the proof of the second part of 2.) Note that in these proofs no effort is made to find some sort of “best” δ . The problem is one which has a yes or a no answer. Either is it or it is not continuous.

Now begin on 3.). If $\mathbf{f}$ is continuous at $\mathbf{x}$, $\mathbf{f}(\mathbf{x}) \in D(\mathbf{g}) \subseteq \mathbb{R}^p$, and $\mathbf{g}$ is continuous at $\mathbf{f}(\mathbf{x})$, then $\mathbf{g} \circ \mathbf{f}$ is continuous at $\mathbf{x}$. Let $\varepsilon > 0$ be given. Then there exists $\eta > 0$ such that if $|\mathbf{y} - \mathbf{f}(\mathbf{x})| < \eta$ and $\mathbf{y} \in D(\mathbf{g})$, it follows that $|\mathbf{g}(\mathbf{y}) - \mathbf{g}(\mathbf{f}(\mathbf{x}))| < \varepsilon$. It follows from continuity of $\mathbf{f}$ at $\mathbf{x}$ that there exists $\delta > 0$ such that if $|\mathbf{x} - \mathbf{z}| < \delta$ and $\mathbf{z} \in D(\mathbf{f})$, then $|\mathbf{f}(\mathbf{z}) - \mathbf{f}(\mathbf{x})| < \eta$. Then if $|\mathbf{x} - \mathbf{z}| < \delta$ and $\mathbf{z} \in D(\mathbf{g} \circ \mathbf{f}) \subseteq D(\mathbf{f})$, all the above hold and so

$$|\mathbf{g}(\mathbf{f}(\mathbf{z})) - \mathbf{g}(\mathbf{f}(\mathbf{x}))| < \varepsilon.$$

This proves part 3.)

Part 4.) says: If $\mathbf{f} = (f_1, \dots, f_q) : D(\mathbf{f}) \rightarrow \mathbb{R}^q$, then $\mathbf{f}$ is continuous if and only if each f_k is a continuous real valued function. Then

$$\begin{aligned}
|f_k(\mathbf{x}) - f_k(\mathbf{y})| &\leq |\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \\
&\equiv \left(\sum_{i=1}^q |f_i(\mathbf{x}) - f_i(\mathbf{y})|^2 \right)^{1/2} \\
&\leq \sum_{i=1}^q |f_i(\mathbf{x}) - f_i(\mathbf{y})|.
\end{aligned} \tag{17.9}$$

Suppose first that $\mathbf{f}$ is continuous at $\mathbf{x}$. Then there exists $\delta > 0$ such that if $|\mathbf{x} - \mathbf{y}| < \delta$, then $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$. The first part of the above inequality then shows that for

each $k = 1, \dots, q$, $|f_k(\mathbf{x}) - f_k(\mathbf{y})| < \varepsilon$. This shows the only if part. Now suppose each function, f_k is continuous. Then if $\varepsilon > 0$ is given, there exists $\delta_k > 0$ such that whenever $|\mathbf{x} - \mathbf{y}| < \delta_k$

$$|f_k(\mathbf{x}) - f_k(\mathbf{y})| < \varepsilon/q.$$

Now let $0 < \delta \leq \min(\delta_1, \dots, \delta_q)$. For $|\mathbf{x} - \mathbf{y}| < \delta$, the above inequality holds for all k and so the last part of 17.9 implies

$$\begin{aligned} |\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| &\leq \sum_{i=1}^q |f_i(\mathbf{x}) - f_i(\mathbf{y})| \\ &< \sum_{i=1}^q \frac{\varepsilon}{q} = \varepsilon. \end{aligned}$$

This proves part 4.)

To verify part 5.), let $\varepsilon > 0$ be given and let $\delta = \varepsilon$. Then if $|\mathbf{x} - \mathbf{y}| < \delta$, the triangle inequality implies

$$\begin{aligned} |f(\mathbf{x}) - f(\mathbf{y})| &= ||\mathbf{x}| - |\mathbf{y}|| \\ &\leq |\mathbf{x} - \mathbf{y}| < \delta = \varepsilon. \end{aligned}$$

This proves part 5.) and completes the proof of the theorem.

17.8.1 The Nested Interval Lemma

Here is a multidimensional version of the nested interval lemma.

Lemma 17.8.2 *Let $I_k = \prod_{i=1}^p [a_i^k, b_i^k] \equiv \{\mathbf{x} \in \mathbb{R}^p : x_i \in [a_i^k, b_i^k]\}$ and suppose that for all $k = 1, 2, \dots$,*

$$I_k \supseteq I_{k+1}.$$

Then there exists a point, $\mathbf{c} \in \mathbb{R}^p$ which is an element of every I_k .

Proof: Since $I_k \supseteq I_{k+1}$, it follows that for each $i = 1, \dots, p$, $[a_i^k, b_i^k] \supseteq [a_i^{k+1}, b_i^{k+1}]$. This implies that for each i ,

$$a_i^k \leq a_i^{k+1}, \quad b_i^k \geq b_i^{k+1}. \quad (17.10)$$

Consequently, if $k \leq l$,

$$a_i^l \leq b_i^l \leq b_i^k. \quad (17.11)$$

Now define

$$c_i \equiv \sup \{a_i^l : l = 1, 2, \dots\}$$

By the first inequality in 17.10,

$$c_i = \sup \{a_i^l : l = k, k+1, \dots\} \quad (17.12)$$

for each $k = 1, 2, \dots$. Therefore, picking any k , 17.11 shows that b_i^k is an upper bound for the set, $\{a_i^l : l = k, k+1, \dots\}$ and so it is at least as large as the least upper bound of this set which is the definition of c_i given in 17.12. Thus, for each i and each k ,

$$a_i^k \leq c_i \leq b_i^k.$$

Defining $\mathbf{c} \equiv (c_1, \dots, c_p)$, $\mathbf{c} \in I_k$ for all k . This proves the lemma.

If you don't like the proof, you could prove the lemma for the one variable case first and then do the following.

Lemma 17.8.3 Let $I_k = \prod_{i=1}^p [a_i^k, b_i^k] \equiv \{\mathbf{x} \in \mathbb{R}^p : x_i \in [a_i^k, b_i^k]\}$ and suppose that for all $k = 1, 2, \dots$,

$$I_k \supseteq I_{k+1}.$$

Then there exists a point, $\mathbf{c} \in \mathbb{R}^p$ which is an element of every I_k .

Proof: For each $i = 1, \dots, p$, $[a_i^k, b_i^k] \supseteq [a_i^{k+1}, b_i^{k+1}]$ and so by the nested interval theorem for one dimensional problems, there exists a point $c_i \in [a_i^k, b_i^k]$ for all k . Then letting $\mathbf{c} \equiv (c_1, \dots, c_p)$ it follows $\mathbf{c} \in I_k$ for all k . This proves the lemma.

17.8.2 The Extreme Value Theorem

Definition 17.8.4 A set, $C \subseteq \mathbb{R}^p$ is said to be **bounded** if $C \subseteq \prod_{i=1}^p [a_i, b_i]$ for some choice of intervals, $[a_i, b_i]$ where $-\infty < a_i < b_i < \infty$. The **diameter** of a set, S , is defined as

$$\text{diam}(S) \equiv \sup \{|\mathbf{x} - \mathbf{y}| : \mathbf{x}, \mathbf{y} \in S\}.$$

A function, $\mathbf{f}$ having values in $\mathbb{R}^p$ is said to be bounded if the set of values of $\mathbf{f}$ is a bounded set.

Thus $\text{diam}(S)$ is just a careful description of what you would think of as the diameter. It measures how stretched out the set is.

Lemma 17.8.5 Let $C \subseteq \mathbb{R}^p$ be closed and bounded and let $f : C \rightarrow \mathbb{R}$ be continuous. Then f is bounded.

Proof: Suppose not. Since C is bounded, it follows $C \subseteq \prod_{i=1}^p [a_i, b_i] \equiv I_0$ for some closed intervals, $[a_i, b_i]$. Consider all sets of the form $\prod_{i=1}^p [c_i, d_i]$ where $[c_i, d_i]$ equals either $[a_i, \frac{a_i+b_i}{2}]$ or $[c_i, d_i] = [\frac{a_i+b_i}{2}, b_i]$. Thus there are 2^p of these sets because there are two choices for the i^{th} slot for $i = 1, \dots, p$. Also, if $\mathbf{x}$ and $\mathbf{y}$ are two points in one of these sets,

$$|x_i - y_i| \leq 2^{-1} |b_i - a_i|.$$

Observe that $\text{diam}(I_0) = \left(\sum_{i=1}^p |b_i - a_i|^2\right)^{1/2}$ because for $\mathbf{x}, \mathbf{y} \in I_0$, $|x_i - y_i| \leq |a_i - b_i|$ for each $i = 1, \dots, p$,

$$\begin{aligned} |\mathbf{x} - \mathbf{y}| &= \left(\sum_{i=1}^p |x_i - y_i|^2\right)^{1/2} \\ &\leq 2^{-1} \left(\sum_{i=1}^p |b_i - a_i|^2\right)^{1/2} \equiv 2^{-1} \text{diam}(I_0). \end{aligned}$$

Denote by $\{J_1, \dots, J_{2^p}\}$ these sets determined above. It follows the diameter of each set is no larger than $2^{-1} \text{diam}(I_0)$. In particular, since $\mathbf{d} \equiv (d_1, \dots, d_p)$ and $\mathbf{c} \equiv (c_1, \dots, c_p)$ are two such points, for each J_k ,

$$\text{diam}(J_k) \equiv \left(\sum_{i=1}^p |d_i - c_i|^2\right)^{1/2} \leq 2^{-1} \text{diam}(I_0)$$

Since the union of these sets equals all of I_0 , it follows

$$C = \bigcup_{k=1}^{2^p} J_k \cap C.$$

If f is not bounded on C , it follows that for some k , f is not bounded on $J_k \cap C$. Let $I_1 \equiv J_k$ and let $C_1 = C \cap I_1$. Now do to I_1 and C_1 what was done to I_0 and C to obtain $I_2 \subseteq I_1$, and for $\mathbf{x}, \mathbf{y} \in I_2$,

$$|\mathbf{x} - \mathbf{y}| \leq 2^{-1} \text{diam}(I_1) \leq 2^{-2} \text{diam}(I_2),$$

and f is unbounded on $I_2 \cap C_1 \equiv C_2$. Continue in this way obtaining sets, I_k such that $I_k \supseteq I_{k+1}$ and $\text{diam}(I_k) \leq 2^{-k} \text{diam}(I_0)$ and f is unbounded on $I_k \cap C$. By the nested interval lemma, there exists a point, $\mathbf{c}$ which is contained in each I_k .

Claim: $\mathbf{c} \in C$.

Proof of claim: Suppose $\mathbf{c} \notin C$. Since C is a closed set, there exists $r > 0$ such that $B(\mathbf{c}, r)$ is contained completely in $\mathbb{R}^p \setminus C$. In other words, $B(\mathbf{c}, r)$ contains no points of C . Let k be so large that $\text{diam}(I_0) 2^{-k} < r$. Then since $\mathbf{c} \in I_k$, and any two points of I_k are closer than $\text{diam}(I_0) 2^{-k}$, I_k must be contained in $B(\mathbf{c}, r)$ and so has no points of C in it, contrary to the manner in which the I_k are defined in which f is unbounded on $I_k \cap C$. Therefore, $\mathbf{c} \in C$ as claimed.

Now for k large enough, and $\mathbf{x} \in C \cap I_k$, the continuity of f implies $|f(\mathbf{c}) - f(\mathbf{x})| < 1$ contradicting the manner in which I_k was chosen since this inequality implies f is bounded on $I_k \cap C$. This proves the theorem.

Here is a proof of the extreme value theorem.

Theorem 17.8.6 *Let C be closed and bounded and let $f : C \rightarrow \mathbb{R}$ be continuous. Then f achieves its maximum and its minimum on C . This means there exist, $\mathbf{x}_1, \mathbf{x}_2 \in C$ such that for all $\mathbf{x} \in C$,*

$$f(\mathbf{x}_1) \leq f(\mathbf{x}) \leq f(\mathbf{x}_2).$$

Proof: Let $M = \sup \{f(\mathbf{x}) : \mathbf{x} \in C\}$. Then by Lemma 17.8.5, M is a finite number. Is $f(\mathbf{x}_2) = M$ for some $\mathbf{x}_2$? if not, you could consider the function,

$$g(\mathbf{x}) \equiv \frac{1}{M - f(\mathbf{x})}$$

and g would be a continuous and unbounded function defined on C , contrary to Lemma 17.8.5. Therefore, there exists $\mathbf{x}_2 \in C$ such that $f(\mathbf{x}_2) = M$. A similar argument applies to show that the existence of $\mathbf{x}_1 \in C$ such that

$$f(\mathbf{x}_1) = \inf \{f(\mathbf{x}) : \mathbf{x} \in C\}.$$

This proves the theorem.

17.8.3 Sequences And Completeness

Definition 17.8.7 *A function whose domain is defined as a set of the form*

$$\{k, k+1, k+2, \dots\}$$

for k an integer is known as a sequence. Thus you can consider $f(k), f(k+1), f(k+2)$, etc. Usually the domain of the sequence is either $\mathbb{N}$, the natural numbers consisting of $\{1, 2, 3, \dots\}$ or the nonnegative integers, $\{0, 1, 2, 3, \dots\}$. Also, it is traditional to write f_1, f_2 , etc. instead of $f(1), f(2), f(3)$ etc. when referring to sequences. In the above context, f_k is called the first term, f_{k+1} the second and so forth. It is also common to write the sequence, not as f but as $\{f_i\}_{i=k}^{\infty}$ or just $\{f_i\}$ for short. The letter used for the name of the sequence is not important. Thus it is all right to let a be the name of a sequence or to refer to it as $\{a_i\}$. When the sequence has values in $\mathbb{R}^p$, it is customary to write it in bold face. Thus $\{\mathbf{a}_i\}$ would refer to a sequence having values in $\mathbb{R}^p$ for some $p > 1$.

Example 17.8.8 Let $\{a_k\}_{k=1}^{\infty}$ be defined by $a_k \equiv k^2 + 1$.

This gives a sequence. In fact, $a_7 = a(7) = 7^2 + 1 = 50$ just from using the formula for the k^{th} term of the sequence.

It is nice when sequences come in this way from a formula for the k^{th} term. However, this is often not the case. Sometimes sequences are defined recursively. This happens, when the first several terms of the sequence are given and then a rule is specified which determines a_{n+1} from knowledge of $a_1, \dots, a_n$. This rule which specifies a_{n+1} from knowledge of a_k for $k \leq n$ is known as a recurrence relation.

Example 17.8.9 Let $a_1 = 1$ and $a_2 = 1$. Assuming $a_1, \dots, a_{n+1}$ are known, $a_{n+2} \equiv a_n + a_{n+1}$.

Thus the first several terms of this sequence, listed in order, are 1, 1, 2, 3, 5, 8, $\dots$. This particular sequence is called the Fibonacci sequence and is important in the study of reproducing rabbits.

Example 17.8.10 Let $\mathbf{a}_k = (k, \sin(k))$. Thus this sequence has values in $\mathbb{R}^2$.

Definition 17.8.11 Let $\{\mathbf{a}_n\}$ be a sequence and let $n_1 < n_2 < n_3, \dots$ be any strictly increasing list of integers such that n_1 is at least as large as the first index used to define the sequence $\{\mathbf{a}_n\}$. Then if $\mathbf{b}_k \equiv \mathbf{a}_{n_k}$, $\{\mathbf{b}_k\}$ is called a subsequence of $\{\mathbf{a}_n\}$.

For example, suppose $a_n = (n^2 + 1)$. Thus $a_1 = 2$, $a_3 = 10$, etc. If

$$n_1 = 1, n_2 = 3, n_3 = 5, \dots, n_k = 2k - 1,$$

then letting $b_k = a_{n_k}$, it follows

$$b_k = ((2k - 1)^2 + 1) = 4k^2 - 4k + 2.$$

Definition 17.8.12 A sequence, $\{\mathbf{a}_k\}$ is said to **converge** to $\mathbf{a}$ if for every $\varepsilon > 0$ there exists n_ε such that if $n > n_\varepsilon$, then $|\mathbf{a} - \mathbf{a}_n| < \varepsilon$. The usual notation for this is $\lim_{n \rightarrow \infty} \mathbf{a}_n = \mathbf{a}$ although it is often written as $\mathbf{a}_n \rightarrow \mathbf{a}$.

The following theorem says the limit, if it exists, is unique.

Theorem 17.8.13 If a sequence, $\{\mathbf{a}_n\}$ converges to $\mathbf{a}$ and to $\mathbf{b}$ then $\mathbf{a} = \mathbf{b}$.

Proof: There exists n_ε such that if $n > n_\varepsilon$ then $|\mathbf{a}_n - \mathbf{a}| < \frac{\varepsilon}{2}$ and if $n > n_\varepsilon$, then $|\mathbf{a}_n - \mathbf{b}| < \frac{\varepsilon}{2}$. Then pick such an n .

$$|\mathbf{a} - \mathbf{b}| < |\mathbf{a} - \mathbf{a}_n| + |\mathbf{a}_n - \mathbf{b}| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since ε is arbitrary, this proves the theorem.

The following is the definition of a Cauchy sequence in $\mathbb{R}^p$.

Definition 17.8.14 $\{\mathbf{a}_n\}$ is a Cauchy sequence if for all $\varepsilon > 0$, there exists n_ε such that whenever $n, m \geq n_\varepsilon$,

$$|\mathbf{a}_n - \mathbf{a}_m| < \varepsilon.$$

A sequence is Cauchy means the terms are “bunching up to each other” as m, n get large.

Theorem 17.8.15 The set of terms in a Cauchy sequence in $\mathbb{R}^p$ is bounded in the sense that for all n , $|\mathbf{a}_n| < M$ for some $M < \infty$.

Proof: Let $\varepsilon = 1$ in the definition of a Cauchy sequence and let $n > n_1$. Then from the definition,

$$|\mathbf{a}_n - \mathbf{a}_{n_1}| < 1.$$

It follows that for all $n > n_1$,

$$|\mathbf{a}_n| < 1 + |\mathbf{a}_{n_1}|.$$

Therefore, for all n ,

$$|\mathbf{a}_n| \leq 1 + |\mathbf{a}_{n_1}| + \sum_{k=1}^{n_1} |\mathbf{a}_k|.$$

This proves the theorem.

Theorem 17.8.16 *If a sequence $\{\mathbf{a}_n\}$ in $\mathbb{R}^p$ converges, then the sequence is a Cauchy sequence. Also, if some subsequence of a Cauchy sequence converges, then the original sequence converges.*

Proof: Let $\varepsilon > 0$ be given and suppose $\mathbf{a}_n \rightarrow \mathbf{a}$. Then from the definition of convergence, there exists n_ε such that if $n > n_\varepsilon$, it follows that

$$|\mathbf{a}_n - \mathbf{a}| < \frac{\varepsilon}{2}$$

Therefore, if $m, n \geq n_\varepsilon + 1$, it follows that

$$|\mathbf{a}_n - \mathbf{a}_m| \leq |\mathbf{a}_n - \mathbf{a}| + |\mathbf{a} - \mathbf{a}_m| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

showing that, since $\varepsilon > 0$ is arbitrary, $\{\mathbf{a}_n\}$ is a Cauchy sequence. It remains to show that the last claim. Suppose then that $\{\mathbf{a}_n\}$ is a Cauchy sequence and $\mathbf{a} = \lim_{k \rightarrow \infty} \mathbf{a}_{n_k}$ where $\{\mathbf{a}_{n_k}\}_{k=1}^\infty$ is a subsequence. Let $\varepsilon > 0$ be given. Then there exists K such that if $k, l \geq K$, then $|\mathbf{a}_k - \mathbf{a}_l| < \frac{\varepsilon}{2}$. Then if $k > K$, it follows $n_k > K$ because $n_1, n_2, n_3, \dots$ is strictly increasing as the subscript increases. Also, there exists K_1 such that if $k > K_1$, $|\mathbf{a}_{n_k} - \mathbf{a}| < \frac{\varepsilon}{2}$. Then letting $n > \max(K, K_1)$, pick $k > \max(K, K_1)$. Then

$$|\mathbf{a} - \mathbf{a}_n| \leq |\mathbf{a} - \mathbf{a}_{n_k}| + |\mathbf{a}_{n_k} - \mathbf{a}_n| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

This proves the theorem.

Definition 17.8.17 *A set, K in $\mathbb{R}^p$ is said to be **sequentially compact** if every sequence in K has a subsequence which converges to a point of K .*

Theorem 17.8.18 *If $I_0 = \prod_{i=1}^p [a_i, b_i]$ where $a_i \leq b_i$, then I_0 is sequentially compact.*

Proof: Let $\{\mathbf{a}_i\}_{i=1}^\infty \subseteq I_0$ and consider all sets of the form $\prod_{i=1}^p [c_i, d_i]$ where $[c_i, d_i]$ equals either $[a_i, \frac{a_i+b_i}{2}]$ or $[c_i, d_i] = [\frac{a_i+b_i}{2}, b_i]$. Thus there are 2^p of these sets because there are two choices for the i^{th} slot for $i = 1, \dots, p$. Also, if $\mathbf{x}$ and $\mathbf{y}$ are two points in one of these sets,

$$|x_i - y_i| \leq 2^{-1} |b_i - a_i|.$$

$$\text{diam}(I_0) = \left(\sum_{i=1}^p |b_i - a_i|^2 \right)^{1/2},$$

$$\begin{aligned} |\mathbf{x} - \mathbf{y}| &= \left(\sum_{i=1}^p |x_i - y_i|^2 \right)^{1/2} \\ &\leq 2^{-1} \left(\sum_{i=1}^p |b_i - a_i|^2 \right)^{1/2} \equiv 2^{-1} \text{diam}(I_0). \end{aligned}$$

In particular, since $\mathbf{d} \equiv (d_1, \dots, d_p)$ and $\mathbf{c} \equiv (c_1, \dots, c_p)$ are two such points,

$$D_1 \equiv \left(\sum_{i=1}^p |d_i - c_i|^2 \right)^{1/2} \leq 2^{-1} \text{diam}(I_0)$$

Denote by $\{J_1, \dots, J_{2^p}\}$ these sets determined above. Since the union of these sets equals all of $I_0 \equiv I$, it follows that for some J_k , the sequence, $\{\mathbf{a}_i\}$ is contained in J_k for infinitely many k . Let that one be called I_1 . Next do for I_1 what was done for I_0 to get $I_2 \subseteq I_1$ such that the diameter is half that of I_1 and I_2 contains $\{\mathbf{a}_k\}$ for infinitely many values of k . Continue in this way obtaining a nested sequence of intervals, $\{I_k\}$ such that $I_k \supseteq I_{k+1}$, and if $\mathbf{x}, \mathbf{y} \in I_k$, then $|\mathbf{x} - \mathbf{y}| \leq 2^{-k} \text{diam}(I_0)$, and I_n contains $\{\mathbf{a}_k\}$ for infinitely many values of k for each n . Then by the nested interval lemma, there exists $\mathbf{c}$ such that $\mathbf{c}$ is contained in each I_k . Pick $\mathbf{a}_{n_1} \in I_1$. Next pick $n_2 > n_1$ such that $\mathbf{a}_{n_2} \in I_2$. If $\mathbf{a}_{n_1}, \dots, \mathbf{a}_{n_k}$ have been chosen, let $\mathbf{a}_{n_{k+1}} \in I_{k+1}$ and $n_{k+1} > n_k$. This can be done because in the construction, I_n contains $\{\mathbf{a}_k\}$ for infinitely many k . Thus the distance between $\mathbf{a}_{n_k}$ and $\mathbf{c}$ is no larger than $2^{-k} \text{diam}(I_0)$ and so $\lim_{k \rightarrow \infty} \mathbf{a}_{n_k} = \mathbf{c} \in I_0$. This proves the theorem.

Corollary 17.8.19 *Let K be a closed and bounded set of points in $\mathbb{R}^p$. Then K is sequentially compact.*

Proof: Since K is closed and bounded, there exists a closed rectangle, $\prod_{k=1}^p [a_k, b_k]$ which contains K . Now let $\{\mathbf{x}_k\}$ be a sequence of points in K . By Theorem 17.8.18 there exists a subsequence, $\{\mathbf{x}_{n_k}\}$ such that $\mathbf{x}_{n_k} \rightarrow \mathbf{x} \in \prod_{k=1}^p [a_k, b_k]$. However, K is closed and each of the points of the sequence is in K so $\mathbf{x} \in K$. This proves the corollary.

Theorem 17.8.20 *Every Cauchy sequence in $\mathbb{R}^p$ converges.*

Proof: Let $\{\mathbf{a}_k\}$ be a Cauchy sequence. By Theorem 17.8.15 there is some interval, $\prod_{i=1}^p [a_i, b_i]$ containing all the terms of $\{\mathbf{a}_k\}$. Therefore, by Theorem 17.8.18 a subsequence converges to a point of this interval. By Theorem 17.8.16 the original sequence converges. This proves the theorem.

17.8.4 Continuity And The Limit Of A Sequence

Just as in the case of a function of one variable, there is a very useful way of thinking of continuity in terms of limits of sequences found in the following theorem. In words, it says a function is continuous if it takes convergent sequences to convergent sequences whenever possible.

Theorem 17.8.21 *A function $\mathbf{f} : D(\mathbf{f}) \rightarrow \mathbb{R}^q$ is continuous at $\mathbf{x} \in D(\mathbf{f})$ if and only if, whenever $\mathbf{x}_n \rightarrow \mathbf{x}$ with $\mathbf{x}_n \in D(\mathbf{f})$, it follows $\mathbf{f}(\mathbf{x}_n) \rightarrow \mathbf{f}(\mathbf{x})$.*

Proof: Suppose first that $\mathbf{f}$ is continuous at $\mathbf{x}$ and let $\mathbf{x}_n \rightarrow \mathbf{x}$. Let $\varepsilon > 0$ be given. By continuity, there exists $\delta > 0$ such that if $|\mathbf{y} - \mathbf{x}| < \delta$, then $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$. However, there exists n_δ such that if $n \geq n_\delta$, then $|\mathbf{x}_n - \mathbf{x}| < \delta$ and so for all n this large,

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_n)| < \varepsilon$$

which shows $\mathbf{f}(\mathbf{x}_n) \rightarrow \mathbf{f}(\mathbf{x})$.

Now suppose the condition about taking convergent sequences to convergent sequences holds at $\mathbf{x}$. Suppose $\mathbf{f}$ fails to be continuous at $\mathbf{x}$. Then there exists $\varepsilon > 0$ and $\mathbf{x}_n \in D(\mathbf{f})$ such that $|\mathbf{x} - \mathbf{x}_n| < \frac{1}{n}$, yet

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_n)| \geq \varepsilon.$$

But this is clearly a contradiction because, although $\mathbf{x}_n \rightarrow \mathbf{x}$, $\mathbf{f}(\mathbf{x}_n)$ fails to converge to $\mathbf{f}(\mathbf{x})$. It follows $\mathbf{f}$ must be continuous after all. This proves the theorem.

As in the case of a function of one variable, there is a concept of uniform continuity.

Definition 17.8.22 A function $\mathbf{f} : D(\mathbf{f}) \rightarrow \mathbb{R}^q$ is uniformly continuous if for every $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $\mathbf{x}, \mathbf{y}$ are points of $D(\mathbf{f})$ such that $|\mathbf{x} - \mathbf{y}| < \delta$, it follows $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$.

Theorem 17.8.23 Let $\mathbf{f} : K \rightarrow \mathbb{R}^q$ be continuous at every point of K where K is a closed and bounded set in $\mathbb{R}^p$. Then $\mathbf{f}$ is uniformly continuous.

Proof: Suppose not. Then there exists $\varepsilon > 0$ and sequences $\{\mathbf{x}_j\}$ and $\{\mathbf{y}_j\}$ of points in K such that

$$|\mathbf{x}_j - \mathbf{y}_j| < \frac{1}{j}$$

but $|\mathbf{f}(\mathbf{x}_j) - \mathbf{f}(\mathbf{y}_j)| \geq \varepsilon$. Then by Corollary 17.8.19 on Page 428 which says K is sequentially compact, there is a subsequence $\{\mathbf{x}_{n_k}\}$ of $\{\mathbf{x}_j\}$ which converges to a point, $\mathbf{x} \in K$. Then since $|\mathbf{x}_{n_k} - \mathbf{y}_{n_k}| < \frac{1}{n_k}$, it follows that $\{\mathbf{y}_{n_k}\}$ also converges to $\mathbf{x}$. Therefore,

$$\varepsilon \leq \lim_{k \rightarrow \infty} |\mathbf{f}(\mathbf{x}_{n_k}) - \mathbf{f}(\mathbf{y}_{n_k})| = |\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x})| = 0,$$

a contradiction. Therefore, $\mathbf{f}$ is uniformly continuous as claimed.

17.9 Exercises

1. Suppose $\{\mathbf{x}_n\}$ is a sequence contained in a closed set C such that $\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}$. Show that $\mathbf{x} \in C$. **Hint:** Recall that a set is closed if and only if the complement of the set is open. That is if and only if $\mathbb{R}^n \setminus C$ is open.
2. Show using Problem 1 and Theorem 17.8.18 that every closed and bounded set is sequentially compact. **Hint:** If C is such a set, then $C \subseteq I_0 \equiv \prod_{i=1}^n [a_i, b_i]$. Now if $\{\mathbf{x}_n\}$ is a sequence in C , it must also be a sequence in I_0 . Apply Problem 1 and Theorem 17.8.18.
3. Prove the extreme value theorem, a continuous function achieves its maximum and minimum on any closed and bounded set, C , using the result of Problem 2. **Hint:** Suppose $\lambda = \sup \{f(\mathbf{x}) : \mathbf{x} \in C\}$. Then there exists $\{\mathbf{x}_n\} \subseteq C$ such that $f(\mathbf{x}_n) \rightarrow \lambda$. Now select a convergent subsequence using Problem 2. Do the same for the minimum.
4. Let C be a closed and bounded set and suppose $\mathbf{f} : C \rightarrow \mathbb{R}^m$ is continuous. Show that $\mathbf{f}$ must also be **uniformly continuous**. This means: For every $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $\mathbf{x}, \mathbf{y} \in C$ and $|\mathbf{x} - \mathbf{y}| < \delta$, it follows $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| < \varepsilon$. This is a good time to review the definition of continuity so you will see the difference. **Hint:** Suppose it is not so. Then there exists $\varepsilon > 0$ and $\{\mathbf{x}_k\}$ and $\{\mathbf{y}_k\}$ such that $|\mathbf{x}_k - \mathbf{y}_k| < \frac{1}{k}$ but $|\mathbf{f}(\mathbf{x}_k) - \mathbf{f}(\mathbf{y}_k)| \geq \varepsilon$. Now use Problem 2 to obtain a convergent subsequence.
5. From Problem 2 every closed and bounded set is sequentially compact. Are these the only sets which are sequentially compact? Explain.
6. A set whose elements are open sets, $\mathcal{C}$ is called an **open cover** of H if $\cup \mathcal{C} \supseteq H$. In other words, $\mathcal{C}$ is an open cover of H if every point of H is in at least one set of $\mathcal{C}$. Show that if $\mathcal{C}$ is an open cover of a closed and bounded set H then

there exists $\delta > 0$ such that whenever $\mathbf{x} \in H$, $B(\mathbf{x}, \delta)$ is contained in some set of $\mathcal{C}$. This number, δ is called a **Lebesgue number**. **Hint:** If there is no Lebesgue number for H , let $H \subseteq I = \prod_{i=1}^n [a_i, b_i]$. Use the process of chopping the intervals in half to get a sequence of nested intervals, I_k contained in I where $\text{diam}(I_k) \leq 2^{-k} \text{diam}(I)$ and there is no Lebesgue number for the open cover on $H_k \equiv H \cap I_k$. Now use the nested interval theorem to get $\mathbf{c}$ in all these H_k . For some $r > 0$ it follows $B(\mathbf{c}, r)$ is contained in some open set of U . But for large k , it must be that $H_k \subseteq B(\mathbf{c}, r)$ which contradicts the construction. You fill in the details.

7. A set is **compact** if for every open cover of the set, there exists a finite subset of the open cover which also covers the set. Show every closed and bounded set in $\mathbb{R}^p$ is compact. Next show that if a set in $\mathbb{R}^p$ is compact, then it must be closed and bounded. This is called the Heine Borel theorem.
8. Suppose S is a nonempty set in $\mathbb{R}^p$. Define

$$\text{dist}(\mathbf{x}, S) \equiv \inf \{|\mathbf{x} - \mathbf{y}| : \mathbf{y} \in S\}.$$

Show that

$$|\text{dist}(\mathbf{x}, S) - \text{dist}(\mathbf{y}, S)| \leq |\mathbf{x} - \mathbf{y}|.$$

Hint: Suppose $\text{dist}(\mathbf{x}, S) < \text{dist}(\mathbf{y}, S)$. If these are equal there is nothing to show. Explain why there exists $\mathbf{z} \in S$ such that $|\mathbf{x} - \mathbf{z}| < \text{dist}(\mathbf{x}, S) + \varepsilon$. Now explain why

$$|\text{dist}(\mathbf{x}, S) - \text{dist}(\mathbf{y}, S)| = \text{dist}(\mathbf{y}, S) - \text{dist}(\mathbf{x}, S) \leq |\mathbf{y} - \mathbf{z}| - (|\mathbf{x} - \mathbf{z}| - \varepsilon)$$

Now use the triangle inequality and observe that ε is arbitrary.

9. Suppose H is a closed set and $H \subseteq U \subseteq \mathbb{R}^p$, an open set. Show there exists a continuous function defined on $\mathbb{R}^p$, f such that $f(\mathbb{R}^p) \subseteq [0, 1]$, $f(\mathbf{x}) = 0$ if $\mathbf{x} \notin U$ and $f(\mathbf{x}) = 1$ if $\mathbf{x} \in H$. **Hint:** Try something like

$$\frac{\text{dist}(\mathbf{x}, U^C)}{\text{dist}(\mathbf{x}, U^C) + \text{dist}(\mathbf{x}, H)},$$

where $U^C \equiv \mathbb{R}^p \setminus U$, a closed set. You need to explain why the denominator is never equal to zero. The rest is supplied by Problem 8. This is a special case of a major theorem called Urysohn's lemma.

Vector Valued Functions Of One Variable

18.0.1 Outcomes

1. Identify a curve given its parametrization.
2. Determine combinations of vector functions such as sums, vector products, and scalar products.
3. Define limit, derivative, and integral for vector functions.
4. Evaluate limits, derivatives and integrals of vector functions.
5. Find the line tangent to a curve at a given point.
6. Recall, derive and apply rules to combinations of vector functions for the following:
 - (a) limits
 - (b) differentiation
 - (c) integration
7. Describe what is meant by arc length.
8. Evaluate the arc length of a curve.
9. Evaluate the work done by a varying force over a curved path.

18.1 Limits Of A Vector Valued Function Of One Variable

The above discussion considered expressions like

$$\frac{\mathbf{f}(t_0 + h) - \mathbf{f}(t_0)}{h}$$

and determined what they get close to as h gets small. In other words it is desired to consider

$$\lim_{h \rightarrow 0} \frac{\mathbf{f}(t_0 + h) - \mathbf{f}(t_0)}{h}$$

Specializing to functions of one variable, one can give a meaning to

$$\lim_{s \rightarrow t+} \mathbf{f}(s), \lim_{s \rightarrow t-} \mathbf{f}(s), \lim_{s \rightarrow \infty} \mathbf{f}(s),$$

and

$$\lim_{s \rightarrow -\infty} \mathbf{f}(s).$$

Definition 18.1.1 *In the case where $D(\mathbf{f})$ is only assumed to satisfy $D(\mathbf{f}) \supseteq (t, t+r)$,*

$$\lim_{s \rightarrow t+} \mathbf{f}(s) = \mathbf{L}$$

if and only if for all $\varepsilon > 0$ there exists $\delta > 0$ such that if

$$0 < s - t < \delta,$$

then

$$|\mathbf{f}(s) - \mathbf{L}| < \varepsilon.$$

In the case where $D(\mathbf{f})$ is only assumed to satisfy $D(\mathbf{f}) \supseteq (t-r, t)$,

$$\lim_{s \rightarrow t-} \mathbf{f}(s) = \mathbf{L}$$

if and only if for all $\varepsilon > 0$ there exists $\delta > 0$ such that if

$$0 < t - s < \delta,$$

then

$$|\mathbf{f}(s) - \mathbf{L}| < \varepsilon.$$

One can also consider limits as a variable “approaches” infinity. Of course nothing is “close” to infinity and so this requires a slightly different definition.

$$\lim_{t \rightarrow \infty} \mathbf{f}(t) = \mathbf{L}$$

if for every $\varepsilon > 0$ there exists l such that whenever $t > l$,

$$|\mathbf{f}(t) - \mathbf{L}| < \varepsilon \tag{18.1}$$

and

$$\lim_{t \rightarrow -\infty} \mathbf{f}(t) = \mathbf{L}$$

if for every $\varepsilon > 0$ there exists l such that whenever $t < l$, 18.1 holds.

Note that in all of this the definitions are identical to the case of scalar valued functions. The only difference is that here $|\cdot|$ refers to the norm or length in $\mathbb{R}^p$ where maybe $p > 1$.

Example 18.1.2 *Let $\mathbf{f}(t) = (\cos t, \sin t, t^2 + 1, \ln(t))$. Find $\lim_{t \rightarrow \pi/2} \mathbf{f}(t)$.*

Use Theorem 17.5.7 on Page 416 and the continuity of the functions to write this limit equals

$$\begin{aligned} & \left(\lim_{t \rightarrow \pi/2} \cos t, \lim_{t \rightarrow \pi/2} \sin t, \lim_{t \rightarrow \pi/2} (t^2 + 1), \lim_{t \rightarrow \pi/2} \ln(t) \right) \\ &= \left(0, 1, \ln\left(\frac{\pi^2}{4} + 1\right), \ln\left(\frac{\pi}{2}\right) \right). \end{aligned}$$

Example 18.1.3 *Let $\mathbf{f}(t) = (\frac{\sin t}{t}, t^2, t + 1)$. Find $\lim_{t \rightarrow 0} \mathbf{f}(t)$.*

Recall that $\lim_{t \rightarrow 0} \frac{\sin t}{t} = 1$. Then from Theorem 17.5.7 on Page 416, $\lim_{t \rightarrow 0} \mathbf{f}(t) = (1, 0, 1)$.

18.2 The Derivative And Integral

The following definition is on the derivative and integral of a vector valued function of one variable.

Definition 18.2.1 *The derivative of a function, $\mathbf{f}'(t)$, is defined as the following limit whenever the limit exists. If the limit does not exist, then neither does $\mathbf{f}'(t)$.*

$$\lim_{h \rightarrow 0} \frac{\mathbf{f}(t+h) - \mathbf{f}(t)}{h} \equiv \mathbf{f}'(t)$$

The function of h on the left is called the difference quotient just as it was for a scalar valued function. If $\mathbf{f}(t) = (f_1(t), \dots, f_p(t))$ and $\int_a^b f_i(t) dt$ exists for each $i = 1, \dots, p$, then $\int_a^b \mathbf{f}(t) dt$ is defined as the vector,

$$\left(\int_a^b f_1(t) dt, \dots, \int_a^b f_p(t) dt \right).$$

This is what is meant by saying $\mathbf{f} \in R([a, b])$.

This is exactly like the definition for a scalar valued function. As before,

$$\mathbf{f}'(x) = \lim_{y \rightarrow x} \frac{\mathbf{f}(y) - \mathbf{f}(x)}{y - x}.$$

As in the case of a scalar valued function, differentiability implies continuity but not the other way around.

Theorem 18.2.2 *If $\mathbf{f}'(t)$ exists, then $\mathbf{f}$ is continuous at t .*

Proof: Suppose $\varepsilon > 0$ is given and choose $\delta_1 > 0$ such that if $|h| < \delta_1$,

$$\left| \frac{\mathbf{f}(t+h) - \mathbf{f}(t)}{h} - \mathbf{f}'(t) \right| < 1.$$

then for such h , the triangle inequality implies

$$|\mathbf{f}(t+h) - \mathbf{f}(t)| < |h| + |\mathbf{f}'(t)| |h|.$$

Now letting $\delta < \min\left(\delta_1, \frac{\varepsilon}{1+|\mathbf{f}'(t)|}\right)$ it follows if $|h| < \delta$, then

$$|\mathbf{f}(t+h) - \mathbf{f}(t)| < \varepsilon.$$

Letting $y = h + t$, this shows that if $|y - t| < \delta$,

$$|\mathbf{f}(y) - \mathbf{f}(t)| < \varepsilon$$

which proves $\mathbf{f}$ is continuous at t . This proves the theorem.

As in the scalar case, there is a fundamental theorem of calculus.

Theorem 18.2.3 *If $\mathbf{f} \in R([a, b])$ and if $\mathbf{f}$ is continuous at $t \in (a, b)$, then*

$$\frac{d}{dt} \left(\int_a^t \mathbf{f}(s) ds \right) = \mathbf{f}(t).$$

Proof: Say $\mathbf{f}(t) = (f_1(t), \dots, f_p(t))$. Then it follows

$$\frac{1}{h} \int_a^{t+h} \mathbf{f}(s) ds - \frac{1}{h} \int_a^t \mathbf{f}(s) ds = \left(\frac{1}{h} \int_t^{t+h} f_1(s) ds, \dots, \frac{1}{h} \int_t^{t+h} f_p(s) ds \right)$$

and $\lim_{h \rightarrow 0} \frac{1}{h} \int_t^{t+h} f_i(s) ds = f_i(t)$ for each $i = 1, \dots, p$ from the fundamental theorem of calculus for scalar valued functions. Therefore,

$$\lim_{h \rightarrow 0} \frac{1}{h} \int_a^{t+h} \mathbf{f}(s) ds - \frac{1}{h} \int_a^t \mathbf{f}(s) ds = (f_1(t), \dots, f_p(t)) = \mathbf{f}(t)$$

and this proves the claim.

Example 18.2.4 Let $\mathbf{f}(x) = \mathbf{c}$ where $\mathbf{c}$ is a constant. Find $\mathbf{f}'(x)$.

The difference quotient,

$$\frac{\mathbf{f}(x+h) - \mathbf{f}(x)}{h} = \frac{\mathbf{c} - \mathbf{c}}{h} = \mathbf{0}$$

Therefore,

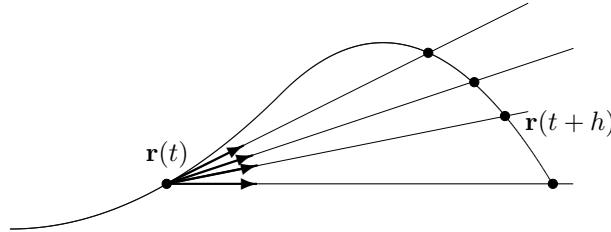
$$\lim_{h \rightarrow 0} \frac{\mathbf{f}(x+h) - \mathbf{f}(x)}{h} = \lim_{h \rightarrow 0} \mathbf{0} = \mathbf{0}$$

Example 18.2.5 Let $\mathbf{f}(t) = (at, bt)$ where a, b are constants. Find $\mathbf{f}'(t)$.

From the above discussion this derivative is just the vector valued functions whose components consist of the derivatives of the components of $\mathbf{f}$. Thus $\mathbf{f}'(t) = (a, b)$.

18.2.1 Geometric And Physical Significance Of The Derivative

Suppose $\mathbf{r}$ is a vector valued function of a parameter, t not necessarily time and consider the following picture of the points traced out by $\mathbf{r}$.



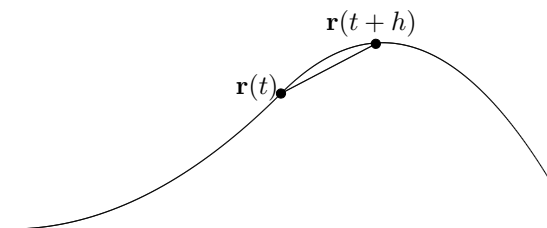
In this picture there are unit vectors in the direction of the vector from $\mathbf{r}(t)$ to $\mathbf{r}(t+h)$. You can see that it is reasonable to suppose these unit vectors, if they converge, converge to a unit vector, $\mathbf{T}$ which is tangent to the curve at the point $\mathbf{r}(t)$. Now each of these unit vectors is of the form

$$\frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{|\mathbf{r}(t+h) - \mathbf{r}(t)|} \equiv \mathbf{T}_h.$$

Thus $\mathbf{T}_h \rightarrow \mathbf{T}$, a unit tangent vector to the curve at the point $\mathbf{r}(t)$. Therefore,

$$\begin{aligned} \mathbf{r}'(t) &\equiv \lim_{h \rightarrow 0} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h} = \lim_{h \rightarrow 0} \frac{|\mathbf{r}(t+h) - \mathbf{r}(t)|}{h} \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{|\mathbf{r}(t+h) - \mathbf{r}(t)|} \\ &= \lim_{h \rightarrow 0} \frac{|\mathbf{r}(t+h) - \mathbf{r}(t)|}{h} \mathbf{T}_h = |\mathbf{r}'(t)| \mathbf{T}. \end{aligned}$$

In the case that t is time, the expression $|\mathbf{r}(t+h) - \mathbf{r}(t)|$ is a good approximation for the distance traveled by the object on the time interval $[t, t+h]$. The real distance would be the length of the curve joining the two points but if h is very small, this is essentially equal to $|\mathbf{r}(t+h) - \mathbf{r}(t)|$ as suggested by the picture below.



Therefore,

$$\frac{|\mathbf{r}(t+h) - \mathbf{r}(t)|}{h}$$

gives for small h , the approximate distance travelled on the time interval, $[t, t+h]$ divided by the length of time, h . Therefore, this expression is really the average speed of the object on this small time interval and so the limit as $h \rightarrow 0$, deserves to be called the instantaneous speed of the object. Thus $|\mathbf{r}'(t)| \mathbf{T}$ represents the speed times a unit direction vector, $\mathbf{T}$ which defines the direction in which the object is moving. Thus $\mathbf{r}'(t)$ is the velocity of the object. This is the physical significance of the derivative when t is time.

How do you go about computing $\mathbf{r}'(t)$? Letting $\mathbf{r}(t) = (r_1(t), \dots, r_q(t))$, the expression

$$\frac{\mathbf{r}(t_0+h) - \mathbf{r}(t_0)}{h} \quad (18.2)$$

is equal to

$$\left(\frac{r_1(t_0+h) - r_1(t_0)}{h}, \dots, \frac{r_q(t_0+h) - r_q(t_0)}{h} \right).$$

Then as h converges to 0, 18.2 converges to

$$\mathbf{v} \equiv (v_1, \dots, v_q)$$

where $v_k = r'_k(t)$. This by Theorem 17.5.7 on Page 416, which says that the term in 18.2 gets close to a vector, $\mathbf{v}$ if and only if all the coordinate functions of the term in 18.2 get close to the corresponding coordinate functions of $\mathbf{v}$.

In the case where t is time, this simply says the velocity vector equals the vector whose components are the derivatives of the components of the displacement vector, $\mathbf{r}(t)$.

In any case, the vector, $\mathbf{T}$ determines a direction vector which is tangent to the curve at the point, $\mathbf{r}(t)$ and so it is possible to find parametric equations for the line tangent to the curve at various points.

Example 18.2.6 Let $\mathbf{r}(t) = (\sin t, t^2, t+1)$ for $t \in [0, 5]$. Find a tangent line to the curve parameterized by $\mathbf{r}$ at the point $\mathbf{r}(2)$.

From the above discussion, a direction vector has the same direction as $\mathbf{r}'(2)$. Therefore, it suffices to simply use $\mathbf{r}'(2)$ as a direction vector for the line. $\mathbf{r}'(2) = (\cos 2, 4, 1)$. Therefore, a parametric equation for the tangent line is

$$(\sin 2, 4, 3) + t(\cos 2, 4, 1) = (x, y, z).$$

Example 18.2.7 Let $\mathbf{r}(t) = (\sin t, t^2, t + 1)$ for $t \in [0, 5]$. Find the velocity vector when $t = 1$.

From the above discussion, this is simply $\mathbf{r}'(1) = (\cos 1, 2, 1)$.

18.2.2 Differentiation Rules

There are rules which relate the derivative to the various operations done with vectors such as the dot product, the cross product, and vector addition and scalar multiplication.

Theorem 18.2.8 Let $a, b \in \mathbb{R}$ and suppose $\mathbf{f}'(t)$ and $\mathbf{g}'(t)$ exist. Then the following formulas are obtained.

$$(\mathbf{af} + \mathbf{bg})'(t) = a\mathbf{f}'(t) + b\mathbf{g}'(t). \quad (18.3)$$

$$(\mathbf{f} \cdot \mathbf{g})'(t) = \mathbf{f}'(t) \cdot \mathbf{g}(t) + \mathbf{f}(t) \cdot \mathbf{g}'(t) \quad (18.4)$$

If $\mathbf{f}, \mathbf{g}$ have values in $\mathbb{R}^3$, then

$$(\mathbf{f} \times \mathbf{g})'(t) = \mathbf{f}(t) \times \mathbf{g}'(t) + \mathbf{f}'(t) \times \mathbf{g}(t) \quad (18.5)$$

The formulas, 18.4, and 18.5 are referred to as the product rule.

Proof: The first formula is left for you to prove. Consider the second, 18.4.

$$\begin{aligned} & \lim_{h \rightarrow 0} \frac{\mathbf{f} \cdot \mathbf{g}(t+h) - \mathbf{f}\mathbf{g}(t)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\mathbf{f}(t+h) \cdot \mathbf{g}(t+h) - \mathbf{f}(t+h) \cdot \mathbf{g}(t)}{h} + \frac{\mathbf{f}(t+h) \cdot \mathbf{g}(t) - \mathbf{f}(t) \cdot \mathbf{g}(t)}{h} \\ &= \lim_{h \rightarrow 0} \left(\mathbf{f}(t+h) \cdot \frac{(\mathbf{g}(t+h) - \mathbf{g}(t))}{h} + \frac{(\mathbf{f}(t+h) - \mathbf{f}(t))}{h} \cdot \mathbf{g}(t) \right) \\ &= \lim_{h \rightarrow 0} \sum_{k=1}^n f_k(t+h) \frac{(g_k(t+h) - g_k(t))}{h} + \sum_{k=1}^n \frac{(f_k(t+h) - f_k(t))}{h} g_k(t) \\ &= \sum_{k=1}^n f_k(t) g'_k(t) + \sum_{k=1}^n f'_k(t) g_k(t) \\ &= \mathbf{f}'(t) \cdot \mathbf{g}(t) + \mathbf{f}(t) \cdot \mathbf{g}'(t). \end{aligned}$$

Formula 18.5 is left as an exercise which follows from the product rule and the definition of the cross product in terms of components given on Page 382.

Example 18.2.9 Let

$$\mathbf{r}(t) = (t^2, \sin t, \cos t)$$

and let $\mathbf{p}(t) = (t, \ln(t+1), 2t)$. Find $(\mathbf{r}(t) \times \mathbf{p}(t))'$.

From 18.5 this equals $(2t, \cos t, -\sin t) \times (t, \ln(t+1), 2t) + (t^2, \sin t, \cos t) \times \left(1, \frac{1}{t+1}, 2\right)$.

Example 18.2.10 Let $\mathbf{r}(t) = (t^2, \sin t, \cos t)$ Find $\int_0^\pi \mathbf{r}(t) dt$.

This equals $(\int_0^\pi t^2 dt, \int_0^\pi \sin t dt, \int_0^\pi \cos t dt) = (\frac{1}{3}\pi^3, 2, 0)$.

Example 18.2.11 An object has position $\mathbf{r}(t) = \left(t^3, \frac{t}{1+t}, \sqrt{t^2+2}\right)$ kilometers where t is given in hours. Find the velocity of the object in kilometers per hour when $t = 1$.

Recall the velocity at time t was $\mathbf{r}'(t)$. Therefore, find $\mathbf{r}'(t)$ and plug in $t = 1$ to find the velocity.

$$\begin{aligned}\mathbf{r}'(t) &= \left(3t^2, \frac{1(1+t)-t}{(1+t)^2}, \frac{1}{2}(t^2+2)^{-1/2} 2t\right) \\ &= \left(3t^2, \frac{1}{(1+t)^2}, \frac{1}{\sqrt{t^2+2}}t\right)\end{aligned}$$

When $t = 1$, the velocity is

$$\mathbf{r}'(1) = \left(3, \frac{1}{4}, \frac{1}{\sqrt{3}}\right) \text{ kilometers per hour.}$$

Obviously, this can be continued. That is, you can consider the possibility of taking the derivative of the derivative and then the derivative of that and so forth. The main thing to consider about this is the notation and it is exactly like it was in the case of a scalar valued function presented earlier. Thus $\mathbf{r}''(t)$ denotes the second derivative.

When you are given a vector valued function of one variable, sometimes it is possible to give a simple description of the curve which results. Usually it is not possible to do this!

Example 18.2.12 Describe the curve which results from the vector valued function, $\mathbf{r}(t) = (\cos 2t, \sin 2t, t)$ where $t \in \mathbb{R}$.

The first two components indicate that for $\mathbf{r}(t) = (x(t), y(t), z(t))$, the pair, $(x(t), y(t))$ traces out a circle. While it is doing so, $z(t)$ is moving at a steady rate in the positive direction. Therefore, the curve which results is a cork skew shaped thing called a helix.

As an application of the theorems for differentiating curves, here is an interesting application. It is also a situation where the curve can be identified as something familiar.

Example 18.2.13 Sound waves have the angle of incidence equal to the angle of reflection. Suppose you are in a large room and you make a sound. The sound waves spread out and you would expect your sound to be inaudible very far away. But what if the room were shaped so that the sound is reflected off the wall toward a single point, possibly far away from you? Then you might have the interesting phenomenon of someone far away hearing what you said quite clearly. How should the room be designed?

Suppose you are located at the point $\mathbf{P}_0$ and the point where your sound is to be reflected is $\mathbf{P}_1$. Consider a plane which contains the two points and let $\mathbf{r}(t)$ denote a parametrization of the intersection of this plane with the walls of the room. Then the condition that the angle of reflection equals the angle of incidence reduces to saying the angle between $\mathbf{P}_0 - \mathbf{r}(t)$ and $-\mathbf{r}'(t)$ equals the angle between $\mathbf{P}_1 - \mathbf{r}(t)$ and $\mathbf{r}'(t)$. Draw a picture to see this. Therefore,

$$\frac{(\mathbf{P}_0 - \mathbf{r}(t)) \cdot (-\mathbf{r}'(t))}{|\mathbf{P}_0 - \mathbf{r}(t)| |\mathbf{r}'(t)|} = \frac{(\mathbf{P}_1 - \mathbf{r}(t)) \cdot (\mathbf{r}'(t))}{|\mathbf{P}_1 - \mathbf{r}(t)| |\mathbf{r}'(t)|}.$$

This reduces to

$$\frac{(\mathbf{r}(t) - \mathbf{P}_0) \cdot (-\mathbf{r}'(t))}{|\mathbf{r}(t) - \mathbf{P}_0|} = \frac{(\mathbf{r}(t) - \mathbf{P}_1) \cdot (\mathbf{r}'(t))}{|\mathbf{r}(t) - \mathbf{P}_1|} \quad (18.6)$$

Now

$$\frac{(\mathbf{r}(t) - \mathbf{P}_1) \cdot (\mathbf{r}'(t))}{|\mathbf{r}(t) - \mathbf{P}_1|} = \frac{d}{dt} |\mathbf{r}(t) - \mathbf{P}_1|$$

and a similar formula holds for $\mathbf{P}_1$ replaced with $\mathbf{P}_0$. This is because

$$|\mathbf{r}(t) - \mathbf{P}_1| = \sqrt{(\mathbf{r}(t) - \mathbf{P}_1) \cdot (\mathbf{r}(t) - \mathbf{P}_1)}$$

and so using the chain rule and product rule,

$$\begin{aligned} \frac{d}{dt} |\mathbf{r}(t) - \mathbf{P}_1| &= \frac{1}{2} ((\mathbf{r}(t) - \mathbf{P}_1) \cdot (\mathbf{r}(t) - \mathbf{P}_1))^{-1/2} 2((\mathbf{r}(t) - \mathbf{P}_1) \cdot \mathbf{r}'(t)) \\ &= \frac{(\mathbf{r}(t) - \mathbf{P}_1) \cdot (\mathbf{r}'(t))}{|\mathbf{r}(t) - \mathbf{P}_1|}. \end{aligned}$$

Therefore, from 18.6,

$$\frac{d}{dt} (|\mathbf{r}(t) - \mathbf{P}_1|) + \frac{d}{dt} (|\mathbf{r}(t) - \mathbf{P}_0|) = 0$$

showing that $|\mathbf{r}(t) - \mathbf{P}_1| + |\mathbf{r}(t) - \mathbf{P}_0| = C$ for some constant, C . This implies the curve of intersection of the plane with the room is an ellipse having $\mathbf{P}_0$ and $\mathbf{P}_1$ as the foci.

18.2.3 Leibniz's Notation

Leibniz's notation also generalizes routinely. For example, $\frac{dy}{dt} = \mathbf{y}'(t)$ with other similar notations holding.

18.3 Exercises

- Find the following limits if possible

$$(a) \lim_{x \rightarrow 0+} \left(\frac{|x|}{x}, \sin x/x, \cos x \right)$$

$$(b) \lim_{x \rightarrow 0+} \left(\frac{x}{|x|}, \sec x, e^x \right)$$

$$(c) \lim_{x \rightarrow 4} \left(\frac{x^2-16}{x+4}, x+7, \frac{\tan 4x}{5x} \right)$$

$$(d) \lim_{x \rightarrow \infty} \left(\frac{x}{1+x^2}, \frac{x^2}{1+x^2}, \frac{\sin x^2}{x} \right)$$

- Find

$$\lim_{x \rightarrow 2} \left(\frac{x^2-4}{x+2}, x^2+2x-1, \frac{x^2-4}{x-2} \right).$$

- Prove from the definition that $\lim_{x \rightarrow a} (\sqrt[3]{x}, x+1) = (\sqrt[3]{a}, a+1)$ for all $a \in \mathbb{R}$. **Hint:** You might want to use the formula for the difference of two cubes,

$$a^3 - b^3 = (a-b)(a^2 + ab + b^2).$$

- Let

$$\mathbf{r}(t) = (4+t^2, \sqrt{t^2+1}t^3, t^3)$$

describe the position of an object in $\mathbb{R}^3$ as a function of t where t is measured in seconds and $\mathbf{r}(t)$ is measured in meters. Is the velocity of this object ever equal to zero? If so, find the value of t at which this occurs and the point in $\mathbb{R}^3$ at which the velocity is zero.

- Let $\mathbf{r}(t) = (\sin 2t, t^2, 2t+1)$ for $t \in [0, 4]$. Find a tangent line to the curve parameterized by $\mathbf{r}$ at the point $\mathbf{r}(2)$.

- Let $\mathbf{r}(t) = (t, \sin t^2, t+1)$ for $t \in [0, 5]$. Find a tangent line to the curve parameterized by $\mathbf{r}$ at the point $\mathbf{r}(2)$.

- Let $\mathbf{r}(t) = (\sin t, t^2, \cos(t^2))$ for $t \in [0, 5]$. Find a tangent line to the curve parameterized by $\mathbf{r}$ at the point $\mathbf{r}(2)$.

- Let $\mathbf{r}(t) = (\sin t, \cos(t^2), t+1)$ for $t \in [0, 5]$. Find the velocity when $t = 3$.

9. Let $\mathbf{r}(t) = (\sin t, t^2, t+1)$ for $t \in [0, 5]$. Find the velocity when $t = 3$.
10. Let $\mathbf{r}(t) = (t, \ln(t^2 + 1), t+1)$ for $t \in [0, 5]$. Find the velocity when $t = 3$.
11. Suppose an object has position $\mathbf{r}(t) \in \mathbb{R}^3$ where $\mathbf{r}$ is differentiable and suppose also that $|\mathbf{r}'(t)| = c$ where c is a constant.
 - (a) Show first that this condition does not require $\mathbf{r}(t)$ to be a constant. **Hint:** You can do this either mathematically or by giving a physical example.
 - (b) Show that you can conclude that $\mathbf{r}'(t) \cdot \mathbf{r}(t) = 0$. That is, the velocity is always perpendicular to the displacement.
12. Prove 18.5 from the component description of the cross product.
13. Prove 18.5 from the formula $(\mathbf{f} \times \mathbf{g})_i = \varepsilon_{ijk} f_j g_k$.
14. Prove 18.5 directly from the definition of the derivative without considering components.
15. A bezier curve in $\mathbb{R}^p$ is a vector valued function of the form

$$\mathbf{y}(t) = \sum_{k=0}^n \binom{n}{k} \mathbf{x}_k (1-t)^{n-k} t^k$$
 where here the $\binom{n}{k}$ are the binomial coefficients and $\mathbf{x}_k$ are $n+1$ points in $\mathbb{R}^n$. Show that $\mathbf{y}(0) = \mathbf{x}_0$, $\mathbf{y}(1) = \mathbf{x}_n$, and find $\mathbf{y}'(0)$ and $\mathbf{y}'(1)$. Recall that $\binom{n}{0} = \binom{n}{n} = 1$ and $\binom{n}{n-1} = \binom{n}{1} = n$. Curves of this sort are important in various computer programs.
16. Suppose $\mathbf{r}(t)$, $\mathbf{s}(t)$, and $\mathbf{p}(t)$ are three differentiable functions of t which have values in $\mathbb{R}^3$. Find a formula for $(\mathbf{r}(t) \times \mathbf{s}(t) \cdot \mathbf{p}(t))'$.
17. If $\mathbf{r}'(t) = \mathbf{0}$ for all $t \in (a, b)$, show that there exists a constant vector, $\mathbf{c}$ such that $\mathbf{r}(t) = \mathbf{c}$ for all $t \in (a, b)$.
18. If $\mathbf{F}'(t) = \mathbf{f}(t)$ for all $t \in (a, b)$ and $\mathbf{F}$ is continuous on $[a, b]$, show that $\int_a^b \mathbf{f}(t) dt = \mathbf{F}(b) - \mathbf{F}(a)$.
19. Verify that if $\boldsymbol{\Omega} \times \mathbf{u} = \mathbf{0}$ for all $\mathbf{u}$, then $\boldsymbol{\Omega} = \mathbf{0}$.

18.4 Line Integrals

The concept of the integral can be extended to functions which are not defined on an interval of the real line but on some curve in $\mathbb{R}^n$. This is done by defining things in such a way that the more general concept reduces to the earlier notion. First it is necessary to consider what is meant by arc length.

18.4.1 Arc Length And Orientations

The application of the integral considered here is the concept of the **length of a curve**.

Definition 18.4.1 C is a **smooth curve** in $\mathbb{R}^n$ if there exists an interval, $[a, b] \subseteq \mathbb{R}$ and functions $x_i : [a, b] \rightarrow \mathbb{R}$ such that the following conditions hold

1. x_i is continuous on $[a, b]$.
2. x'_i exists and is continuous and bounded on $[a, b]$, with $x'_i(a)$ defined as the derivative from the right,

$$\lim_{h \rightarrow 0+} \frac{x_i(a+h) - x_i(a)}{h},$$

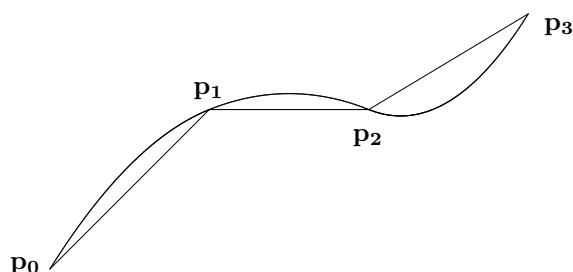
and $x'_i(b)$ defined similarly as the derivative from the left.

3. For $\mathbf{p}(t) \equiv (x_1(t), \dots, x_n(t))$, $t \rightarrow \mathbf{p}(t)$ is one to one on (a, b) .

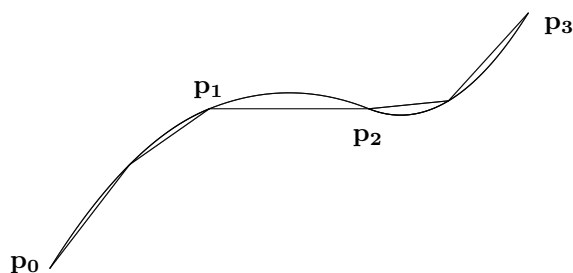
$$4. |\mathbf{p}'(t)| \equiv \left(\sum_{i=1}^n |x'_i(t)|^2 \right)^{1/2} \neq 0 \text{ for all } t \in [a, b].$$

$$5. C = \cup \{ (x_1(t), \dots, x_n(t)) : t \in [a, b] \}.$$

The functions, $x_i(t)$, defined above are giving the coordinates of a point in $\mathbb{R}^n$ and the list of these functions is called a **parametrization** for the smooth curve. Note the natural direction of the interval also gives a direction for moving along the curve. Such a direction is called an orientation. The integral is used to define what is meant by the length of such a smooth curve. Consider such a smooth curve having parametrization $(x_1, \dots, x_n)$. Forming a partition of $[a, b]$, $a = t_0 < \dots < t_n = b$ and letting $\mathbf{p}_i = (x_1(t_i), \dots, x_n(t_i))$, you could consider the polygon formed by lines from $\mathbf{p}_0$ to $\mathbf{p}_1$ and from $\mathbf{p}_1$ to $\mathbf{p}_2$ and from $\mathbf{p}_3$ to $\mathbf{p}_4$ etc. to be an approximation to the curve, C . The following picture illustrates what is meant by this.



Now consider what happens when the partition is refined by including more points. You can see from the following picture that the polygonal approximation would appear to be even better and that as more points are added in the partition, the sum of the lengths of the line segments seems to get close to something which deserves to be defined as the length of the curve, C .



Thus the length of the curve is approximated by

$$\sum_{k=1}^n |\mathbf{p}(t_k) - \mathbf{p}(t_{k-1})|.$$

Since the functions in the parametrization are differentiable, it is reasonable to expect this to be close to

$$\sum_{k=1}^n |\mathbf{p}'(t_{k-1})| (t_k - t_{k-1})$$

which is seen to be a Riemann sum for the integral

$$\int_a^b |\mathbf{p}'(t)| dt$$

and it is this integral which is defined as the length of the curve.

Would the same length be obtained if another parametrization were used? This is a very important question because the length of the curve should depend only on the curve itself and not on the method used to trace out the curve. The answer to this question is that the length of the curve does not depend on parametrization. The proof is somewhat technical so is given in the last section of this chapter.

Does the definition of length given above correspond to the usual definition of length in the case when the curve is a line segment? It is easy to see that it does so by considering two points in $\mathbb{R}^n$, $\mathbf{p}$ and $\mathbf{q}$. A parametrization for the line segment joining these two points is

$$f_i(t) \equiv tp_i + (1-t)q_i, \quad t \in [0, 1].$$

Using the definition of length of a smooth curve just given, the length according to this definition is

$$\int_0^1 \left(\sum_{i=1}^n (p_i - q_i)^2 \right)^{1/2} dt = |\mathbf{p} - \mathbf{q}|.$$

Thus this new definition which is valid for smooth curves which may not be straight line segments gives the usual length for straight line segments.

The proof that curve length is well defined for a smooth curve contains a result which deserves to be stated as a corollary. It is proved in Lemma 18.6.13 on Page 451 but the proof is mathematically fairly advanced so it is presented later.

Corollary 18.4.2 *Let C be a smooth curve and let $\mathbf{f} : [a, b] \rightarrow C$ and $\mathbf{g} : [c, d] \rightarrow C$ be two parameterizations satisfying 1 - 5. Then $\mathbf{g}^{-1} \circ \mathbf{f}$ is either strictly increasing or strictly decreasing.*

Definition 18.4.3 *If $\mathbf{g}^{-1} \circ \mathbf{f}$ is increasing, then $\mathbf{f}$ and $\mathbf{g}$ are said to be equivalent parameterizations and this is written as $\mathbf{f} \sim \mathbf{g}$. It is also said that the two parameterizations give the same orientation for the curve when $\mathbf{f} \sim \mathbf{g}$.*

When the parameterizations are equivalent, they preserve the direction, of motion along the curve and this also shows there are exactly two orientations of the curve since either $\mathbf{g}^{-1} \circ \mathbf{f}$ is increasing or it is decreasing. This is not hard to believe. In simple language, the message is that there are exactly two directions of motion along a curve. The difficulty is in proving this is actually the case.

Lemma 18.4.4 *The following hold for $\sim$.*

$$\mathbf{f} \sim \mathbf{f}, \quad (18.7)$$

$$\text{If } \mathbf{f} \sim \mathbf{g} \text{ then } \mathbf{g} \sim \mathbf{f}, \quad (18.8)$$

$$\text{If } \mathbf{f} \sim \mathbf{g} \text{ and } \mathbf{g} \sim \mathbf{h}, \text{ then } \mathbf{f} \sim \mathbf{h}. \quad (18.9)$$

Proof: Formula 18.7 is obvious because $\mathbf{f}^{-1} \circ \mathbf{f}(t) = t$ so it is clearly an increasing function. If $\mathbf{f} \sim \mathbf{g}$ then $\mathbf{f}^{-1} \circ \mathbf{g}$ is increasing. Now $\mathbf{g}^{-1} \circ \mathbf{f}$ must also be increasing because it is the inverse of $\mathbf{f}^{-1} \circ \mathbf{g}$. This verifies 18.8. To see 18.9, $\mathbf{f}^{-1} \circ \mathbf{h} = (\mathbf{f}^{-1} \circ \mathbf{g}) \circ (\mathbf{g}^{-1} \circ \mathbf{h})$ and so since both of these functions are increasing, it follows $\mathbf{f}^{-1} \circ \mathbf{h}$ is also increasing. This proves the lemma.

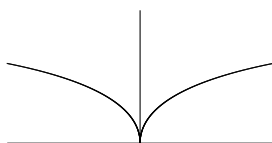
The symbol, $\sim$ is called an equivalence relation. If C is such a smooth curve just described, and if $\mathbf{f} : [a, b] \rightarrow C$ is a parameterization of C , consider $\mathbf{g}(t) \equiv \mathbf{f}((a+b)-t)$, also a parameterization of C . Now by Corollary 18.4.2, if $\mathbf{h}$ is a parametrization, then if $\mathbf{f}^{-1} \circ \mathbf{h}$ is not increasing, it must be the case that $\mathbf{g}^{-1} \circ \mathbf{h}$ is increasing. Consequently, either $\mathbf{h} \sim \mathbf{g}$ or $\mathbf{h} \sim \mathbf{f}$. These parameterizations, $\mathbf{h}$, which satisfy $\mathbf{h} \sim \mathbf{f}$ are called the

equivalence class determined by $\mathbf{f}$ and those $\mathbf{h} \sim \mathbf{g}$ are called the equivalence class determined by $\mathbf{g}$. These two classes are called **orientations** of C . They give the direction of motion on C . You see that going from $\mathbf{f}$ to $\mathbf{g}$ corresponds to tracing out the curve in the opposite direction.

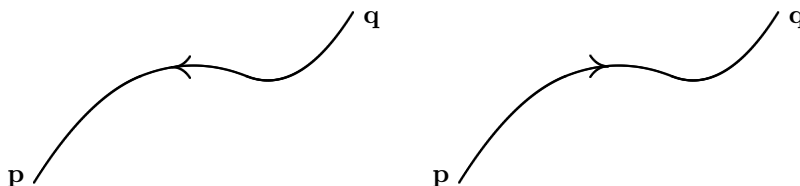
Sometimes people wonder why it is required, in the definition of a smooth curve that $\mathbf{p}'(t) \neq \mathbf{0}$. Imagine t is time and $\mathbf{p}(t)$ gives the location of a point in space. If $\mathbf{p}'(t)$ is allowed to equal zero, the point can stop and change directions abruptly, producing a pointy place in C . Here is an example.

Example 18.4.5 Graph the curve (t^3, t^2) for $t \in [-1, 1]$.

In this case, $t = x^{1/3}$ and so $y = x^{2/3}$. Thus the graph of this curve looks like the picture below. Note the pointy place. Such a curve should not be considered smooth.



So what is the thing to remember from all this? First, there are certain conditions which must be satisfied for a curve to be smooth. These are listed above. Next, if you have any curve, there are two directions you can move over this curve, each called an orientation. This is illustrated in the following picture.



Either you move from $\mathbf{p}$ to $\mathbf{q}$ or you move from $\mathbf{q}$ to $\mathbf{p}$.

Definition 18.4.6 A curve C is piecewise smooth if there exist points on this curve, $\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_n$ such that, denoting $C_{\mathbf{p}_{k-1}\mathbf{p}_k}$ the part of the curve joining $\mathbf{p}_{k-1}$ and $\mathbf{p}_k$, it follows $C_{\mathbf{p}_{k-1}\mathbf{p}_k}$ is a smooth curve and $\cup_{k=1}^n C_{\mathbf{p}_{k-1}\mathbf{p}_k} = C$. In other words, it is piecewise smooth if it consists of a finite number of smooth curves linked together.

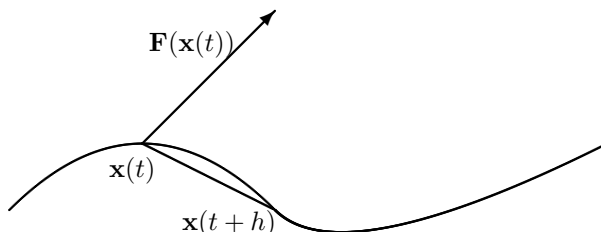
Note that Example 18.4.5 is an example of a piecewise smooth curve although it is not smooth.

18.4.2 Line Integrals And Work

Let C be a smooth curve contained in $\mathbb{R}^p$. A curve, C is an “**oriented curve**” if the only parameterizations considered are those which lie in exactly one of the two equivalence classes, each of which is called an “**orientation**”. In simple language, orientation specifies a direction over which motion along the curve is to take place. Thus, it specifies the order in which the points of C are encountered. The pair of concepts consisting of the set of points making up the curve along with a direction of motion along the curve is called an **oriented curve**.

Definition 18.4.7 Suppose $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^p$ is given for each $\mathbf{x} \in C$ where C is a smooth oriented curve and suppose $\mathbf{x} \rightarrow \mathbf{F}(\mathbf{x})$ is continuous. The mapping $\mathbf{x} \rightarrow \mathbf{F}(\mathbf{x})$ is called a **vector field**. In the case that $\mathbf{F}(\mathbf{x})$ is a force, it is called a **force field**.

Next the concept of work done by a force field, $\mathbf{F}$ on an object as it moves along the curve, C , in the direction determined by the given orientation of the curve will be defined. This is new. Earlier the work done by a force which acts on an object moving in a straight line was discussed but here the object moves over a curve. In order to define what is meant by the work, consider the following picture.



In this picture, the work done by a constant force, $\mathbf{F}$ on an object which moves from the point $\mathbf{x}(t)$ to the point $\mathbf{x}(t+h)$ along the straight line shown would equal $\mathbf{F} \cdot (\mathbf{x}(t+h) - \mathbf{x}(t))$. It is reasonable to assume this would be a good approximation to the work done in moving along the curve joining $\mathbf{x}(t)$ and $\mathbf{x}(t+h)$ provided h is small enough. Also, provided h is small,

$$\mathbf{x}(t+h) - \mathbf{x}(t) \approx \mathbf{x}'(t)h$$

where the wriggly equal sign indicates the two quantities are close. In the notation of Leibniz, one writes dt for h and

$$dW = \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t) dt$$

or in other words,

$$\frac{dW}{dt} = \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t).$$

Defining the total work done by the force at $t = 0$, corresponding to the first endpoint of the curve, to equal zero, the work would satisfy the following initial value problem.

$$\frac{dW}{dt} = \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t), \quad W(a) = 0.$$

This motivates the following definition of work.

Definition 18.4.8 Let $\mathbf{F}(\mathbf{x})$ be given above. Then the work done by this force field on an object moving over the curve C in the direction determined by the specified orientation is defined as

$$\int_C \mathbf{F} \cdot d\mathbf{R} \equiv \int_a^b \mathbf{F}(\mathbf{x}(t)) \cdot \mathbf{x}'(t) dt$$

where the function, $\mathbf{x}$ is one of the allowed parameterizations of C in the given orientation of C . In other words, there is an interval, $[a, b]$ and as t goes from a to b , $\mathbf{x}(t)$ moves in the direction determined from the given orientation of the curve.

Theorem 18.4.9 The symbol, $\int_C \mathbf{F} \cdot d\mathbf{R}$, is well defined in the sense that every parametrization in the given orientation of C gives the same value for $\int_C \mathbf{F} \cdot d\mathbf{R}$.

Proof: Suppose $\mathbf{g} : [c, d] \rightarrow C$ is another allowed parametrization. Thus $\mathbf{g}^{-1} \circ \mathbf{f}$ is an increasing function, ϕ . Then since ϕ is increasing,

$$\int_c^d \mathbf{F}(\mathbf{g}(s)) \cdot \mathbf{g}'(s) ds = \int_a^b \mathbf{F}(\mathbf{g}(\phi(t))) \cdot \mathbf{g}'(\phi(t)) \phi'(t) dt$$

$$= \int_a^b \mathbf{F}(\mathbf{f}(t)) \cdot \frac{d}{dt} (\mathbf{g}(\mathbf{g}^{-1} \circ \mathbf{f}(t))) dt = \int_a^b \mathbf{F}(\mathbf{f}(t)) \cdot \mathbf{f}'(t) dt.$$

This proves the theorem.

Regardless the physical interpretation of $\mathbf{F}$, this is called the **line integral**. When $\mathbf{F}$ is interpreted as a force, the line integral measures the extent to which the motion over the curve in the indicated direction is aided by the force. If the net effect of the force on the object is to impede rather than to aid the motion, this will show up as the work being negative.

Does the concept of work as defined here coincide with the earlier concept of work when the object moves over a straight line when acted on by a constant force?

Let $\mathbf{p}$ and $\mathbf{q}$ be two points in $\mathbb{R}^n$ and suppose $\mathbf{F}$ is a constant force acting on an object which moves from $\mathbf{p}$ to $\mathbf{q}$ along the straight line joining these points. Then the work done is $\mathbf{F} \cdot (\mathbf{q} - \mathbf{p})$. Is the same thing obtained from the above definition? Let $\mathbf{x}(t) \equiv \mathbf{p} + t(\mathbf{q} - \mathbf{p})$, $t \in [0, 1]$ be a parameterization for this oriented curve, the straight line in the direction from $\mathbf{p}$ to $\mathbf{q}$. Then $\mathbf{x}'(t) = \mathbf{q} - \mathbf{p}$ and $\mathbf{F}(\mathbf{x}(t)) = \mathbf{F}$. Therefore, the above definition yields

$$\int_0^1 \mathbf{F} \cdot (\mathbf{q} - \mathbf{p}) dt = \mathbf{F} \cdot (\mathbf{q} - \mathbf{p}).$$

Therefore, the new definition adds to but does not contradict the old one.

Example 18.4.10 Suppose for $t \in [0, \pi]$ the position of an object is given by $\mathbf{r}(t) = t\mathbf{i} + \cos(2t)\mathbf{j} + \sin(2t)\mathbf{k}$. Also suppose there is a force field defined on $\mathbb{R}^3$, $\mathbf{F}(x, y, z) \equiv 2xy\mathbf{i} + x^2\mathbf{j} + \mathbf{k}$. Find

$$\int_C \mathbf{F} \cdot d\mathbf{R}$$

where C is the curve traced out by this object which has the orientation determined by the direction of increasing t .

To find this line integral use the above definition and write

$$\begin{aligned} \int_C \mathbf{F} \cdot d\mathbf{R} &= \int_0^\pi (2t(\cos(2t)), t^2, 1) \cdot \\ &\quad (1, -2\sin(2t), 2\cos(2t)) dt \end{aligned}$$

In evaluating this replace the x in the formula for $\mathbf{F}$ with t , the y in the formula for $\mathbf{F}$ with $\cos(2t)$ and the z in the formula for $\mathbf{F}$ with $\sin(2t)$ because these are the values of these variables which correspond to the value of t . Taking the dot product, this equals the following integral.

$$\int_0^\pi (2t \cos 2t - 2(\sin 2t)t^2 + 2 \cos 2t) dt = \pi^2$$

Example 18.4.11 Let C denote the oriented curve obtained by $\mathbf{r}(t) = (t, \sin t, t^3)$ where the orientation is determined by increasing t for $t \in [0, 2]$. Also let $\mathbf{F} = (x, y, xz + z)$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$.

You use the definition.

$$\begin{aligned} \int_C \mathbf{F} \cdot d\mathbf{R} &= \int_0^2 (t, \sin(t), (t+1)t^3) \cdot (1, \cos(t), 3t^2) dt \\ &= \int_0^2 (t + \sin(t)\cos(t) + 3(t+1)t^5) dt \\ &= \frac{1251}{14} - \frac{1}{2} \cos^2(2). \end{aligned}$$

Suppose you have a curve specified by $\mathbf{r}(s) = (x(s), y(s), z(s))$ and it has the property that $|\mathbf{r}'(s)| = 1$ for all $s \in [0, b]$. Then the length of this curve for s between 0 and s_1 is

$$\int_0^{s_1} |\mathbf{r}'(s)| ds = \int_0^{s_1} 1 ds = s_1.$$

This parameter is therefore called arc length because the length of the curve up to s equals s . Now you can always change the parameter to be arc length.

Proposition 18.4.12 *Suppose C is an oriented smooth curve parameterized by $\mathbf{r}(t)$ for $t \in [a, b]$. Then letting l denote the total length of C , there exists $\mathbf{R}(s)$, $s \in [0, l]$ another parameterization for this curve which preserves the orientation and such that $|\mathbf{R}'(s)| = 1$ so that s is arc length.*

Prove: Let $\phi(t) \equiv \int_a^t |\mathbf{r}'(\tau)| d\tau \equiv s$. Then s is an increasing function of t because

$$\frac{ds}{dt} = \phi'(t) = |\mathbf{r}'(t)| > 0.$$

Now define $\mathbf{R}(s) \equiv \mathbf{r}(\phi^{-1}(s))$. Then

$$\begin{aligned} \mathbf{R}'(s) &= \mathbf{r}'(\phi^{-1}(s)) (\phi^{-1})'(s) \\ &= \frac{\mathbf{r}'(\phi^{-1}(s))}{|\mathbf{r}'(\phi^{-1}(s))|} \end{aligned}$$

and so $|\mathbf{R}'(s)| = 1$ as claimed. $\mathbf{R}(l) = \mathbf{r}(\phi^{-1}(l)) = \mathbf{r}\left(\phi^{-1}\left(\int_a^b |\mathbf{r}'(\tau)| d\tau\right)\right) = \mathbf{r}(b)$ and $\mathbf{R}(0) = \mathbf{r}(\phi^{-1}(0)) = \mathbf{r}(a)$ and $\mathbf{R}$ delivers the same set of points in the same order as $\mathbf{r}$ because $\frac{ds}{dt} > 0$.

The arc length parameter is just like any other parameter, in so far as considerations of line integrals are concerned, because it was shown above that line integrals are independent of parametrization. However, when things are defined in terms of the arc length parametrization, it is clear they depend only on geometric properties of the curve itself and for this reason, the arc length parametrization is important in differential geometry.

18.4.3 Another Notation For Line Integrals

Definition 18.4.13 *Let $\mathbf{F}(x, y, z) = (P(x, y, z), Q(x, y, z), R(x, y, z))$ and let C be an oriented curve. Then another way to write $\int_C \mathbf{F} \cdot d\mathbf{R}$ is*

$$\int_C Pdx + Qdy + Rdz$$

This last is referred to as the integral of a **differential form**, $Pdx + Qdy + Rdz$. The study of differential forms is important. Formally, $d\mathbf{R} = (dx, dy, dz)$ and so the integrand in the above is formally $\mathbf{F} \cdot d\mathbf{R}$. Other occurrences of this notation are handled similarly in 2 or higher dimensions.

18.5 Exercises

1. Suppose for $t \in [0, \pi]$ the position of an object is given by $\mathbf{r}(t) = t\mathbf{i} + \cos(2t)\mathbf{j} + \sin(2t)\mathbf{k}$. Also suppose there is a force field defined on $\mathbb{R}^3$,

$$\mathbf{F}(x, y, z) \equiv 2xy\mathbf{i} + (x^2 + 2zy)\mathbf{j} + y^2\mathbf{k}.$$

Find the work,

$$\int_C \mathbf{F} \cdot d\mathbf{R}$$

where C is the curve traced out by this object which has the orientation determined by the direction of increasing t .

2. Here is a vector field, $(y, x + z^2, 2yz)$ and here is the parametrization of a curve, C . $\mathbf{R}(t) = (\cos 2t, 2 \sin 2t, t)$ where t goes from 0 to $\pi/4$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$.
3. If f and g are both increasing functions, show that $f \circ g$ is an increasing function also. Assume anything you like about the domains of the functions.
4. Suppose for $t \in [0, 3]$ the position of an object is given by $\mathbf{r}(t) = t\mathbf{i} + t\mathbf{j} + t\mathbf{k}$. Also suppose there is a force field defined on $\mathbb{R}^3$, $\mathbf{F}(x, y, z) \equiv yz\mathbf{i} + xz\mathbf{j} + xy\mathbf{k}$. Find

$$\int_C \mathbf{F} \cdot d\mathbf{R}$$

where C is the curve traced out by this object which has the orientation determined by the direction of increasing t . Repeat the problem for $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j} + t\mathbf{k}$.

5. Suppose for $t \in [0, 1]$ the position of an object is given by $\mathbf{r}(t) = t\mathbf{i} + t\mathbf{j} + t\mathbf{k}$. Also suppose there is a force field defined on $\mathbb{R}^3$, $\mathbf{F}(x, y, z) \equiv z\mathbf{i} + xz\mathbf{j} + xy\mathbf{k}$. Find

$$\int_C \mathbf{F} \cdot d\mathbf{R}$$

where C is the curve traced out by this object which has the orientation determined by the direction of increasing t . Repeat the problem for $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j} + t\mathbf{k}$.

6. Let $\mathbf{F}(x, y, z)$ be a given force field and suppose it acts on an object having mass, m on a curve with parametrization, $(x(t), y(t), z(t))$ for $t \in [a, b]$. Show directly that the work done equals the difference in the kinetic energy. **Hint:**

$$\begin{aligned} \int_a^b \mathbf{F}(x(t), y(t), z(t)) \cdot (x'(t), y'(t), z'(t)) dt = \\ \int_a^b m(x''(t), y''(t), z''(t)) \cdot (x'(t), y'(t), z'(t)) dt, \end{aligned}$$

etc.

7. Suppose for $t \in [0, 2\pi]$ the position of an object is given by

$$\mathbf{r}(t) = 2t\mathbf{i} + \cos(t)\mathbf{j} + \sin(t)\mathbf{k}.$$

Also suppose there is a force field defined on $\mathbb{R}^3$,

$$\mathbf{F}(x, y, z) \equiv 2xy\mathbf{i} + (x^2 + 2zy)\mathbf{j} + y^2\mathbf{k}.$$

Find the work,

$$\int_C \mathbf{F} \cdot d\mathbf{R}$$

where C is the curve traced out by this object which has the orientation determined by the direction of increasing t .

8. Here is a vector field, $(y, x^2 + z, 2yz)$ and here is the parametrization of a curve, C . $\mathbf{R}(t) = (\cos 2t, 2 \sin 2t, t)$ where t goes from 0 to $\pi/4$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$.

9. Suppose for $t \in [0, 1]$ the position of an object is given by $\mathbf{r}(t) = t\mathbf{i} + t\mathbf{j} + t\mathbf{k}$. Also suppose there is a force field defined on $\mathbb{R}^3$, $\mathbf{F}(x, y, z) \equiv yz\mathbf{i} + xz\mathbf{j} + xy\mathbf{k}$. Find

$$\int_C \mathbf{F} \cdot d\mathbf{R}$$

where C is the curve traced out by this object which has the orientation determined by the direction of increasing t . Repeat the problem for $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j} + t\mathbf{k}$.

You should get the same answer in this case. This is because the vector field happens to be conservative. (More on this later.)

18.6 Independence Of Parametrization*



Recall that if $\mathbf{p}(t) : t \in [a, b]$ was a parametrization of a smooth curve, C , the length of C is defined as

$$\int_a^b |\mathbf{p}'(t)| dt$$

If some other parametrization were used to trace out C , would the same answer be obtained? To answer this question in a satisfactory manner requires some hard calculus.

18.6.1 Hard Calculus

Definition 18.6.1 A sequence $\{a_n\}_{n=1}^\infty$ converges to a ,

$$\lim_{n \rightarrow \infty} a_n = a \text{ or } a_n \rightarrow a$$

if and only if for every $\varepsilon > 0$ there exists n_ε such that whenever $n \geq n_\varepsilon$,

$$|a_n - a| < \varepsilon.$$

In words the definition says that given any measure of closeness, ε , the terms of the sequence are eventually all this close to a . Note the similarity with the concept of limit. Here, the word “eventually” refers to n being sufficiently large. Earlier, it referred to y being sufficiently close to x on one side or another or else x being sufficiently large in either the positive or negative directions. The limit of a sequence, if it exists, is unique.

Theorem 18.6.2 If $\lim_{n \rightarrow \infty} a_n = a$ and $\lim_{n \rightarrow \infty} a_n = a_1$ then $a_1 = a$.

Proof: Suppose $a_1 \neq a$. Then let $0 < \varepsilon < |a_1 - a|/2$ in the definition of the limit. It follows there exists n_ε such that if $n \geq n_\varepsilon$, then $|a_n - a| < \varepsilon$ and $|a_n - a_1| < \varepsilon$. Therefore, for such n ,

$$\begin{aligned} |a_1 - a| &\leq |a_1 - a_n| + |a_n - a| \\ &< \varepsilon + \varepsilon < |a_1 - a|/2 + |a_1 - a|/2 = |a_1 - a|, \end{aligned}$$

a contradiction.

Definition 18.6.3 Let $\{a_n\}$ be a sequence and let $n_1 < n_2 < n_3, \dots$ be any strictly increasing list of integers such that n_1 is at least as large as the first index used to define the sequence $\{a_n\}$. Then if $b_k \equiv a_{n_k}$, $\{b_k\}$ is called a subsequence of $\{a_n\}$.

Theorem 18.6.4 Let $\{x_n\}$ be a sequence with $\lim_{n \rightarrow \infty} x_n = x$ and let $\{x_{n_k}\}$ be a subsequence. Then $\lim_{k \rightarrow \infty} x_{n_k} = x$.

Proof: Let $\varepsilon > 0$ be given. Then there exists n_ε such that if $n > n_\varepsilon$, then $|x_n - x| < \varepsilon$. Suppose $k > n_\varepsilon$. Then $n_k \geq k > n_\varepsilon$ and so

$$|x_{n_k} - x| < \varepsilon$$

showing $\lim_{k \rightarrow \infty} x_{n_k} = x$ as claimed.

There is a very useful way of thinking of continuity in terms of limits of sequences found in the following theorem. In words, it says a function is continuous if it takes convergent sequences to convergent sequences whenever possible.

Theorem 18.6.5 A function $f : D(f) \rightarrow \mathbb{R}$ is continuous at $x \in D(f)$ if and only if, whenever $x_n \rightarrow x$ with $x_n \in D(f)$, it follows $f(x_n) \rightarrow f(x)$.

Proof: Suppose first that f is continuous at x and let $x_n \rightarrow x$. Let $\varepsilon > 0$ be given. By continuity, there exists $\delta > 0$ such that if $|y - x| < \delta$, then $|f(y) - f(x)| < \varepsilon$. However, there exists n_δ such that if $n \geq n_\delta$, then $|x_n - x| < \delta$ and so for all n this large,

$$|f(x) - f(x_n)| < \varepsilon$$

which shows $f(x_n) \rightarrow f(x)$.

Now suppose the condition about taking convergent sequences to convergent sequences holds at x . Suppose f fails to be continuous at x . Then there exists $\varepsilon > 0$ and $x_n \in D(f)$ such that $|x - x_n| < \frac{1}{n}$, yet

$$|f(x) - f(x_n)| \geq \varepsilon.$$

But this is clearly a contradiction because, although $x_n \rightarrow x$, $f(x_n)$ fails to converge to $f(x)$. It follows f must be continuous after all. This proves the theorem.

Definition 18.6.6 A set, $K \subseteq \mathbb{R}$ is sequentially compact if whenever $\{a_n\} \subseteq K$ is a sequence, there exists a subsequence, $\{a_{n_k}\}$ such that this subsequence converges to a point of K .

The following theorem is part of a major advanced calculus theorem known as the Heine Borel theorem.

Theorem 18.6.7 Every closed interval, $[a, b]$ is sequentially compact.

Proof: Let $\{x_n\} \subseteq [a, b] \equiv I_0$. Consider the two intervals $[a, \frac{a+b}{2}]$ and $[\frac{a+b}{2}, b]$ each of which has length $(b-a)/2$. At least one of these intervals contains x_n for infinitely many values of n . Call this interval I_1 . Now do for I_1 what was done for I_0 . Split it in half and let I_2 be the interval which contains x_n for infinitely many values of n . Continue this way obtaining a sequence of nested intervals $I_0 \supseteq I_1 \supseteq I_2 \supseteq I_3 \dots$ where the length of I_n is $(b-a)/2^n$. Now pick n_1 such that $x_{n_1} \in I_1$, n_2 such that $n_2 > n_1$ and $x_{n_2} \in I_2$, n_3 such that $n_3 > n_2$ and $x_{n_3} \in I_3$, etc. (This can be done because in each case the intervals contained x_n for infinitely many values of n .) By the nested interval lemma there exists a point, c contained in all these intervals. Furthermore,

$$|x_{n_k} - c| < (b-a)2^{-k}$$

and so $\lim_{k \rightarrow \infty} x_{n_k} = c \in [a, b]$. This proves the theorem.

Lemma 18.6.8 *Let $\phi : [a, b] \rightarrow \mathbb{R}$ be a continuous function and suppose ϕ is 1-1 on (a, b) . Then ϕ is either strictly increasing or strictly decreasing on $[a, b]$. Furthermore, ϕ^{-1} is continuous.*

Proof: First it is shown that ϕ is either strictly increasing or strictly decreasing on (a, b) .

If ϕ is not strictly decreasing on (a, b) , then there exists $x_1 < y_1$, $x_1, y_1 \in (a, b)$ such that

$$(\phi(y_1) - \phi(x_1))(y_1 - x_1) > 0.$$

If for some other pair of points, $x_2 < y_2$ with $x_2, y_2 \in (a, b)$, the above inequality does not hold, then since ϕ is 1-1,

$$(\phi(y_2) - \phi(x_2))(y_2 - x_2) < 0.$$

Let $x_t \equiv tx_1 + (1-t)x_2$ and $y_t \equiv ty_1 + (1-t)y_2$. Then $x_t < y_t$ for all $t \in [0, 1]$ because

$$tx_1 \leq ty_1 \text{ and } (1-t)x_2 \leq (1-t)y_2$$

with strict inequality holding for at least one of these inequalities since not both t and $(1-t)$ can equal zero. Now define

$$h(t) \equiv (\phi(y_t) - \phi(x_t))(y_t - x_t).$$

Since h is continuous and $h(0) < 0$, while $h(1) > 0$, there exists $t \in (0, 1)$ such that $h(t) = 0$. Therefore, both x_t and y_t are points of (a, b) and $\phi(y_t) - \phi(x_t) = 0$ contradicting the assumption that ϕ is one to one. It follows ϕ is either strictly increasing or strictly decreasing on (a, b) .

This property of being either strictly increasing or strictly decreasing on (a, b) carries over to $[a, b]$ by the continuity of ϕ . Suppose ϕ is strictly increasing on (a, b) , a similar argument holding for ϕ strictly decreasing on (a, b) . If $x > a$, then pick $y \in (a, x)$ and from the above, $\phi(y) < \phi(x)$. Now by continuity of ϕ at a ,

$$\phi(a) = \lim_{x \rightarrow a+} \phi(x) \leq \phi(y) < \phi(x).$$

Therefore, $\phi(a) < \phi(x)$ whenever $x \in (a, b)$. Similarly $\phi(b) > \phi(x)$ for all $x \in (a, b)$.

It only remains to verify ϕ^{-1} is continuous. Suppose then that $s_n \rightarrow s$ where s_n and s are points of $\phi([a, b])$. It is desired to verify that $\phi^{-1}(s_n) \rightarrow \phi^{-1}(s)$. If this does not happen, there exists $\varepsilon > 0$ and a subsequence, still denoted by s_n such that $|\phi^{-1}(s_n) - \phi^{-1}(s)| \geq \varepsilon$. Using the sequential compactness of $[a, b]$ there exists a further subsequence, still denoted by n , such that $\phi^{-1}(s_n) \rightarrow t_1 \in [a, b]$, $t_1 \neq \phi^{-1}(s)$. Then by continuity of ϕ , it follows $s_n \rightarrow \phi(t_1)$ and so $s = \phi(t_1)$. Therefore, $t_1 = \phi^{-1}(s)$ after all. This proves the lemma.

Corollary 18.6.9 *Let $f : (a, b) \rightarrow \mathbb{R}$ be one to one and continuous. Then $f(a, b)$ is an open interval, (c, d) and $f^{-1} : (c, d) \rightarrow (a, b)$ is continuous.*

Proof: Since f is either strictly increasing or strictly decreasing, it follows that $f(a, b)$ is an open interval, (c, d) . Assume f is decreasing. Now let $x \in (a, b)$. Why is f^{-1} continuous at $f(x)$? Since f is decreasing, if $f(x) < f(y)$, then $y \equiv f^{-1}(f(y)) < x \equiv f^{-1}(f(x))$ and so f^{-1} is also decreasing. Let $\varepsilon > 0$ be given. Let $\varepsilon > \eta > 0$ and $(x - \eta, x + \eta) \subseteq (a, b)$. Then $f(x) \in (f(x + \eta), f(x - \eta))$. Let

$$\delta = \min(f(x) - f(x + \eta), f(x - \eta) - f(x)).$$

Then if

$$|f(z) - f(x)| < \delta,$$

it follows

$$z \equiv f^{-1}(f(z)) \in (x - \eta, x + \eta) \subseteq (x - \varepsilon, x + \varepsilon)$$

so

$$|f^{-1}(f(z)) - x| = |f^{-1}(f(z)) - f^{-1}(f(x))| < \varepsilon.$$

This proves the theorem in the case where f is strictly decreasing. The case where f is increasing is similar.

Theorem 18.6.10 *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous and one to one. Suppose $f'(x_1)$ exists for some $x_1 \in [a, b]$ and $f'(x_1) \neq 0$. Then $(f^{-1})'(f(x_1))$ exists and is given by the formula, $(f^{-1})'(f(x_1)) = \frac{1}{f'(x_1)}$.*

Proof: By Lemma 18.6.8 f is either strictly increasing or strictly decreasing and f^{-1} is continuous on $[a, b]$. Therefore there exists $\eta > 0$ such that if $0 < |f(x_1) - f(x)| < \eta$, then

$$0 < |x_1 - x| = |f^{-1}(f(x_1)) - f^{-1}(f(x))| < \delta$$

where δ is small enough that for $0 < |x_1 - x| < \delta$,

$$\left| \frac{x - x_1}{f(x) - f(x_1)} - \frac{1}{f'(x_1)} \right| < \varepsilon.$$

It follows that if $0 < |f(x_1) - f(x)| < \eta$,

$$\left| \frac{f^{-1}(f(x)) - f^{-1}(f(x_1))}{f(x) - f(x_1)} - \frac{1}{f'(x_1)} \right| = \left| \frac{x - x_1}{f(x) - f(x_1)} - \frac{1}{f'(x_1)} \right| < \varepsilon$$

Therefore, since $\varepsilon > 0$ is arbitrary,

$$\lim_{y \rightarrow f(x_1)} \frac{f^{-1}(y) - f^{-1}(f(x_1))}{y - f(x_1)} = \frac{1}{f'(x_1)}$$

and this proves the theorem.

The following obvious corollary comes from the above by not bothering with end points.

Corollary 18.6.11 *Let $f : (a, b) \rightarrow \mathbb{R}$ be continuous and one to one. Suppose $f'(x_1)$ exists for some $x_1 \in (a, b)$ and $f'(x_1) \neq 0$. Then $(f^{-1})'(f(x_1))$ exists and is given by the formula, $(f^{-1})'(f(x_1)) = \frac{1}{f'(x_1)}$.*

This is one of those theorems which is very easy to remember if you neglect the difficult questions and simply focus on formal manipulations. Consider the following.

$$f^{-1}(f(x)) = x.$$

Now use the chain rule on both sides to write

$$(f^{-1})'(f(x)) f'(x) = 1,$$

and then divide both sides by $f'(x)$ to obtain

$$(f^{-1})'(f(x)) = \frac{1}{f'(x)}.$$

Of course this gives the conclusion of the above theorem rather effortlessly and it is formal manipulations like this which aid in remembering formulas such as the one given in the theorem.

18.6.2 Independence Of Parametrization

Theorem 18.6.12 *Let $\phi : [a, b] \rightarrow [c, d]$ be one to one and suppose ϕ' exists and is continuous on $[a, b]$. Then if f is a continuous function defined on $[a, b]$ which is Riemann integrable¹,*

$$\int_c^d f(s) ds = \int_a^b f(\phi(t)) |\phi'(t)| dt$$

Proof: Let $F'(s) = f(s)$. (For example, let $F(s) = \int_a^s f(r) dr$.) Then the first integral equals $F(d) - F(c)$ by the fundamental theorem of calculus. By Lemma 2.15.6, ϕ is either strictly increasing or strictly decreasing. Suppose ϕ is strictly decreasing. Then $\phi(a) = d$ and $\phi(b) = c$. Therefore, $\phi' \leq 0$ and the second integral equals

$$\begin{aligned} - \int_a^b f(\phi(t)) \phi'(t) dt &= \int_b^a \frac{d}{dt} (F(\phi(t))) dt \\ &= F(\phi(a)) - F(\phi(b)) = F(d) - F(c). \end{aligned}$$

The case when ϕ is increasing is similar but easier. This proves the theorem.

Lemma 18.6.13 *Let $\mathbf{f} : [a, b] \rightarrow C$, $\mathbf{g} : [c, d] \rightarrow C$ be parameterizations of a smooth curve which satisfy conditions 1 - 5. Then $\phi(t) \equiv \mathbf{g}^{-1} \circ \mathbf{f}(t)$ is 1-1 on (a, b) , continuous on $[a, b]$, and either strictly increasing or strictly decreasing on $[a, b]$.*

Proof: It is obvious ϕ is 1-1 on (a, b) from the conditions $\mathbf{f}$ and $\mathbf{g}$ satisfy. It only remains to verify continuity on $[a, b]$ because then the final claim follows from Lemma 18.6.8. If ϕ is not continuous on $[a, b]$, then there exists a sequence, $\{t_n\} \subseteq [a, b]$ such that $t_n \rightarrow t$ but $\phi(t_n)$ fails to converge to $\phi(t)$. Therefore, for some $\varepsilon > 0$ there exists a subsequence, still denoted by n such that $|\phi(t_n) - \phi(t)| \geq \varepsilon$. Using the sequential compactness of $[c, d]$, (See Theorem 18.6.7 on Page 448.) there is a further subsequence, still denoted by n such that $\{\phi(t_n)\}$ converges to a point, s , of $[c, d]$ which is not equal to $\phi(t)$. Thus $\mathbf{g}^{-1} \circ \mathbf{f}(t_n) \rightarrow s$ and still $t_n \rightarrow t$. Therefore, the continuity of $\mathbf{f}$ and $\mathbf{g}$ imply $\mathbf{f}(t_n) \rightarrow \mathbf{f}(t)$ and $\mathbf{f}(t_n) \rightarrow \mathbf{f}(t)$. Therefore, $\mathbf{g}(s) = \mathbf{f}(t)$ and so $s = \mathbf{g}^{-1} \circ \mathbf{f}(t) = \phi(t)$, a contradiction. Therefore, ϕ is continuous as claimed.

Theorem 18.6.14 *The length of a smooth curve is not dependent on which parametrization is used.*

Proof: Let C be the curve and suppose $\mathbf{f} : [a, b] \rightarrow C$ and $\mathbf{g} : [c, d] \rightarrow C$ both satisfy conditions 1 - 5. Is it true that $\int_a^b |\mathbf{f}'(t)| dt = \int_c^d |\mathbf{g}'(s)| ds$?

Let $\phi(t) \equiv \mathbf{g}^{-1} \circ \mathbf{f}(t)$ for $t \in [a, b]$. I want to show that ϕ is C^1 on an interval of the form $[a + \delta, b - \delta]$. By the above lemma, ϕ is either strictly increasing or strictly decreasing on $[a, b]$. Suppose for the sake of simplicity that it is strictly increasing. The decreasing case is handled similarly.

Let $s_0 \in \phi([a + \delta, b - \delta]) \subset (c, d)$. Then by assumption 4, $g'_i(s_0) \neq 0$ for some i . By continuity of g'_i , it follows $g'_i(s) \neq 0$ for all $s \in I$ where I is an open interval contained in $[c, d]$ which contains s_0 . It follows from the mean value theorem that on this interval, g_i is either strictly increasing or strictly decreasing. Therefore, $J \equiv g_i(I)$ is also an open interval and you can define a differentiable function, $h_i : J \rightarrow I$ by

$$h_i(g_i(s)) = s.$$

This implies that for $s \in I$,

$$h'_i(g_i(s)) = \frac{1}{g'_i(s)}. \quad (18.10)$$

¹Recall that all continuous functions of this sort are Riemann integrable.

Now letting $s = \phi(t)$ for $s \in I$, it follows $t \in J_1$, an open interval. Also, for s and t related this way, $\mathbf{f}(t) = \mathbf{g}(s)$ and so in particular, for $s \in I$,

$$g_i(s) = f_i(t).$$

Consequently,

$$s = h_i(g_i(s)) = h_i(f_i(t)) = \phi(t)$$

and so, for $t \in J_1$,

$$\phi'(t) = h'_i(f_i(t)) f'_i(t) = h'_i(g_i(s)) f'_i(t) = \frac{f'_i(t)}{g'_i(\phi(t))} \quad (18.11)$$

which shows that ϕ' exists and is continuous on J_1 , an open interval containing $\phi^{-1}(s_0)$. Since s_0 is arbitrary, this shows ϕ' exists on $[a + \delta, b - \delta]$ and is continuous there.

Now $\mathbf{f}(t) = \mathbf{g} \circ (\mathbf{g}^{-1} \circ \mathbf{f})(t) = \mathbf{g}(\phi(t))$ and it was just shown that ϕ' is a continuous function on $[a - \delta, b + \delta]$. It follows from the chain rule

$$\mathbf{f}'(t) = \mathbf{g}'(\phi(t)) \phi'(t)$$

and so, by Theorem 18.6.12,

$$\begin{aligned} \int_{\phi(a+\delta)}^{\phi(b-\delta)} |\mathbf{g}'(s)| ds &= \int_{a+\delta}^{b-\delta} |\mathbf{g}'(\phi(t))| |\phi'(t)| dt \\ &= \int_{a+\delta}^{b-\delta} |\mathbf{f}'(t)| dt. \end{aligned}$$

Now using the continuity of ϕ , $\mathbf{g}'$, and $\mathbf{f}'$ on $[a, b]$ and letting $\delta \rightarrow 0+$ in the above, yields

$$\int_c^d |\mathbf{g}'(s)| ds = \int_a^b |\mathbf{f}'(t)| dt$$

and this proves the theorem.

Motion On A Space Curve

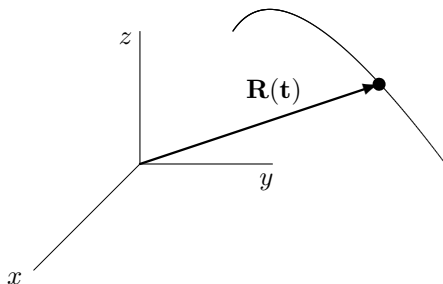
19.0.3 Outcomes

1. Recall the definitions of unit tangent, unit normal, and osculating plane.
2. Calculate the curvature for a space curve.
3. Given the position vector function of a moving object, calculate the velocity, speed, and acceleration of the object and write the acceleration in terms of its tangential and normal components.
4. Derive formulas for the curvature of a parameterized curve and the curvature of a plane curve given as a function.

19.1 Space Curves

A fly buzzing around the room, a person riding a roller coaster, and a satellite orbiting the earth all have something in common. They are moving over some sort of curve in three dimensions.

Denote by $\mathbf{R}(t)$ the position vector of the point on the curve which occurs at time t . Assume that $\mathbf{R}'$, $\mathbf{R}''$ exist and is continuous. Thus $\mathbf{R}' = \mathbf{v}$, the velocity and $\mathbf{R}'' = \mathbf{a}$ is the acceleration.



Lemma 19.1.1 Define $\mathbf{T}(t) \equiv \mathbf{R}'(t) / |\mathbf{R}'(t)|$. Then $|\mathbf{T}(t)| = 1$ and if $\mathbf{T}'(t) \neq 0$, then there exists a unit vector, $\mathbf{N}(t)$ perpendicular to $\mathbf{T}(t)$ and a scalar valued function, $\kappa(t)$, with $\mathbf{T}'(t) = \kappa(t) |\mathbf{v}| \mathbf{N}(t)$.

Proof: It follows from the definition that $|\mathbf{T}| = 1$. Therefore, $\mathbf{T} \cdot \mathbf{T} = 1$ and so, upon differentiating both sides,

$$\mathbf{T}' \cdot \mathbf{T} + \mathbf{T} \cdot \mathbf{T}' = 2\mathbf{T}' \cdot \mathbf{T} = 0.$$

Therefore, $\mathbf{T}'$ is perpendicular to $\mathbf{T}$. Let

$$\mathbf{N}(t) \equiv \frac{\mathbf{T}'}{|\mathbf{T}'|}.$$

Then letting $|\mathbf{T}'| \equiv \kappa(t) |\mathbf{v}(t)|$, it follows

$$\mathbf{T}'(t) = \kappa(t) |\mathbf{v}(t)| \mathbf{N}(t).$$

This proves the lemma.

Definition 19.1.2 The vector, $\mathbf{T}(t)$ is called the **unit tangent vector** and the vector, $\mathbf{N}(t)$ is called the **principal normal**. The function, $\kappa(t)$ in the above lemma is called the **curvature**. The **radius of curvature** is defined as $\rho = 1/\kappa$. The plane determined by the two vectors, $\mathbf{T}$ and $\mathbf{N}$ is called the **osculating**¹ **plane**. It identifies a particular plane which is in a sense tangent to this space curve.

The important thing about this is that it is possible to write the acceleration as the sum of two vectors, one perpendicular to the direction of motion and the other in the direction of motion.

Theorem 19.1.3 For $\mathbf{R}(t)$ the position vector of a space curve, the acceleration is given by the formula

$$\begin{aligned} \mathbf{a} &= \frac{d|\mathbf{v}|}{dt} \mathbf{T} + \kappa |\mathbf{v}|^2 \mathbf{N} \\ &\equiv a_T \mathbf{T} + a_N \mathbf{N}. \end{aligned} \quad (19.1)$$

Furthermore, $a_T^2 + a_N^2 = |\mathbf{a}|^2$.

Proof:

$$\begin{aligned} \mathbf{a} &= \frac{d\mathbf{v}}{dt} = \frac{d}{dt} (\mathbf{R}') = \frac{d}{dt} (|\mathbf{v}| \mathbf{T}) \\ &= \frac{d|\mathbf{v}|}{dt} \mathbf{T} + |\mathbf{v}| \mathbf{T}' \\ &= \frac{d|\mathbf{v}|}{dt} \mathbf{T} + |\mathbf{v}|^2 \kappa \mathbf{N}. \end{aligned}$$

This proves the first part.

For the second part,

$$\begin{aligned} |\mathbf{a}|^2 &= (a_T \mathbf{T} + a_N \mathbf{N}) \cdot (a_T \mathbf{T} + a_N \mathbf{N}) \\ &= a_T^2 \mathbf{T} \cdot \mathbf{T} + 2a_N a_T \mathbf{T} \cdot \mathbf{N} + a_N^2 \mathbf{N} \cdot \mathbf{N} \\ &= a_T^2 + a_N^2 \end{aligned}$$

because $\mathbf{T} \cdot \mathbf{N} = 0$. This proves the theorem.

Finally, it is well to point out that the curvature is a property of the curve itself, and does not depend on the parameterization of the curve. If the curve is given by two different vector valued functions, $\mathbf{R}(t)$ and $\mathbf{R}(\tau)$, then from the formula above for the curvature,

$$\kappa(t) = \frac{|\mathbf{T}'(t)|}{|\mathbf{v}(t)|} = \frac{\left| \frac{d\mathbf{T}}{d\tau} \frac{d\tau}{dt} \right|}{\left| \frac{d\mathbf{R}}{d\tau} \frac{d\tau}{dt} \right|} = \frac{\left| \frac{d\mathbf{T}}{d\tau} \right|}{\left| \frac{d\mathbf{R}}{d\tau} \right|} \equiv \kappa(\tau).$$

From this, it is possible to give an important formula from physics. Suppose an object orbits a point at constant speed, v . In the above notation, $|\mathbf{v}| = v$. What is the centripetal acceleration of this object? You may know from a physics class that the answer is v^2/r where r is the radius. This follows from the above quite easily. The parametrization of the object which is as described is

¹To osculate means to kiss. Thus this plane could be called the kissing plane. However, that does not sound formal enough so we call it the osculating plane.

$$\mathbf{R}(t) = \left(r \cos\left(\frac{v}{r}t\right), r \sin\left(\frac{v}{r}t\right) \right).$$

Therefore, $\mathbf{T} = \left(-\sin\left(\frac{v}{r}t\right), \cos\left(\frac{v}{r}t\right) \right)$ and $\mathbf{T}' = \left(-\frac{v}{r} \cos\left(\frac{v}{r}t\right), -\frac{v}{r} \sin\left(\frac{v}{r}t\right) \right)$. Thus,

$$\kappa = |\mathbf{T}'(t)|/v = \frac{1}{r}.$$

It follows

$$\mathbf{a} = \frac{dv}{dt} \mathbf{T} + v^2 \kappa \mathbf{N} = \frac{v^2}{r} \mathbf{N}.$$

The vector, $\mathbf{N}$ points from the object toward the center of the circle because it is a positive multiple of the vector, $\left(-\frac{v}{r} \cos\left(\frac{v}{r}t\right), -\frac{v}{r} \sin\left(\frac{v}{r}t\right) \right)$.

Formula 19.1 also yields an easy way to find the curvature. Take the cross product of both sides with $\mathbf{v}$, the velocity. Then

$$\begin{aligned} \mathbf{a} \times \mathbf{v} &= \frac{d|\mathbf{v}|}{dt} \mathbf{T} \times \mathbf{v} + |\mathbf{v}|^2 \kappa \mathbf{N} \times \mathbf{v} \\ &= \frac{d|\mathbf{v}|}{dt} \mathbf{T} \times \mathbf{v} + |\mathbf{v}|^3 \kappa \mathbf{N} \times \mathbf{T} \end{aligned}$$

Now $\mathbf{T}$ and $\mathbf{v}$ have the same direction so the first term on the right equals zero. Taking the magnitude of both sides, and using the fact that $\mathbf{N}$ and $\mathbf{T}$ are two perpendicular unit vectors,

$$|\mathbf{a} \times \mathbf{v}| = |\mathbf{v}|^3 \kappa$$

and so

$$\kappa = \frac{|\mathbf{a} \times \mathbf{v}|}{|\mathbf{v}|^3}. \quad (19.2)$$

Example 19.1.4 Let $\mathbf{R}(t) = (\cos(t), t, t^2)$ for $t \in [0, 3]$. Find the speed, velocity, curvature, and write the acceleration in terms of normal and tangential components.

First of all $\mathbf{v}(t) = (-\sin t, 1, 2t)$ and so the speed is given by

$$|\mathbf{v}| = \sqrt{\sin^2(t) + 1 + 4t^2}.$$

Therefore,

$$a_T = \frac{d}{dt} \left(\sqrt{\sin^2(t) + 1 + 4t^2} \right) = \frac{\sin(t) \cos(t) + 4t}{\sqrt{(2 + 4t^2 - \cos^2 t)}}.$$

It remains to find a_N . To do this, you can find the curvature first if you like.

$$\mathbf{a}(t) = \mathbf{R}''(t) = (-\cos t, 0, 2).$$

Then

$$\begin{aligned} \kappa &= \frac{|(-\cos t, 0, 2) \times (-\sin t, 1, 2t)|}{\left(\sqrt{\sin^2(t) + 1 + 4t^2} \right)^3} \\ &= \frac{\sqrt{4 + (-2 \sin(t) + 2(\cos(t))t)^2 + \cos^2(t)}}{\left(\sqrt{\sin^2(t) + 1 + 4t^2} \right)^3} \end{aligned}$$

Then

$$a_N = \kappa |\mathbf{v}|^2$$

$$\begin{aligned}
&= \frac{\sqrt{4 + (-2 \sin(t) + 2(\cos(t))t)^2 + \cos^2(t)}}{\left(\sqrt{\sin^2(t) + 1 + 4t^2}\right)^3} (\sin^2(t) + 1 + 4t^2) \\
&= \frac{\sqrt{4 + (-2 \sin(t) + 2(\cos(t))t)^2 + \cos^2(t)}}{\sqrt{\sin^2(t) + 1 + 4t^2}}.
\end{aligned}$$

You can observe the formula $a_N^2 + a_T^2 = |\mathbf{a}|^2$ holds. Indeed $a_N^2 + a_T^2 =$

$$\begin{aligned}
&\left(\frac{\sqrt{4 + (-2 \sin(t) + 2(\cos(t))t)^2 + \cos^2(t)}}{\sqrt{\sin^2(t) + 1 + 4t^2}}\right)^2 + \left(\frac{\sin(t) \cos(t) + 4t}{\sqrt{(2 + 4t^2 - \cos^2(t))}}\right)^2 \\
&= \frac{4 + (-2 \sin t + 2(\cos t)t)^2 + \cos^2 t}{\sin^2 t + 1 + 4t^2} + \frac{(\sin t \cos t + 4t)^2}{2 + 4t^2 - \cos^2 t} = \cos^2 t + 4 = |\mathbf{a}|^2
\end{aligned}$$

19.1.1 Some Simple Techniques

Recall the formula for acceleration is

$$\mathbf{a} = a_T \mathbf{T} + a_N \mathbf{N} \quad (19.3)$$

where $a_T = \frac{d|\mathbf{v}|}{dt}$ and $a_N = \kappa |\mathbf{v}|^2$. Of course one way to find a_T and a_N is to just find $|\mathbf{v}|$, $\frac{d|\mathbf{v}|}{dt}$ and κ and plug in. However, there is another way which might be easier. Take the dot product of both sides with $\mathbf{T}$. This gives,

$$\mathbf{a} \cdot \mathbf{T} = a_T \mathbf{T} \cdot \mathbf{T} + a_N \mathbf{N} \cdot \mathbf{T} = a_T.$$

Thus

$$\mathbf{a} = (\mathbf{a} \cdot \mathbf{T}) \mathbf{T} + a_N \mathbf{N}$$

and so

$$\mathbf{a} - (\mathbf{a} \cdot \mathbf{T}) \mathbf{T} = a_N \mathbf{N} \quad (19.4)$$

and taking norms of both sides,

$$|\mathbf{a} - (\mathbf{a} \cdot \mathbf{T}) \mathbf{T}| = a_N.$$

Also from 19.4,

$$\frac{\mathbf{a} - (\mathbf{a} \cdot \mathbf{T}) \mathbf{T}}{|\mathbf{a} - (\mathbf{a} \cdot \mathbf{T}) \mathbf{T}|} = \frac{a_N \mathbf{N}}{a_N |\mathbf{N}|} = \mathbf{N}.$$

Also recall

$$\kappa = \frac{|\mathbf{a} \times \mathbf{v}|}{|\mathbf{v}|^3}, \quad a_T^2 + a_N^2 = |\mathbf{a}|^2$$

This is usually easier than computing $\mathbf{T}'/|\mathbf{T}'|$. To illustrate the use of these simple observations, consider the example worked above which was fairly messy. I will make it easier by selecting a value of t and by using the above simplifying techniques.

Example 19.1.5 Let $\mathbf{R}(t) = (\cos(t), t, t^2)$ for $t \in [0, 3]$. Find the speed, velocity, curvature, and write the acceleration in terms of normal and tangential components when $t = 0$. Also find $\mathbf{N}$ at the point where $t = 0$.

First I need to find the velocity and acceleration. Thus

$$\mathbf{v} = (-\sin t, 1, 2t), \quad \mathbf{a} = (-\cos t, 0, 2)$$

and consequently,

$$\mathbf{T} = \frac{(-\sin t, 1, 2t)}{\sqrt{\sin^2(t) + 1 + 4t^2}}.$$

When $t = 0$, this reduces to

$$\mathbf{v}(0) = (0, 1, 0), \quad \mathbf{a} = (-1, 0, 2), \quad |\mathbf{v}(0)| = 1, \quad \mathbf{T} = (0, 1, 0),$$

and consequently,

$$\mathbf{T} = (0, 1, 0).$$

Then the tangential component of acceleration when $t = 0$ is

$$a_T = (-1, 0, 2) \cdot (0, 1, 0) = 0$$

Now $|\mathbf{a}|^2 = 5$ and so $a_N = \sqrt{5}$ because $a_T^2 + a_N^2 = |\mathbf{a}|^2$. Thus $\sqrt{5} = \kappa |\mathbf{v}(0)|^2 = \kappa \cdot 1 = \kappa$. Next let's find $\mathbf{N}$. From $\mathbf{a} = a_T \mathbf{T} + a_N \mathbf{N}$ it follows

$$(-1, 0, 2) = 0 \cdot \mathbf{T} + \sqrt{5} \mathbf{N}$$

and so

$$\mathbf{N} = \frac{1}{\sqrt{5}}(-1, 0, 2).$$

This was pretty easy.

Example 19.1.6 Find a formula for the curvature of the curve given by the graph of $y = f(x)$ for $x \in [a, b]$. Assume whatever you like about smoothness of f .

You need to write this as a parametric curve. This is most easily accomplished by letting $t = x$. Thus a parametrization is

$$(t, f(t), 0) : t \in [a, b].$$

Then you can use the formula given above. The acceleration is $(0, f''(t), 0)$ and the velocity is $(1, f'(t), 0)$. Therefore,

$$\mathbf{a} \times \mathbf{v} = (0, f''(t), 0) \times (1, f'(t), 0) = (0, 0, -f''(t)).$$

Therefore, the curvature is given by

$$\frac{|\mathbf{a} \times \mathbf{v}|}{|\mathbf{v}|^3} = \frac{|f''(t)|}{(1 + f'(t)^2)^{3/2}}.$$

Sometimes curves don't come to you parametrically. This is unfortunate when it occurs but you can sometimes find a parametric description of such curves. It should be emphasized that it is only sometimes when you can actually find a parametrization. General systems of nonlinear equations cannot be solved using algebra.

Example 19.1.7 Find a parametrization for the intersection of the surfaces

$$y + 3z = 2x^2 + 4 \quad \text{and} \quad y + 2z = x + 1.$$

You need to solve for x and y in terms of x . This yields

$$z = 2x^2 - x + 3, \quad y = -4x^2 + 3x - 5.$$

Therefore, letting $t = x$, the parametrization is

$$(x, y, z) = (t, -4t^2 - 5 + 3t, -t + 3 + 2t^2).$$

Example 19.1.8 Find a parametrization for the straight line joining $(3, 2, 4)$ and $(1, 10, 5)$.

$(x, y, z) = (3, 2, 4) + t(-2, 8, 1) = (3 - 2t, 2 + 8t, 4 + t)$ where $t \in [0, 1]$. Note where this came from. The vector, $(-2, 8, 1)$ is obtained from $(1, 10, 5) - (3, 2, 4)$. Now you should check to see this works.

19.2 Geometry Of Space Curves*

If you are interested in more on space curves, you should read this section. Otherwise, proceed to the exercises. Denote by $\mathbf{R}(s)$ the function which takes s to a point on this curve where s is arc length. Thus $\mathbf{R}(s)$ equals the point on the curve which occurs when you have traveled a distance of s along the curve from one end. This is known as the parametrization of the curve in terms of arc length. Note also that it incorporates an orientation on the curve because there are exactly two ends you could begin measuring length from. In this section, assume anything about smoothness and continuity to make the following manipulations valid. In particular, assume that $\mathbf{R}'$ exists and is continuous.

Lemma 19.2.1 Define $\mathbf{T}(s) \equiv \mathbf{R}'(s)$. Then $|\mathbf{T}(s)| = 1$ and if $\mathbf{T}'(s) \neq 0$, then there exists a unit vector, $\mathbf{N}(s)$ perpendicular to $\mathbf{T}(s)$ and a scalar valued function, $\kappa(s)$ with $\mathbf{T}'(s) = \kappa(s)\mathbf{N}(s)$.

Proof: First, $s = \int_0^s |\mathbf{R}'(r)| dr$ because of the definition of arc length. Therefore, from the fundamental theorem of calculus, $1 = |\mathbf{R}'(s)| = |\mathbf{T}(s)|$. Therefore, $\mathbf{T} \cdot \mathbf{T} = 1$ and so upon differentiating this on both sides, yields $\mathbf{T}' \cdot \mathbf{T} + \mathbf{T} \cdot \mathbf{T}' = 0$ which shows $\mathbf{T} \cdot \mathbf{T}' = 0$. Therefore, the vector, $\mathbf{T}'$ is perpendicular to the vector, $\mathbf{T}$. In case $\mathbf{T}'(s) \neq 0$, let $\mathbf{N}(s) = \frac{\mathbf{T}'(s)}{|\mathbf{T}'(s)|}$ and so $\mathbf{T}'(s) = |\mathbf{T}'(s)|\mathbf{N}(s)$, showing the scalar valued function is $\kappa(s) = |\mathbf{T}'(s)|$. This proves the lemma.

The radius of curvature is defined as $\rho = \frac{1}{\kappa}$. Thus at points where there is a lot of curvature, the radius of curvature is small and at points where the curvature is small, the radius of curvature is large. The plane determined by the two vectors, $\mathbf{T}$ and $\mathbf{N}$ is called the osculating plane. It identifies a particular plane which is in a sense tangent to this space curve. In the case where $|\mathbf{T}'(s)| = 0$ near the point of interest, $\mathbf{T}(s)$ equals a constant and so the space curve is a straight line which it would be supposed has no curvature. Also, the principal normal is undefined in this case. This makes sense because if there is no curving going on, there is no special direction normal to the curve at such points which could be distinguished from any other direction normal to the curve. In the case where $|\mathbf{T}'(s)| = 0$, $\kappa(s) = 0$ and the radius of curvature would be considered infinite.

Definition 19.2.2 The vector, $\mathbf{T}(s)$ is called the unit tangent vector and the vector, $\mathbf{N}(s)$ is called the **principal normal**. The function, $\kappa(s)$ in the above lemma is called the **curvature**. When $\mathbf{T}'(s) \neq 0$ so the principal normal is defined, the vector, $\mathbf{B}(s) \equiv \mathbf{T}(s) \times \mathbf{N}(s)$ is called the **binormal**.

The binormal is normal to the osculating plane and $\mathbf{B}'$ tells how fast this vector changes. Thus it measures the rate at which the curve twists.

Lemma 19.2.3 *Let $\mathbf{R}(s)$ be a parametrization of a space curve with respect to arc length and let the vectors, $\mathbf{T}$, $\mathbf{N}$, and $\mathbf{B}$ be as defined above. Then $\mathbf{B}' = \mathbf{T} \times \mathbf{N}'$ and there exists a scalar function, $\tau(s)$ such that $\mathbf{B}' = \tau\mathbf{N}$.*

Proof: From the definition of $\mathbf{B} = \mathbf{T} \times \mathbf{N}$, and you can differentiate both sides and get $\mathbf{B}' = \mathbf{T}' \times \mathbf{N} + \mathbf{T} \times \mathbf{N}'$. Now recall that $\mathbf{T}'$ is a multiple called curvature multiplied by $\mathbf{N}$ so the vectors, $\mathbf{T}'$ and $\mathbf{N}$ have the same direction and $\mathbf{B}' = \mathbf{T} \times \mathbf{N}'$. Therefore, $\mathbf{B}'$ is either zero or is perpendicular to $\mathbf{T}$. But also, from the definition of $\mathbf{B}$, $\mathbf{B}$ is a unit vector and so $\mathbf{B}(s) \cdot \mathbf{B}(s) = 1$. Differentiating this, $\mathbf{B}'(s) \cdot \mathbf{B}(s) + \mathbf{B}(s) \cdot \mathbf{B}'(s) = 0$ showing that $\mathbf{B}'$ is perpendicular to $\mathbf{B}$ also. Therefore, $\mathbf{B}'$ is a vector which is perpendicular to both vectors, $\mathbf{T}$ and $\mathbf{B}$ and since this is in three dimensions, $\mathbf{B}'$ must be some scalar multiple of $\mathbf{N}$ and it is this multiple called τ . Thus $\mathbf{B}' = \tau\mathbf{N}$ as claimed.

Lets go over this last claim a little more. The following situation is obtained. There are two vectors, $\mathbf{T}$ and $\mathbf{B}$ which are perpendicular to each other and both $\mathbf{B}'$ and $\mathbf{N}$ are perpendicular to these two vectors, hence perpendicular to the plane determined by them. Therefore, $\mathbf{B}'$ must be a multiple of $\mathbf{N}$. Take a piece of paper, draw two unit vectors on it which are perpendicular. Then you can see that any two vectors which are perpendicular to this plane must be multiples of each other.

The scalar function, τ is called the torsion. In case $\mathbf{T}' = 0$, none of this is defined because in this case there is not a well defined osculating plane. The conclusion of the following theorem is called the Serret Frenet formulas.

Theorem 19.2.4 (Serret Frenet) *Let $\mathbf{R}(s)$ be the parametrization with respect to arc length of a space curve and $\mathbf{T}(s) = \mathbf{R}'(s)$ is the unit tangent vector. Suppose $|\mathbf{T}'(s)| \neq 0$ so the principal normal, $\mathbf{N}(s) = \frac{\mathbf{T}'(s)}{|\mathbf{T}'(s)|}$ is defined. The binormal is the vector $\mathbf{B} \equiv \mathbf{T} \times \mathbf{N}$ so $\mathbf{T}, \mathbf{N}, \mathbf{B}$ forms a right handed system of unit vectors each of which is perpendicular to every other. Then the following system of differential equations holds in $\mathbb{R}^3$.*

$$\mathbf{B}' = \tau\mathbf{N}, \quad \mathbf{T}' = \kappa\mathbf{N}, \quad \mathbf{N}' = -\kappa\mathbf{T} - \tau\mathbf{B}$$

where κ is the curvature and is nonnegative and τ is the **torsion**.

Proof: $\kappa \geq 0$ because $\kappa = |\mathbf{T}'(s)|$. The first two equations are already established. To get the third, note that $\mathbf{B} \times \mathbf{T} = \mathbf{N}$ which follows because $\mathbf{T}, \mathbf{N}, \mathbf{B}$ is given to form a right handed system of unit vectors each perpendicular to the others. (Use your right hand.) Now take the derivative of this expression. thus

$$\begin{aligned} \mathbf{N}' &= \mathbf{B}' \times \mathbf{T} + \mathbf{B} \times \mathbf{T}' \\ &= \tau\mathbf{N} \times \mathbf{T} + \kappa\mathbf{B} \times \mathbf{N}. \end{aligned}$$

Now recall again that $\mathbf{T}, \mathbf{N}, \mathbf{B}$ is a right hand system. Thus

$$\mathbf{N} \times \mathbf{T} = -\mathbf{B}, \quad \mathbf{B} \times \mathbf{N} = -\mathbf{T}.$$

This establishes the Frenet Serret formulas.

This is an important example of a system of differential equations in $\mathbb{R}^3$. It is a remarkable result because it says that from knowledge of the two scalar functions, τ and κ , and initial values for $\mathbf{B}, \mathbf{T}$, and $\mathbf{N}$ when $s = 0$ you can obtain the binormal, unit tangent, and principal normal vectors. It is just the solution of an initial value problem although this is for a vector valued rather than scalar valued function. Having done this, you can reconstruct the entire space curve starting at some point, $\mathbf{R}_0$ because $\mathbf{R}'(s) = \mathbf{T}(s)$ and so $\mathbf{R}(s) = \mathbf{R}_0 + \int_0^s \mathbf{T}'(r) dr$.

The vectors, $\mathbf{B}, \mathbf{T}$, and $\mathbf{N}$ are vectors which are functions of position on the space curve. Often, especially in applications, you deal with a space curve which is parameterized by a function of t where t is time. Thus a value of t would correspond to a point

on this curve and you could let $\mathbf{B}(t)$, $\mathbf{T}(t)$, and $\mathbf{N}(t)$ be the binormal, unit tangent, and principal normal at this point of the curve. The following example is typical.

Example 19.2.5 Given the circular helix, $\mathbf{R}(t) = (a \cos t)\mathbf{i} + (a \sin t)\mathbf{j} + (bt)\mathbf{k}$, find the arc length, $s(t)$, the unit tangent vector, $\mathbf{T}(t)$, the principal normal, $\mathbf{N}(t)$, the binormal, $\mathbf{B}(t)$, the curvature, $\kappa(t)$, and the torsion, $\tau(t)$. Here $t \in [0, T]$.

The arc length is $s(t) = \int_0^t (\sqrt{a^2 + b^2}) \, dr = (\sqrt{a^2 + b^2})t$. Now the tangent vector is obtained using the chain rule as

$$\begin{aligned}\mathbf{T} &= \frac{d\mathbf{R}}{ds} = \frac{d\mathbf{R}}{dt} \frac{dt}{ds} = \frac{1}{\sqrt{a^2 + b^2}} \mathbf{R}'(t) \\ &= \frac{1}{\sqrt{a^2 + b^2}} ((-a \sin t)\mathbf{i} + (a \cos t)\mathbf{j} + b\mathbf{k})\end{aligned}$$

The principal normal:

$$\begin{aligned}\frac{d\mathbf{T}}{ds} &= \frac{d\mathbf{T}}{dt} \frac{dt}{ds} \\ &= \frac{1}{a^2 + b^2} ((-a \cos t)\mathbf{i} + (-a \sin t)\mathbf{j} + 0\mathbf{k})\end{aligned}$$

and so

$$\mathbf{N} = \frac{d\mathbf{T}}{ds} / \left| \frac{d\mathbf{T}}{ds} \right| = -((\cos t)\mathbf{i} + (\sin t)\mathbf{j})$$

The binormal:

$$\begin{aligned}\mathbf{B} &= \frac{1}{\sqrt{a^2 + b^2}} \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -a \sin t & a \cos t & b \\ -\cos t & -\sin t & 0 \end{vmatrix} \\ &= \frac{1}{\sqrt{a^2 + b^2}} ((b \sin t)\mathbf{i} - b \cos t \mathbf{j} + a\mathbf{k})\end{aligned}$$

Now the curvature, $\kappa(t) = \left| \frac{d\mathbf{T}}{ds} \right| = \sqrt{\left(\frac{a \cos t}{a^2 + b^2} \right)^2 + \left(\frac{a \sin t}{a^2 + b^2} \right)^2} = \frac{a}{a^2 + b^2}$. Note the curvature is constant in this example. The final task is to find the torsion. Recall that $\mathbf{B}' = \tau \mathbf{N}$ where the derivative on $\mathbf{B}$ is taken with respect to arc length. Therefore, remembering that t is a function of s ,

$$\begin{aligned}\mathbf{B}'(s) &= \frac{1}{\sqrt{a^2 + b^2}} ((b \cos t)\mathbf{i} + (b \sin t)\mathbf{j}) \frac{dt}{ds} \\ &= \frac{1}{a^2 + b^2} ((b \cos t)\mathbf{i} + (b \sin t)\mathbf{j}) \\ &= \tau (-(\cos t)\mathbf{i} - (\sin t)\mathbf{j}) = \tau \mathbf{N}\end{aligned}$$

and it follows $-b/(a^2 + b^2) = \tau$.

An important application of the usefulness of these ideas involves the decomposition of the acceleration in terms of these vectors of an object moving over a space curve.

Corollary 19.2.6 Let $\mathbf{R}(t)$ be a space curve and denote by $\mathbf{v}(t)$ the velocity, $\mathbf{v}(t) = \mathbf{R}'(t)$ and let $v(t) \equiv |\mathbf{v}(t)|$ denote the speed and let $\mathbf{a}(t)$ denote the acceleration. Then $\mathbf{v} = v\mathbf{T}$ and $\mathbf{a} = \frac{dv}{dt}\mathbf{T} + \kappa v^2 \mathbf{N}$.

Proof: $\mathbf{T} = \frac{d\mathbf{R}}{ds} = \frac{d\mathbf{R}}{dt} \frac{dt}{ds} = \mathbf{v} \frac{dt}{ds}$. Also, $s = \int_0^t v(r) dr$ and so $\frac{ds}{dt} = v$ which implies $\frac{dt}{ds} = \frac{1}{v}$. Therefore, $\mathbf{T} = \mathbf{v}/v$ which implies $\mathbf{v} = v\mathbf{T}$ as claimed.

Now the acceleration is just the derivative of the velocity and so by the Serrat Frenet formulas,

$$\begin{aligned} \mathbf{a} &= \frac{dv}{dt} \mathbf{T} + v \frac{d\mathbf{T}}{dt} \\ &= \frac{dv}{dt} \mathbf{T} + v \frac{d\mathbf{T}}{ds} v = \frac{dv}{dt} \mathbf{T} + v^2 \kappa \mathbf{N} \end{aligned}$$

Note how this decomposes the acceleration into a component tangent to the curve and one which is normal to it. Also note that from the above, $v |\mathbf{T}'| \frac{\mathbf{T}'(t)}{|\mathbf{T}'|} = v^2 \kappa \mathbf{N}$ and so $\frac{|\mathbf{T}'|}{v} = \kappa$ and $\mathbf{N} = \frac{\mathbf{T}'(t)}{|\mathbf{T}'|}$

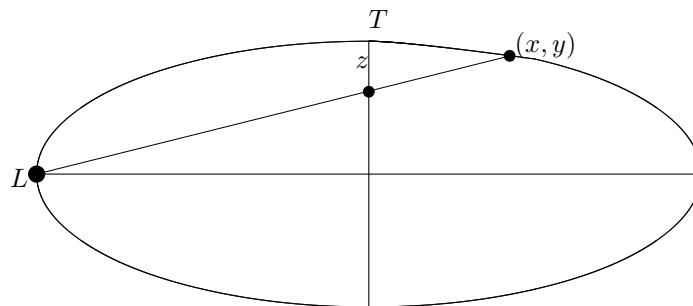
19.3 Exercises

- Find a parametrization for the intersection of the planes $2x + y + 3z = -2$ and $3x - 2y + z = -4$.
- Find a parametrization for the intersection of the plane $3x + y + z = -3$ and the circular cylinder $x^2 + y^2 = 1$.
- Find a parametrization for the intersection of the plane $4x + 2y + 3z = 2$ and the elliptic cylinder $x^2 + 4z^2 = 9$.
- Find a parametrization for the straight line joining $(1, 2, 1)$ and $(-1, 4, 4)$.
- Find a parametrization for the intersection of the surfaces $3y + 3z = 3x^2 + 2$ and $3y + 2z = 3$.
- Find a formula for the curvature of the curve, $y = \sin x$ in the xy plane.
- An object moves over the curve (t, e^t, at) where $t \in \mathbb{R}$ and a is a positive constant. Find the value of t at which the normal component of acceleration is largest if there is such a point.
- Find a formula for the curvature of the space curve in $\mathbb{R}^2$, $(x(t), y(t))$.
- An object moves over the helix, $(\cos 3t, \sin 3t, 5t)$. Find the normal and tangential components of the acceleration of this object as a function of t and write the acceleration in the form $a_T \mathbf{T} + a_N \mathbf{N}$.
- An object moves over the helix, $(\cos t, \sin t, t)$. Find the normal and tangential components of the acceleration of this object as a function of t and write the acceleration in the form $a_T \mathbf{T} + a_N \mathbf{N}$.
- An object moves in $\mathbb{R}^3$ according to the formula, $(\cos 3t, \sin 3t, t^2)$. Find the normal and tangential components of the acceleration of this object as a function of t and write the acceleration in the form $a_T \mathbf{T} + a_N \mathbf{N}$.
- An object moves over the helix, $(\cos t, \sin t, 2t)$. Find the osculating plane at the point of the curve corresponding to $t = \pi/4$.
- An object moves over a circle of radius r according to the formula,

$$\mathbf{r}(t) = (r \cos(\omega t), r \sin(\omega t))$$

where $v = r\omega$. Show that the speed of the object is constant and equals to v . Tell why $a_T = 0$ and find a_N , $\mathbf{N}$.

14. Suppose $|\mathbf{R}(t)| = c$ where c is a constant $\mathbf{R}(t)$. Show the velocity, $\mathbf{R}'(t)$ is always perpendicular to $\mathbf{R}(t)$.
15. An object moves in three dimensions and the only force on the object is a central force. This means that if $\mathbf{r}(t)$ is the position of the object, $\mathbf{a}(t) = k(\mathbf{r}(t))\mathbf{r}(t)$ where k is some function. Show that if this happens, then the motion of the object must be in a plane. **Hint:** First argue that $\mathbf{a} \times \mathbf{r} = \mathbf{0}$. Next show that $(\mathbf{a} \times \mathbf{r}) = (\mathbf{v} \times \mathbf{r})'$. Therefore, $(\mathbf{v} \times \mathbf{r})' = \mathbf{0}$. Explain why this requires $\mathbf{v} \times \mathbf{r} = \mathbf{c}$ for some vector, $\mathbf{c}$ which does not depend on t . Then explain why $\mathbf{c} \cdot \mathbf{r} = 0$. This implies the motion is in a plane. Why? What are some examples of central forces?
16. Let $\mathbf{R}(t) = (\cos t)\mathbf{i} + (\cos t)\mathbf{j} + (\sqrt{2}\sin t)\mathbf{k}$. Find the arc length, s as a function of the parameter, t , if $t = 0$ is taken to correspond to $s = 0$.
17. Let $\mathbf{R}(t) = 2\mathbf{i} + (4t + 2)\mathbf{j} + 4t\mathbf{k}$. Find the arc length, s as a function of the parameter, t , if $t = 0$ is taken to correspond to $s = 0$.
18. Let $\mathbf{R}(t) = e^{5t}\mathbf{i} + e^{-5t}\mathbf{j} + 5\sqrt{2}t\mathbf{k}$. Find the arc length, s as a function of the parameter, t , if $t = 0$ is taken to correspond to $s = 0$.
19. Consider the curve obtained from the graph of $y = f(x)$. Find a formula for the curvature.
20. Consider the curve in the plane $y = e^x$. Find the point on this curve at which the curvature is a maximum.
21. An object moves along the x axis toward $(0, 0)$ and then along the curve $y = x^2$ in the direction of increasing x at constant speed. Is the force acting on the object a continuous function? Explain. Is there any physically reasonable way to make this force continuous by relaxing the requirement that the object move at constant speed? If the curve were part of a railroad track, what would happen at the point where $x = 0$?
22. An object of mass m moving over a space curve is acted on by a force, $\mathbf{F}$. Show the work done by this force equals ma_T (length of the curve). In other words, it is only the tangential component of the force which does work.
23. The edge of an elliptical skating rink represented in the following picture has a light at its left end and satisfies the equation $\frac{x^2}{900} + \frac{y^2}{256} = 1$. (Distances measured in yards.)



A hockey puck slides from the point, T towards the center of the rink at the rate of 2 yards per second. What is the speed of its shadow along the wall when $z = 8$? **Hint:** You need to find $\sqrt{x'^2 + y'^2}$ at the instant described.

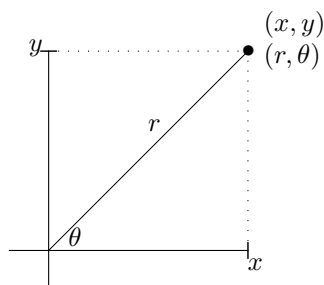
Some Curvilinear Coordinate Systems

20.0.1 Outcomes

1. Recall and use polar coordinates.
2. Graph relations involving polar coordinates.
3. Find the area of regions defined in terms of polar coordinates.
4. Recall and understand the derivation of Kepler's laws.
5. Recall and apply the concept of acceleration in polar coordinates.
6. Recall and use cylindrical and spherical coordinates.

20.1 Polar Coordinates

So far points have been identified in terms of Cartesian coordinates but there are other ways of specifying points in two and three dimensional space. These other ways involve using a list of two or three numbers which have a totally different meaning than Cartesian coordinates to specify a point in two or three dimensional space. In general these lists of numbers which have a different meaning than Cartesian coordinates are called Curvilinear coordinates. Probably the simplest curvilinear coordinate system is that of **polar coordinates**. The idea is suggested in the following picture.



You see in this picture, the number r identifies the distance of the point from the origin, $(0, 0)$ while θ is the angle shown between the positive x axis and the line from the origin to the point. This angle will always be given in radians and is in the interval $[0, 2\pi)$. Thus the given point, indicated by a small dot in the picture, can be described

in terms of the Cartesian coordinates, (x, y) or the polar coordinates, (r, θ) . How are the two coordinates systems related? From the picture,

$$x = r \cos(\theta), \quad y = r \sin(\theta). \quad (20.1)$$

Example 20.1.1 *The polar coordinates of a point in the plane are $(5, \frac{\pi}{6})$. Find the Cartesian or rectangular coordinates of this point.*

From 20.1, $x = 5 \cos(\frac{\pi}{6}) = \frac{5}{2}\sqrt{3}$ and $y = 5 \sin(\frac{\pi}{6}) = \frac{5}{2}$. Thus the Cartesian coordinates are $(\frac{5}{2}\sqrt{3}, \frac{5}{2})$.

Example 20.1.2 *Suppose the Cartesian coordinates of a point are $(3, 4)$. Find the polar coordinates.*

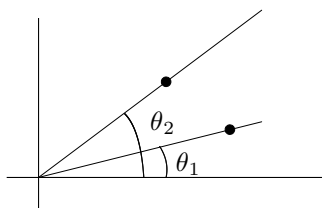
Recall that r is the distance from $(0, 0)$ and so $r = 5 = \sqrt{3^2 + 4^2}$. It remains to identify the angle. Note the point is in the first quadrant, (Both the x and y values are positive.) Therefore, the angle is something between 0 and $\pi/2$ and also $3 = 5 \cos(\theta)$, and $4 = 5 \sin(\theta)$. Therefore, dividing yields $\tan(\theta) = 4/3$. At this point, use a calculator or a table of trigonometric functions to find that at least approximately, $\theta = .927295$ radians.

20.1.1 Graphs In Polar Coordinates

Just as in the case of rectangular coordinates, it is possible to use relations between the polar coordinates to specify points in the plane. The process of sketching their graphs is very similar to that used to sketch graphs of functions in rectangular coordinates. I will only consider the case where the relation between the polar coordinates is of the form, $r = f(\theta)$. To graph such a relation, you can make a table of the form

θ	r
θ_1	$f(\theta_1)$
θ_2	$f(\theta_2)$
$\vdots$	$\vdots$

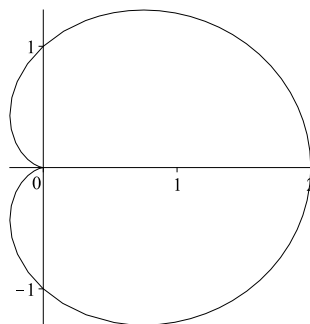
and then graph the resulting points and connect them up with a curve. The following picture illustrates how to begin this process.



To obtain the point in the plane which goes with the pair $(\theta, f(\theta))$, you draw the ray through the origin which makes an angle of θ with the positive x axis. Then you move along this ray a distance of $f(\theta)$ to obtain the point. As in the case with rectangular coordinates, this process is tedious and is best done by a computer algebra system.

Example 20.1.3 *Graph the polar equation, $r = 1 + \cos \theta$.*

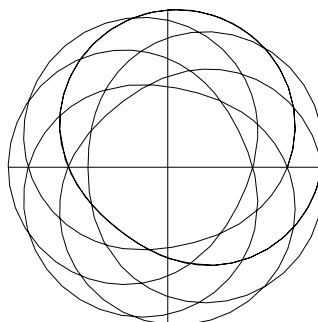
To do this, I will use Maple. The command which produces the polar graph of this is: `> plot(1+cos(t), t=0..2*Pi, coords=polar);` It tells Maple that r is given by $1 + \cos(t)$ and that $t \in [0, 2\pi]$. The variable t is playing the role of θ . It is easier to type t than θ in Maple.



You can also see just from your knowledge of the trig. functions that the graph should look something like this. When $\theta = 0$, $r = 2$ and then as θ increases to $\pi/2$, you see that $\cos \theta$ decreases to 0. Thus the line from the origin to the point on the curve should get shorter as θ goes from 0 to $\pi/2$. Then from $\pi/2$ to π , $\cos \theta$ gets negative eventually equaling -1 at $\theta = \pi$. Thus $r = 0$ at this point. Viewing the graph, you see this is exactly what happens. The above function is called a **cardioid**.

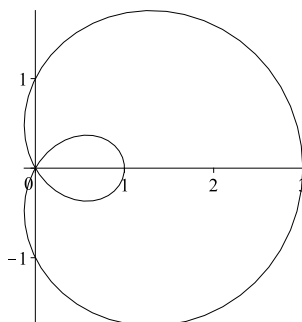
Here is another example. This is the graph obtained from $r = 3 + \sin\left(\frac{7\theta}{6}\right)$.

Example 20.1.4 Graph $r = 3 + \sin\left(\frac{7\theta}{6}\right)$ for $\theta \in [0, 14\pi]$.



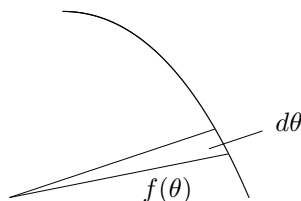
In polar coordinates people sometimes allow r to be negative. When this happens, it means that to obtain the point in the plane, you go in the opposite direction along the ray which starts at the origin and makes an angle of θ with the positive x axis. I do not believe the fussiness occasioned by this extra generality is justified by any sufficiently interesting application so no more will be said about this. It is mainly a fun way to obtain pretty pictures. Here is such an example.

Example 20.1.5 Graph $r = 1 + 2\cos \theta$ for $\theta \in [0, 2\pi]$.



20.2 The Area In Polar Coordinates

How can you find the area of the region determined by $0 \leq r \leq f(\theta)$ for $\theta \in [a, b]$, assuming this is a well defined set of points in the plane? See Example 20.1.5 with $\theta \in [0, 2\pi]$ to see something which it would be better to avoid. I have in mind the situation where every ray through the origin having angle θ for $\theta \in [a, b]$ intersects the graph of $r = f(\theta)$ in exactly one point. To see how to find the area of such a region, consider the following picture.



This is a representation of a small triangle obtained from two rays whose angles differ by only $d\theta$. What is the area of this triangle, dA ? It would be

$$\frac{1}{2} \sin(d\theta) f(\theta)^2 \approx \frac{1}{2} f(\theta)^2 d\theta = dA$$

with the approximation getting better as the angle gets smaller. Thus the area should solve the initial value problem,

$$\frac{dA}{d\theta} = \frac{1}{2} f(\theta)^2, \quad A(a) = 0.$$

Therefore, the total area would be given by the integral,

$$\frac{1}{2} \int_a^b f(\theta)^2 d\theta. \quad (20.2)$$

Example 20.2.1 Find the area of the cardioid, $r = 1 + \cos \theta$ for $\theta \in [0, 2\pi]$.

From the graph of the cardioid presented earlier, you can see the region of interest satisfies the conditions above that every ray intersects the graph in only one point. Therefore, from 20.2 this area is

$$\frac{1}{2} \int_0^{2\pi} (1 + \cos(\theta))^2 d\theta = \frac{3}{2}\pi.$$

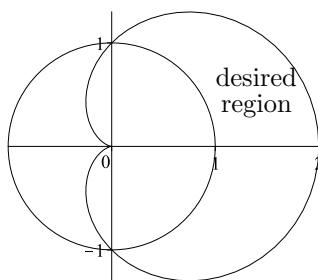
Example 20.2.2 Verify the area of a circle of radius a is πa^2 .

The polar equation is just $r = a$ for $\theta \in [0, 2\pi]$. Therefore, the area should be

$$\frac{1}{2} \int_0^{2\pi} a^2 d\theta = \pi a^2.$$

Example 20.2.3 Find the area of the region inside the cardioid, $r = 1 + \cos \theta$ and outside the circle, $r = 1$ for $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

As is usual in such cases, it is a good idea to graph the curves involved to get an idea what is wanted.



The area of this region would be the area of the part of the cardioid corresponding to $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ minus the area of the part of the circle in the first quadrant. Thus the area is

$$\frac{1}{2} \int_{-\pi/2}^{\pi/2} (1 + \cos(\theta))^2 d\theta - \frac{1}{2} \int_{-\pi/2}^{\pi/2} 1 d\theta = \frac{1}{4}\pi + 2.$$

This example illustrates the following procedure for finding the area between the graphs of two curves given in polar coordinates.

Procedure 20.2.4 Suppose that for all $\theta \in [a, b]$, $0 < g(\theta) < f(\theta)$. To find the area of the region defined in terms of polar coordinates by $g(\theta) < r < f(\theta)$, $\theta \in [a, b]$, you do the following.

$$\frac{1}{2} \int_a^b (f(\theta)^2 - g(\theta)^2) d\theta.$$

20.3 Exercises

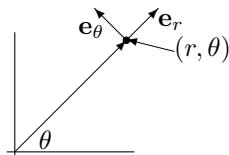
- The following are the polar coordinates of points. Find the rectangular coordinates.
 - $(5, \frac{\pi}{6})$
 - $(3, \frac{\pi}{3})$
 - $(4, \frac{2\pi}{3})$
 - $(2, \frac{3\pi}{4})$
 - $(3, \frac{7\pi}{6})$
 - $(8, \frac{11\pi}{6})$
- The following are the rectangular coordinates of points. Find the polar coordinates of these points.
 - $(\frac{5}{2}\sqrt{2}, \frac{5}{2}\sqrt{2})$
 - $(\frac{3}{2}, \frac{3}{2}\sqrt{3})$
 - $(-\frac{5}{2}\sqrt{2}, \frac{5}{2}\sqrt{2})$
 - $(-\frac{5}{2}, \frac{5}{2}\sqrt{3})$
 - $(-\sqrt{3}, -1)$
 - $(\frac{3}{2}, -\frac{3}{2}\sqrt{3})$
- In general it is a stupid idea to try to use algebra to invert and solve for a set of curvilinear coordinates such as polar or cylindrical coordinates in terms of Cartesian coordinates. Not only is it often very difficult or even impossible to do it¹, but also it takes you in entirely the wrong direction because the whole point of introducing the new coordinates is to write everything in terms of these new coordinates and not in terms of Cartesian coordinates. However, sometimes this inversion can be done. Describe how to solve for r and θ in terms of x and y in polar coordinates.
- Suppose $r = \frac{a}{1+e \sin \theta}$ where $e \geq 0$. By changing to rectangular coordinates, show that this is either a parabola, an ellipse or a hyperbola. Determine the values of e which correspond to the various cases.
- In Example 20.1.4 suppose you graphed it for $\theta \in [0, k\pi]$ where k is a positive integer. What is the small-

¹It is no problem for these simple cases of curvilinear coordinates. However, it is a major difficulty in general. Algebra is simply not adequate to solve systems of nonlinear equations.

- est value of k such that the graph will look exactly like the one presented in the example?
6. Suppose you were to graph $r = 3 + \sin\left(\frac{m}{n}\theta\right)$ where m, n are integers. Can you give some description of what the graph will look like for $\theta \in [0, k\pi]$ for k a very large positive integer? How would things change if you did $r = 3 + \sin(\alpha\theta)$ where α is an irrational number?
 7. Graph $r = 1 + \sin\theta$ for $\theta \in [0, 2\pi]$.
 8. Graph $r = 2 + \sin\theta$ for $\theta \in [0, 2\pi]$.
 9. Graph $r = 1 + 2\sin\theta$ for $\theta \in [0, 2\pi]$.
 10. Graph $r = 2 + \sin(2\theta)$ for $\theta \in [0, 2\pi]$.
 11. Graph $r = 1 + \sin(2\theta)$ for $\theta \in [0, 2\pi]$.
 12. Graph $r = 1 + \sin(3\theta)$ for $\theta \in [0, 2\pi]$.
 13. Find the area of the bounded region determined by $r = 1 + \sin(3\theta)$ for $\theta \in [0, 2\pi]$.
 14. Find the area inside $r = 1 + \sin\theta$ and outside the circle $r = 1/2$.
 15. Find the area inside the circle $r = 1/2$ and outside the region defined by $r = 1 + \sin\theta$.

20.4 The Acceleration In Polar Coordinates

Sometimes you have information about forces which act not in the direction of the coordinate axes but in some other direction. When this is the case, it is often useful to express things in terms of different coordinates which are consistent with these directions. A good example of this is the force exerted by the sun on a planet. This force is always directed toward the sun and so the force vector changes as the planet moves. To discuss this, consider the following simple diagram in which two unit vectors, $\mathbf{e}_r$ and $\mathbf{e}_\theta$ are shown.



The vector, $\mathbf{e}_r = (\cos\theta, \sin\theta)$ and the vector, $\mathbf{e}_\theta = (-\sin\theta, \cos\theta)$. You should convince yourself that the picture above corresponds to this definition of the two vectors. Note that $\mathbf{e}_r$ is a unit vector pointing away from $\mathbf{0}$ and

$$\mathbf{e}_\theta = \frac{d\mathbf{e}_r}{d\theta}, \quad \mathbf{e}_r = -\frac{d\mathbf{e}_\theta}{d\theta}. \quad (20.3)$$

Now consider the position vector from $\mathbf{0}$ of a point in the plane, $\mathbf{r}(t)$. Then

$$\mathbf{r}(t) = r(t) \mathbf{e}_r(\theta(t))$$

where $r(t) = |\mathbf{r}(t)|$. Thus $r(t)$ is just the distance from the origin, $\mathbf{0}$ to the point. What is the velocity and acceleration? Using the chain rule,

$$\frac{d\mathbf{e}_r}{dt} = \frac{d\mathbf{e}_r}{d\theta} \theta'(t), \quad \frac{d\mathbf{e}_\theta}{dt} = \frac{d\mathbf{e}_\theta}{d\theta} \theta'(t)$$

and so from 20.3,

$$\frac{d\mathbf{e}_r}{dt} = \theta'(t) \mathbf{e}_\theta, \quad \frac{d\mathbf{e}_\theta}{dt} = -\theta'(t) \mathbf{e}_r \quad (20.4)$$

Using 20.4 as needed along with the product rule and the chain rule,

$$\begin{aligned} \mathbf{r}'(t) &= r'(t) \mathbf{e}_r + r(t) \frac{d}{dt}(\mathbf{e}_r(\theta(t))) \\ &= r'(t) \mathbf{e}_r + r(t) \theta'(t) \mathbf{e}_\theta. \end{aligned}$$

Next consider the acceleration.

$$\begin{aligned}
 \mathbf{r}''(t) &= r''(t) \mathbf{e}_r + r'(t) \frac{d\mathbf{e}_r}{dt} + r'(t) \theta'(t) \mathbf{e}_\theta + r(t) \theta''(t) \mathbf{e}_\theta + r(t) \theta'(t) \frac{d}{dt}(\mathbf{e}_\theta) \\
 &= r''(t) \mathbf{e}_r + 2r'(t) \theta'(t) \mathbf{e}_\theta + r(t) \theta''(t) \mathbf{e}_\theta + r(t) \theta'(t) (-\mathbf{e}_r) \theta'(t) \\
 &= \left(r''(t) - r(t) \theta'(t)^2 \right) \mathbf{e}_r + \left(2r'(t) \theta'(t) + r(t) \theta''(t) \right) \mathbf{e}_\theta.
 \end{aligned} \tag{20.5}$$

This is a very profound formula. Consider the following examples.

Example 20.4.1 Suppose an object of mass m moves at a uniform speed, s , around a circle of radius R . Find the force acting on the object.

By Newton's second law, the force acting on the object is $m\mathbf{r}''$. In this case, $r(t) = R$, a constant and since the speed is constant, $\theta'' = 0$. Therefore, the term in 20.5 corresponding to $\mathbf{e}_\theta$ equals zero and $m\mathbf{r}'' = -R\theta'(t)^2 \mathbf{e}_r$. The speed of the object is s and so it moves s/R radians in unit time. Thus $\theta'(t) = s/R$ and so

$$m\mathbf{r}'' = -mR \left(\frac{s}{R} \right)^2 \mathbf{e}_r = -m \frac{s^2}{R} \mathbf{e}_r.$$

This is the familiar formula for centripetal force from elementary physics, obtained as a very special case of 20.5.

Example 20.4.2 A platform rotates at a constant speed in the counter clockwise direction and an object of mass m moves from the center of the platform toward the edge at constant speed. What forces act on this object?

Let v denote the constant speed of the object moving toward the edge of the platform. Then

$$r'(t) = v, r''(t) = 0, \theta''(t) = 0,$$

while $\theta'(t) = \omega$, a positive constant. From 20.5

$$m\mathbf{r}''(t) = -mr(t) \omega^2 \mathbf{e}_r + m2v\omega \mathbf{e}_\theta.$$

Thus the object experiences centripetal force from the first term and also a funny force from the second term which is in the direction of rotation of the platform. You can observe this by experiment if you like. Go to a playground and have someone spin one of those merry go rounds while you ride it and move from the center toward the edge. The term $2r'\theta'$ is called the Coriolis force.

Suppose at each point of space, $\mathbf{r}$ is associated a force, $\mathbf{F}(\mathbf{r})$ which a given object of mass m will experience if its position vector is $\mathbf{r}$. This is called a force field. a force field is a central force field if $\mathbf{F}(\mathbf{r}) = g(\mathbf{r}) \mathbf{e}_r$. Thus in a central force field, the force an object experiences will always be directed toward or away from the origin, $\mathbf{0}$. The following simple lemma is very interesting because it says that in a central force field, objects must move in a plane.

Lemma 20.4.3 Suppose an object moves in three dimensions in such a way that the only force acting on the object is a central force. Then the motion of the object is in a plane.

Proof: Let $\mathbf{r}(t)$ denote the position vector of the object. Then from the definition of a central force and Newton's second law,

$$m\mathbf{r}'' = g(\mathbf{r}) \mathbf{r}.$$

Therefore, $m\mathbf{r}'' \times \mathbf{r} = m(\mathbf{r}' \times \mathbf{r})' = g(\mathbf{r}) \mathbf{r} \times \mathbf{r} = \mathbf{0}$. Therefore, $(\mathbf{r}' \times \mathbf{r}) = \mathbf{n}$, a constant vector and so $\mathbf{r} \cdot \mathbf{n} = \mathbf{r} \cdot (\mathbf{r}' \times \mathbf{r}) = 0$ showing that $\mathbf{n}$ is a normal vector to a plane which contains $\mathbf{r}(t)$ for all t . This proves the lemma.

20.5 Planetary Motion

Kepler's laws of planetary motion state that planets move around the sun along an ellipse, the equal area law described above holds, and there is a formula for the time it takes for the planet to move around the sun. These laws, discovered by Kepler, were shown by Newton to be consequences of his law of gravitation which states that the force acting on a mass, m by a mass, M is given by

$$\mathbf{F} = -GMm \left(\frac{1}{r^3} \right) \mathbf{r} = -GMm \left(\frac{1}{r^2} \right) \mathbf{e}_r$$

where r is the distance between centers of mass and $\mathbf{r}$ is the position vector from M to m . Here G is the gravitation constant. This is called an inverse square law. Gravity acts according to this law and so does electrostatic force. The constant, G , is very small when usual units are used and it has been computed using a very delicate experiment. It is now accepted to be

$$6.67 \times 10^{-11} \text{ Newton meter}^2/\text{kilogram}^2.$$

The experiment involved a light source shining on a mirror attached to a quartz fiber from which was suspended a long rod with two equal masses at the ends which were attracted by two larger masses. The gravitation force between the suspended masses and the two large masses caused the fibre to twist ever so slightly and this twisting was measured by observing the deflection of the light reflected from the mirror on a scale placed some distance from the fibre. The constant was first measured successfully by Lord Cavendish in 1798 and the present accepted value was obtained in 1942. Experiments like these are major accomplishments.

In the following argument, M is the mass of the sun and m is the mass of the planet. (It could also be a comet or an asteroid.)

20.5.1 The Equal Area Rule

An object moves in three dimensions in such a way that the only force acting on the object is a central force. Then the object moves in a plane and the radius vector from the origin to the object sweeps out area at a constant rate. This is the equal area rule. In the context of planetary motion it is called Kepler's second law.

Lemma 20.4.3 says the object moves in a plane. From the assumption that the force field is a central force field, it follows from 20.5 that

$$2r'(t)\theta'(t) + r(t)\theta''(t) = 0$$

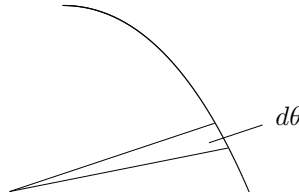
Multiply both sides of this equation by r . This yields

$$2rr'\theta' + r^2\theta'' = (r^2\theta')' = 0. \quad (20.6)$$

Consequently,

$$r^2\theta' = c \quad (20.7)$$

for some constant, C . Now consider the following picture.



In this picture, $d\theta$ is the indicated angle and the two lines determining this angle are position vectors for the object at point t and point $t + dt$. The area of the sector, dA , is essentially $r^2 d\theta$ and so $dA = \frac{1}{2} r^2 d\theta$. Therefore,

$$\frac{dA}{dt} = \frac{1}{2} r^2 \frac{d\theta}{dt} = \frac{c}{2}. \quad (20.8)$$

20.5.2 Inverse Square Law Motion, Kepler's First Law

Consider the first of Kepler's laws, the one which states that planets move along ellipses. From Lemma 20.4.3, the motion is in a plane. Now from 20.5 and Newton's second law,

$$\begin{aligned} & \left(r''(t) - r(t) \theta'(t)^2 \right) \mathbf{e}_r + (2r'(t) \theta'(t) + r(t) \theta''(t)) \mathbf{e}_\theta \\ &= -\frac{GMm}{m} \left(\frac{1}{r^2} \right) \mathbf{e}_r = -k \left(\frac{1}{r^2} \right) \mathbf{e}_r \end{aligned}$$

Thus $k = GM$ and

$$r''(t) - r(t) \theta'(t)^2 = -k \left(\frac{1}{r^2} \right), \quad 2r'(t) \theta'(t) + r(t) \theta''(t) = 0. \quad (20.9)$$

As in 20.6, $(r^2 \theta')' = 0$ and so there exists a constant, c , such that

$$r^2 \theta' = c. \quad (20.10)$$

Now the other part of 20.9 and 20.10 implies

$$r''(t) - r(t) \theta'(t)^2 = r''(t) - r(t) \left(\frac{c^2}{r^4} \right) = -k \left(\frac{1}{r^2} \right). \quad (20.11)$$

It is only r as a function of θ which is of interest. Using the chain rule,

$$r' = \frac{dr}{d\theta} \frac{d\theta}{dt} = \frac{dr}{d\theta} \left(\frac{c}{r^2} \right) \quad (20.12)$$

and so also

$$\begin{aligned} r'' &= \frac{d^2 r}{d\theta^2} \left(\frac{d\theta}{dt} \right) \left(\frac{c}{r^2} \right) + \frac{dr}{d\theta} (-2) (c) (r^{-3}) \frac{dr}{d\theta} \frac{d\theta}{dt} \\ &= \frac{d^2 r}{d\theta^2} \left(\frac{c}{r^2} \right)^2 - 2 \left(\frac{dr}{d\theta} \right)^2 \left(\frac{c^2}{r^5} \right) \end{aligned} \quad (20.13)$$

Using 20.13 and 20.12 in 20.11 yields

$$\frac{d^2 r}{d\theta^2} \left(\frac{c}{r^2} \right)^2 - 2 \left(\frac{dr}{d\theta} \right)^2 \left(\frac{c^2}{r^5} \right) - r(t) \left(\frac{c^2}{r^4} \right) = -k \left(\frac{1}{r^2} \right).$$

Now multiply both sides of this equation by r^4/c^2 to obtain

$$\frac{d^2 r}{d\theta^2} - 2 \left(\frac{dr}{d\theta} \right)^2 \frac{1}{r} - r = \frac{-kr^2}{c^2}. \quad (20.14)$$

This is a nice differential equation for r as a function of θ but it is not clear what its solution is. It turns out to be convenient to define a new dependent variable, $\rho \equiv r^{-1}$ so $r = \rho^{-1}$. Then

$$\frac{dr}{d\theta} = (-1) \rho^{-2} \frac{d\rho}{d\theta}, \quad \frac{d^2 r}{d\theta^2} = 2\rho^{-3} \left(\frac{d\rho}{d\theta} \right)^2 + (-1) \rho^{-2} \frac{d^2 \rho}{d\theta^2}.$$

Substituting this in to 20.14 yields

$$2\rho^{-3} \left(\frac{d\rho}{d\theta} \right)^2 + (-1) \rho^{-2} \frac{d^2\rho}{d\theta^2} - 2 \left(\rho^{-2} \frac{d\rho}{d\theta} \right)^2 \rho - \rho^{-1} = \frac{-k\rho^{-2}}{c^2}.$$

which simplifies to

$$(-1) \rho^{-2} \frac{d^2\rho}{d\theta^2} - \rho^{-1} = \frac{-k\rho^{-2}}{c^2}$$

since those two terms which involve $\left(\frac{d\rho}{d\theta} \right)^2$ cancel. Now multiply both sides by $-\rho^2$ and this yields

$$\frac{d^2\rho}{d\theta^2} + \rho = \frac{k}{c^2}, \quad (20.15)$$

which is a much nicer differential equation. Let $R = \rho - \frac{k}{c^2}$. Then in terms of R , this differential equation is

$$\frac{d^2R}{d\theta^2} + R = 0.$$

Multiply both sides by $\frac{dR}{d\theta}$.

$$\frac{1}{2} \frac{d}{d\theta} \left(\left(\frac{dR}{d\theta} \right)^2 + R^2 \right) = 0$$

and so

$$\left(\frac{dR}{d\theta} \right)^2 + R^2 = \delta^2 \quad (20.16)$$

for some $\delta > 0$. Therefore, there exists an angle, $\psi = \psi(\theta)$ such that

$$R = \delta \sin(\psi), \quad \frac{dR}{d\theta} = \delta \cos(\psi)$$

because 20.16 says $(\frac{1}{\delta} \frac{dR}{d\theta}, \frac{1}{\delta} R)$ is a point on the unit circle. But differentiating, the first of the above equations,

$$\frac{dR}{d\theta} = \delta \cos(\psi) \frac{d\psi}{d\theta} = \delta \cos(\psi)$$

and so $\frac{d\psi}{d\theta} = 1$. Therefore, $\psi = \theta + \phi$. Choosing the coordinate system appropriately, you can assume $\phi = 0$. Therefore,

$$R = \rho - \frac{k}{c^2} = \frac{1}{r} - \frac{k}{c^2} = \delta \sin(\theta)$$

and so, solving for r ,

$$\begin{aligned} r &= \frac{1}{\left(\frac{k}{c^2} \right) + \delta \sin \theta} = \frac{c^2/k}{1 + (c^2/k) \delta \sin \theta} \\ &= \frac{p\varepsilon}{1 + \varepsilon \sin \theta} \end{aligned}$$

where

$$\varepsilon = (c^2/k) \delta \text{ and } p = c^2/k\varepsilon. \quad (20.17)$$

Here all these constants are nonnegative.

Thus

$$r + \varepsilon r \sin \theta = \varepsilon p$$

and so $r = (\varepsilon p - \varepsilon y)$. Then squaring both sides,

$$x^2 + y^2 = (\varepsilon p - \varepsilon y)^2 = \varepsilon^2 p^2 - 2p\varepsilon^2 y + \varepsilon^2 y^2$$

And so

$$x^2 + (1 - \varepsilon^2) y^2 = \varepsilon^2 p^2 - 2p\varepsilon^2 y. \quad (20.18)$$

In case $\varepsilon = 1$, this reduces to the equation of a parabola. If $\varepsilon < 1$, this reduces to the equation of an ellipse and if $\varepsilon > 1$, this is called a hyperbola. This proves that objects which are acted on only by a force of the form given in the above example move along hyperbolas, ellipses or circles. The case where $\varepsilon = 0$ corresponds to a circle. The constant, ε is called the eccentricity. This is called Kepler's first law in the case of a planet.

20.5.3 Kepler's Third Law

Kepler's third law involves the time it takes for the planet to orbit the sun. From 20.18 you can complete the square and obtain

$$x^2 + (1 - \varepsilon^2) \left(y + \frac{p\varepsilon^2}{1 - \varepsilon^2} \right)^2 = \varepsilon^2 p^2 + \frac{p^2 \varepsilon^4}{(1 - \varepsilon^2)} = \frac{\varepsilon^2 p^2}{(1 - \varepsilon^2)},$$

and this yields

$$x^2 / \left(\frac{\varepsilon^2 p^2}{1 - \varepsilon^2} \right) + \left(y + \frac{p\varepsilon^2}{1 - \varepsilon^2} \right)^2 / \left(\frac{\varepsilon^2 p^2}{(1 - \varepsilon^2)^2} \right) = 1. \quad (20.19)$$

Now note this is the equation of an ellipse and that the diameter of this ellipse is

$$\frac{2\varepsilon p}{(1 - \varepsilon^2)} \equiv 2a. \quad (20.20)$$

This follows because

$$\frac{\varepsilon^2 p^2}{(1 - \varepsilon^2)^2} \geq \frac{\varepsilon^2 p^2}{1 - \varepsilon^2}.$$

Now let T denote the time it takes for the planet to make one revolution about the sun. Using this formula, and 20.8 the following equation must hold.

$$\overbrace{\pi \frac{\varepsilon p}{\sqrt{1 - \varepsilon^2}} \frac{\varepsilon p}{(1 - \varepsilon^2)}}^{\text{area of ellipse}} = T \frac{c}{2}$$

Therefore,

$$T = \frac{2}{c} \frac{\pi \varepsilon^2 p^2}{(1 - \varepsilon^2)^{3/2}}$$

and so

$$T^2 = \frac{4\pi^2 \varepsilon^4 p^4}{c^2 (1 - \varepsilon^2)^3}$$

Now using 20.17, recalling that $k = GM$, and 20.20,

$$\begin{aligned} T^2 &= \frac{4\pi^2 \varepsilon^4 p^4}{k \varepsilon p (1 - \varepsilon^2)^3} = \frac{4\pi^2 (\varepsilon p)^3}{k (1 - \varepsilon^2)^3} \\ &= \frac{4\pi^2 a^3}{k} = \frac{4\pi^2 a^3}{GM}. \end{aligned}$$

Written more memorably, this has shown

$$T^2 = \frac{4\pi^2}{GM} \left(\frac{\text{diameter of ellipse}}{2} \right)^3. \quad (20.21)$$

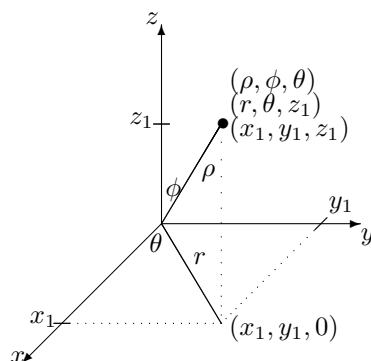
This relationship is known as Kepler's third law.

20.6 Exercises

1. Suppose you know how the spherical coordinates of a moving point change as a function of t . Can you figure out the velocity of the point? Specifically, suppose $\phi(t) = t$, $\theta(t) = 1 + t$, and $\rho(t) = t$. Find the velocity of the object in terms of Cartesian coordinates. **Hint:** You would need to find $x'(t)$, $y'(t)$, and $z'(t)$. Then in terms of Cartesian coordinates, the velocity would be $x'(t)\mathbf{i} + y'(t)\mathbf{j} + z'(t)\mathbf{k}$.
2. Find the length of the cardioid, $r = 1 + \cos \theta$, $\theta \in [0, 2\pi]$. **Hint:** A parameterization is $x(\theta) = (1 + \cos \theta) \cos \theta$, $y(\theta) = (1 + \cos \theta) \sin \theta$.
3. In general, show that the length of the curve given in polar coordinates by $r = f(\theta)$, $\theta \in [a, b]$ equals $\int_a^b \sqrt{f'(\theta)^2 + f(\theta)^2} d\theta$.
4. Suppose the curve given in polar coordinates by $r = f(\theta)$ for $\theta \in [a, b]$ is rotated about the y axis. Find a formula for the resulting surface of revolution.
5. Suppose an object moves in such a way that $r^2\theta'$ is a constant. Show that the only force acting on the object is a central force.
6. Explain why low pressure areas rotate counter clockwise in the Northern hemisphere and clockwise in the Southern hemisphere. **Hint:** Note that from the point of view of an observer fixed in space above the North pole, the low pressure area already has a counter clockwise rotation because of the rotation of the earth and its spherical shape. Now consider 20.7. In the low pressure area stuff will move toward the center so r gets smaller. How are things different in the Southern hemisphere?
7. What are some physical assumptions which are made in the above derivation of Kepler's laws from Newton's laws of motion?
8. The orbit of the earth is pretty nearly circular and the distance from the sun to the earth is about 149×10^6 kilometers. Using 20.21 and the above value of the universal gravitation constant, determine the mass of the sun. The earth goes around it in 365 days. (Actually it is 365.256 days.)
9. It is desired to place a satellite above the equator of the earth which will rotate about the center of mass of the earth every 24 hours. Is it necessary that the orbit be circular? What if you want the satellite to stay above the same point on the earth at all times? If the orbit is to be circular and the satellite is to stay above the same point, at what distance from the center of mass of the earth should the satellite be? You may use that the mass of the earth is 5.98×10^{24} kilograms. Such a satellite is called geosynchronous.

20.7 Spherical And Cylindrical Coordinates

Now consider two three dimensional generalizations of polar coordinates. The following picture serves as motivation for the definition of these two other coordinate systems.



In this picture, ρ is the distance between the origin, the point whose Cartesian coordinates are $(0, 0, 0)$ and the point indicated by a dot and labelled as (x_1, y_1, z_1) , (r, θ, z_1) , and (ρ, ϕ, θ) . The angle between the positive z axis and the line between the origin and the point indicated by a dot is denoted by ϕ , and θ , is the angle between the positive x axis and the line joining the origin to the point $(x_1, y_1, 0)$ as shown, while r is the length of this line. Thus r and θ determine a point in the plane determined by letting $z = 0$ and r and θ are the usual polar coordinates. Thus $r \geq 0$ and $\theta \in [0, 2\pi)$. Letting z_1 denote the usual z coordinate of a point in three dimensions, like the one shown as a dot, (r, θ, z_1) are the cylindrical coordinates of the dotted point. The spherical coordinates are determined by (ρ, ϕ, θ) . When ρ is specified, this indicates that the point of interest is on some sphere of radius ρ which is centered at the origin. Then when ϕ is given, the location of the point is narrowed down to a circle and finally, θ determines which point is on this circle. Let $\phi \in [0, \pi]$, $\theta \in [0, 2\pi)$, and $\rho \in [0, \infty)$. The picture shows how to relate these new coordinate systems to Cartesian coordinates. For Cylindrical coordinates,

$$\begin{aligned}x &= r \cos(\theta), \\y &= r \sin(\theta), \\z &= z\end{aligned}$$

and for spherical coordinates,

$$\begin{aligned}x &= \rho \sin(\phi) \cos(\theta), \\y &= \rho \sin(\phi) \sin(\theta), \\z &= \rho \cos(\phi).\end{aligned}$$

Spherical coordinates should be especially interesting to you because you live on the surface of a sphere. This has been known for several hundred years. You may also know that the standard way to determine position on the earth is to give the longitude and latitude. The latitude corresponds to ϕ and the longitude corresponds to θ .²

Example 20.7.1 Express the surface, $z = \frac{1}{\sqrt{3}}\sqrt{x^2 + y^2}$ in spherical coordinates.

²Actually latitude is determined on maps and in navigation by measuring the angle from the equator rather than the pole but it is essentially the same idea that we have presented here.

This is

$$\rho \cos(\phi) = \frac{1}{\sqrt{3}} \sqrt{(\rho \sin(\phi) \cos(\theta))^2 + (\rho \sin(\phi) \sin(\theta))^2} = \frac{1}{3} \sqrt{3} \rho \sin \phi.$$

Therefore, this reduces to

$$\tan \phi = \sqrt{3}$$

and so this is just $\phi = \pi/3$.

Example 20.7.2 Express the surface, $y = x$ in terms of spherical coordinates.

This says $\rho \sin(\phi) \sin(\theta) = \rho \sin(\phi) \cos(\theta)$. Thus $\sin \theta = \cos \theta$. You could also write $\tan \theta = 1$.

Example 20.7.3 Express the surface, $x^2 + y^2 = 4$ in cylindrical coordinates.

This says $r^2 \cos^2 \theta + r^2 \sin^2 \theta = 4$. Thus $r = 2$.

20.8 Exercises

- The following are the cylindrical coordinates of points, (r, θ, z) . Find the rectangular and spherical coordinates.
 - $(5, \frac{5\pi}{6}, -3)$
 - $(3, \frac{\pi}{3}, 4)$
 - $(4, \frac{2\pi}{3}, 1)$
 - $(2, \frac{3\pi}{4}, -2)$
 - $(3, \frac{3\pi}{2}, -1)$
 - $(8, \frac{11\pi}{6}, -11)$
- The following are the rectangular coordinates of points, (x, y, z) . Find the cylindrical and spherical coordinates of these points.
 - $(\frac{5}{2}\sqrt{2}, \frac{5}{2}\sqrt{2}, -3)$
 - $(\frac{3}{2}, \frac{3}{2}\sqrt{3}, 2)$
 - $(-\frac{5}{2}\sqrt{2}, \frac{5}{2}\sqrt{2}, 11)$
 - $(-\frac{5}{2}, \frac{5}{2}\sqrt{3}, 23)$
 - $(-\sqrt{3}, -1, -5)$
 - $(\frac{3}{2}, -\frac{3}{2}\sqrt{3}, -7)$
- The following are spherical coordinates of points in the form (ρ, ϕ, θ) . Find the rectangular and cylindrical coordinates. The rectangular coordinates (x, y, z) and cylindrical coordinates (r, θ, z) are as listed.
 - $(4, \frac{\pi}{4}, \frac{5\pi}{6})$
 - $(2, \frac{\pi}{3}, \frac{2\pi}{3})$
 - $(3, \frac{5\pi}{6}, \frac{3\pi}{2})$
 - $(4, \frac{\pi}{2}, \frac{7\pi}{4})$
 - $(4, \frac{2\pi}{3}, \frac{\pi}{6})$
 - $(4, \frac{3\pi}{4}, \frac{5\pi}{3})$
- The following are rectangular coordinates of points, (x, y, z) . Find the spherical and cylindrical coordinates.
 - $(\sqrt{2}, \sqrt{6}, 2\sqrt{2})$
 - $(-\frac{1}{2}\sqrt{3}, \frac{3}{2}, 1)$
 - $(-\frac{3}{4}\sqrt{2}, \frac{3}{4}\sqrt{2}, -\frac{3}{2}\sqrt{3})$
 - $(-\sqrt{3}, 1, 2\sqrt{3})$
 - $(-\frac{1}{4}\sqrt{2}, \frac{1}{4}\sqrt{6}, -\frac{1}{2}\sqrt{2})$
- A point has Cartesian coordinates, $(1, 2, 3)$. Find its spherical and cylindrical coordinates using a calculator or other electronic gadget.
- Describe the following surface in rectangular coordinates. $\phi = \pi/4$ where ϕ is the polar angle in spherical coordinates.
- Describe the following surface. $\theta = \pi/4$ where θ is the angle measured from the positive x axis in spherical coordinates.
- Describe the following surface in rectangular coordinates. $r = 5$ where r is one of the cylindrical coordinates.

9. Describe the following surface in rectangular coordinates. $\rho = 4$ where ρ is the distance to the origin.
10. Give the cone, $z = \sqrt{x^2 + y^2}$ in cylindrical coordinates and in spherical coordinates.
11. Write the following in spherical coordinates.
 - (a) $z = x^2 + y^2$.
 - (b) $x^2 - y^2 = 1$
 - (c) $z^2 + x^2 + y^2 = 6$
 - (d) $z = \sqrt{x^2 + y^2}$
- (e) $y = x$
 - (f) $z = x$
12. Write the following in cylindrical coordinates.
 - (a) $z = x^2 + y^2$.
 - (b) $x^2 - y^2 = 1$
 - (c) $z^2 + x^2 + y^2 = 6$
 - (d) $z = \sqrt{x^2 + y^2}$
 - (e) $y = x$
 - (f) $z = x$

Part V

Vector Calculus In Many
Variables

Functions Of Many Variables

21.0.1 Outcomes

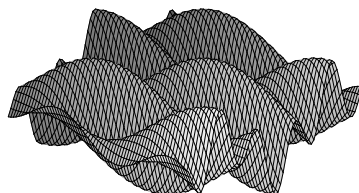
1. Represent a function of two variables by level curves.
2. Identify the characteristics of a function from a graph of its level curves.
3. Recall and use the concept of limit point.
4. Describe the geometrical significance of a directional derivative.
5. Give the relationship between partial derivatives and directional derivatives.
6. Compute partial derivatives and directional derivatives from their definitions.
7. Evaluate higher order partial derivatives.
8. State conditions under which mixed partial derivatives are equal.
9. Verify equations involving partial derivatives.
10. Describe the gradient of a scalar valued function and use to compute the directional derivative.
11. Explain why the directional derivative is maximized in the direction of the gradient and minimized in the direction of minus the gradient.

21.1 The Graph Of A Function Of Two Variables

With vector valued functions of many variables, it doesn't take long before it is impossible to draw meaningful pictures. This is because one needs more than three dimensions to accomplish the task and we can only visualize things in three dimensions. Ultimately, one of the main purposes of calculus is to free us from the tyranny of art. In calculus, we are permitted and even required to think in a meaningful way about things which cannot be drawn. However, it is certainly interesting to consider some things which can be visualized and this will help to formulate and understand more general notions which make sense in contexts which cannot be visualized. One of these is the concept of a scalar valued function of two variables.

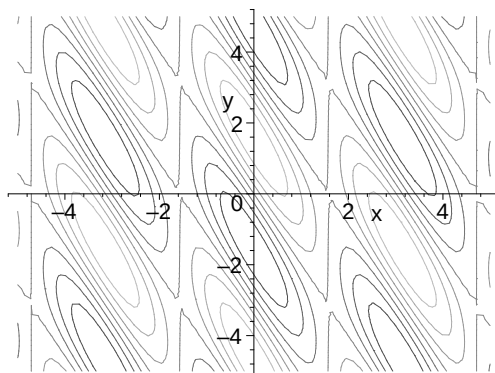
Let $f(x, y)$ denote a scalar valued function of two variables evaluated at the point (x, y) . Its graph consists of the set of points, (x, y, z) such that $z = f(x, y)$. How does one go about depicting such a graph? The usual way is to fix one of the variables, say x and consider the function $z = f(x, y)$ where y is allowed to vary and x is fixed. Graphing this would give a curve which lies in the surface to be depicted. Then do the same thing for other values of x and the result would depict the desired graph. Computers do this

very well. The following is the graph of the function $z = \cos(x) \sin(2x + y)$ drawn using Maple, a computer algebra system.¹



Notice how elaborate this picture is. The lines in the drawing correspond to taking one of the variables constant and graphing the curve which results. The computer did this drawing in seconds but you couldn't do it as well if you spent all day on it. I used a grid consisting of 70 choices for x and 70 choices for y .

Sometimes attempts are made to understand three dimensional objects like the above graph by looking at contour graphs in two dimensions. The contour graph of the above three dimensional graph is below and comes from using the computer algebra system again.



This is in two dimensions and the different lines in two dimensions correspond to points on the three dimensional graph which have the same z value. If you have looked at a weather map, these lines are called isotherms or isobars depending on whether the function involved is temperature or pressure. In a contour geographic map, the contour lines represent constant altitude. If many contour lines are close to each other, this indicates rapid change in the altitude, temperature, pressure, or whatever else may be measured.

A scalar function of three variables, cannot be visualized because four dimensions are required. However, some people like to try and visualize even these examples. This is done by looking at level surfaces in $\mathbb{R}^3$ which are defined as surfaces where the function assumes a constant value. They play the role of contour lines for a function of two variables. As a simple example, consider $f(x, y, z) = x^2 + y^2 + z^2$. The level surfaces of this function would be concentric spheres centered at $\mathbf{0}$. (Why?) Another way to visualize objects in higher dimensions involves the use of color and animation. However, there really are limits to what you can accomplish in this direction. So much for art.

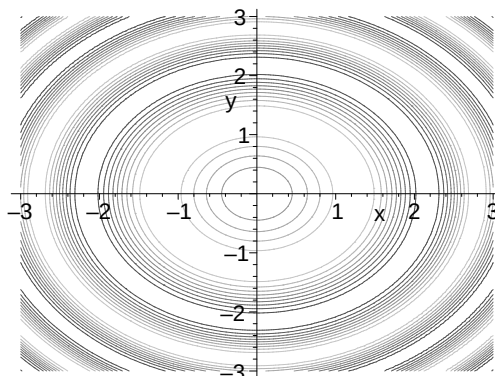
However, the concept of level curves is quite useful because these can be drawn.

Example 21.1.1 Determine from a contour map where the function,

$$f(x, y) = \sin(x^2 + y^2)$$

¹I used Maple and exported the graph as a file which I then imported into this document.

is steepest.



In the picture, the steepest places are where the contour lines are close together because they correspond to various values of the function. You can look at the picture and see where they are close and where they are far. This is the advantage of a contour map.

21.2 Review Of Limits

Recall the concept of limit of a function of many variables. When $\mathbf{f} : D(\mathbf{f}) \rightarrow \mathbb{R}^q$ one can only consider in a meaningful way limits at limit points of the set, $D(\mathbf{f})$.

Definition 21.2.1 Let A denote a nonempty subset of $\mathbb{R}^p$. A point, $\mathbf{x}$ is said to be a **limit point** of the set, A if for every $r > 0$, $B(\mathbf{x}, r)$ contains infinitely many points of A .

Example 21.2.2 Let S denote the set, $\{(x, y, z) \in \mathbb{R}^3 : x, y, z \text{ are all in } \mathbb{N}\}$. Which points are limit points?

This set does not have any because any two of these points are at least as far apart as 1. Therefore, if $\mathbf{x}$ is any point of $\mathbb{R}^3$, $B(\mathbf{x}, 1/4)$ contains at most one point.

Example 21.2.3 Let U be an open set in $\mathbb{R}^3$. Which points of U are limit points of U ?

They all are. From the definition of U being open, if $\mathbf{x} \in U$, There exists $B(\mathbf{x}, r) \subseteq U$ for some $r > 0$. Now consider the line segment $\mathbf{x} + t\mathbf{e}_1$ where $t \in [0, 1/2]$. This describes infinitely many points and they are all in $B(\mathbf{x}, r)$ because

$$|\mathbf{x} + t\mathbf{e}_1 - \mathbf{x}| = tr < r.$$

Therefore, every point of U is a limit point of U .

The case where U is open will be the one of most interest but many other sets have limit points.

Definition 21.2.4 Let $\mathbf{f} : D(\mathbf{f}) \subseteq \mathbb{R}^p \rightarrow \mathbb{R}^q$ where $q, p \geq 1$ be a function and let $\mathbf{x}$ be a limit point of $D(\mathbf{f})$. Then

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}$$

if and only if the following condition holds. For all $\varepsilon > 0$ there exists $\delta > 0$ such that if

$$0 < |\mathbf{y} - \mathbf{x}| < \delta \text{ and } \mathbf{y} \in D(\mathbf{f})$$

then,

$$|\mathbf{L} - \mathbf{f}(\mathbf{y})| < \varepsilon.$$

The condition that $\mathbf{x}$ must be a limit point of $D(\mathbf{f})$ if you are to take a limit at $\mathbf{x}$ is what makes the limit well defined.

Proposition 21.2.5 *Let $\mathbf{f} : D(\mathbf{f}) \subseteq \mathbb{R}^p \rightarrow \mathbb{R}^q$ where $q, p \geq 1$ be a function and let $\mathbf{x}$ be a limit point of $D(\mathbf{f})$. Then if $\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y})$ exists, it must be unique.*

Proof: Suppose $\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}_1$ and $\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}_2$. Then for $\varepsilon > 0$ given, let $\delta_i > 0$ correspond to $\mathbf{L}_i$ in the definition of the limit and let $\delta = \min(\delta_1, \delta_2)$. Since $\mathbf{x}$ is a limit point, there exists $\mathbf{y} \in B(\mathbf{x}, \delta) \cap D(\mathbf{f})$. Therefore,

$$\begin{aligned} |\mathbf{L}_1 - \mathbf{L}_2| &\leq |\mathbf{L}_1 - \mathbf{f}(\mathbf{y})| + |\mathbf{f}(\mathbf{y}) - \mathbf{L}_2| \\ &< \varepsilon + \varepsilon = 2\varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this shows $\mathbf{L}_1 = \mathbf{L}_2$. The following theorem summarized many important interactions involving continuity. Most of this theorem has been proved in Theorem 17.5.5 on Page 415 and Theorem 17.5.7 on Page 416.

Theorem 21.2.6 *Suppose $\mathbf{x}$ is a limit point of $D(\mathbf{f})$ and*

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L}, \lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{g}(\mathbf{y}) = \mathbf{K}$$

where $\mathbf{K}$ and $\mathbf{L}$ are vectors in $\mathbb{R}^p$ for $p \geq 1$. Then if $a, b \in \mathbb{R}$,

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} a\mathbf{f}(\mathbf{y}) + b\mathbf{g}(\mathbf{y}) = a\mathbf{L} + b\mathbf{K}, \quad (21.1)$$

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f} \cdot \mathbf{g}(\mathbf{y}) = \mathbf{L} \cdot \mathbf{K} \quad (21.2)$$

Also, if $\mathbf{h}$ is a continuous function defined near $\mathbf{L}$, then

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{h} \circ \mathbf{f}(\mathbf{y}) = \mathbf{h}(\mathbf{L}). \quad (21.3)$$

For a vector valued function,

$$\mathbf{f}(\mathbf{y}) = (f_1(\mathbf{y}), \dots, f_q(\mathbf{y}))^T,$$

$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{L} = (L_1, \dots, L_k)^T$ if and only if

$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} f_k(\mathbf{y}) = L_k \quad (21.4)$$

for each $k = 1, \dots, p$.

In the case where $\mathbf{f}$ and $\mathbf{g}$ have values in $\mathbb{R}^3$

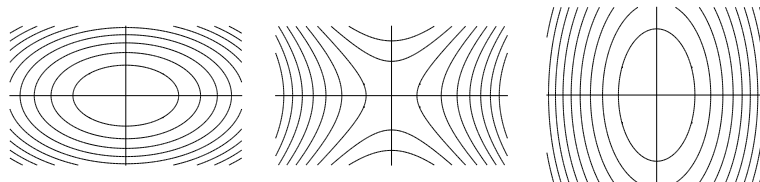
$$\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) \times \mathbf{g}(\mathbf{y}) = \mathbf{L} \times \mathbf{K}. \quad (21.5)$$

Also recall Theorem 17.5.6 on Page 416.

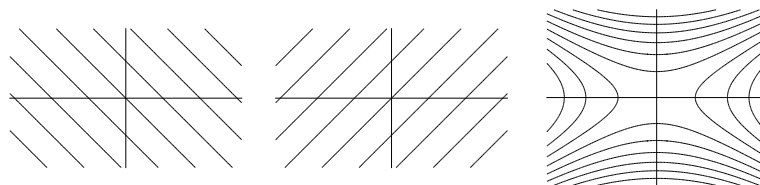
Theorem 21.2.7 *For $\mathbf{f} : D(\mathbf{f}) \rightarrow \mathbb{R}^q$ and $\mathbf{x} \in D(\mathbf{f})$ such that $\mathbf{x}$ is a limit point of $D(\mathbf{f})$, it follows $\mathbf{f}$ is continuous at $\mathbf{x}$ if and only if $\lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{f}(\mathbf{x})$.*

21.3 Exercises

- Sketch the contour graph of the function of two variables $f(x, y) = (x - 1)^2 + (y - 2)^2$.
- Which of the following functions could correspond to the following contour graphs?
 $z = x^2 + 3y^2, z = 3x^2 + y^2, z = x^2 - y^2, z = x + y$.



- Which of the following functions could correspond to the following contour graphs?
 $z = x^2 - 3y^2, z = y^2 - 3x^2, z = x - y, z = x + y$.



- Find the following limits if they exist. If they do not exist, explain why.

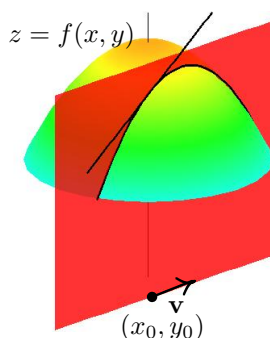
- $\lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - y^2}{x^2 + y^2}$
- $\lim_{(x,y) \rightarrow (0,0)} \frac{2x^3 + xy^2 - x^2 - 2y^2}{x^2 + 2y^2}$
- $\lim_{(x,y) \rightarrow (0,0)} \frac{\sin(x^2 + y^2)}{x^2 + y^2}$
- $\lim_{(x,y) \rightarrow (0,0)} \frac{\sin(x^2 + 2y^2)}{x^2 + 2y^2}$
- $\lim_{(x,y) \rightarrow (0,0)} \frac{\sin(x^2 + 2y^2)}{2x^2 + y^2}$
- $\lim_{(x,y) \rightarrow (0,0)} \frac{(x^2 - y^4)^2}{(x^2 + y^4)^2}$

- *Suppose f is a function defined on a set D and that $\mathbf{a} \in D$ is not a limit point of D . Show that if I define the notion of limit in the same way as above, then $\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x}) = 5$. Show that it is also the case that $\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x}) = 7$. In other words, the concept of limit is totally meaningless. This is why the insistence that the point $\mathbf{a}$ be a limit point of D .
- *Show that the definition of continuity at $\mathbf{a} \in D(\mathbf{f})$ is not dependent on $\mathbf{a}$ being a limit point of $D(\mathbf{f})$. The concept of limit and the concept of continuity are related at those points $\mathbf{a}$ which are limit points of the domain.

21.4 The Directional Derivative And Partial Derivatives

21.4.1 The Directional Derivative

The directional derivative is just what its name suggests. It is the derivative of a function in a particular direction. The following picture illustrates the situation in the case of a function of two variables.



In this picture, $\mathbf{v} \equiv (v_1, v_2)$ is a unit vector in the xy plane and $\mathbf{x}_0 \equiv (x_0, y_0)$ is a point in the xy plane. When (x, y) moves in the direction of $\mathbf{v}$, this results in a change in $z = f(x, y)$ as shown in the picture. The directional derivative in this direction is defined as

$$\lim_{t \rightarrow 0} \frac{f(x_0 + tv_1, y_0 + tv_2) - f(x_0, y_0)}{t}.$$

It tells how fast z is changing in this direction. If you looked at it from the side, you would be getting the slope of the indicated tangent line. A simple example of this is a person climbing a mountain. He could go various directions, some steeper than others. The directional derivative is just a measure of the steepness in a given direction. This motivates the following general definition of the directional derivative.

Definition 21.4.1 Let $f : U \rightarrow \mathbb{R}$ where U is an open set in $\mathbb{R}^n$ and let $\mathbf{v}$ be a unit vector. For $\mathbf{x} \in U$, define the **directional derivative** of f in the direction, $\mathbf{v}$, at the point $\mathbf{x}$ as

$$D_{\mathbf{v}}f(\mathbf{x}) \equiv \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t}.$$

Example 21.4.2 Find the directional derivative of the function, $f(x, y) = x^2y$ in the direction of $\mathbf{i} + \mathbf{j}$ at the point $(1, 2)$.

First you need a unit vector which has the same direction as the given vector. This unit vector is $\mathbf{v} \equiv \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$. Then to find the directional derivative from the definition, write the difference quotient described above. Thus $f(\mathbf{x} + t\mathbf{v}) = \left(1 + \frac{t}{\sqrt{2}}\right)^2 \left(2 + \frac{t}{\sqrt{2}}\right)$ and $f(\mathbf{x}) = 2$. Therefore,

$$\frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t} = \frac{\left(1 + \frac{t}{\sqrt{2}}\right)^2 \left(2 + \frac{t}{\sqrt{2}}\right) - 2}{t},$$

and to find the directional derivative, you take the limit of this as $t \rightarrow 0$. However, this difference quotient equals $\frac{1}{4}\sqrt{2}(10 + 4t\sqrt{2} + t^2)$ and so, letting $t \rightarrow 0$,

$$D_{\mathbf{v}}f(1, 2) = \left(\frac{5}{2}\sqrt{2}\right).$$

There is something you must keep in mind about this. The direction vector must always be a unit vector².

²Actually, there is a more general formulation of the notion of directional derivative known as the Gateaux derivative in which the length of $\mathbf{v}$ is not one but it is not considered here.

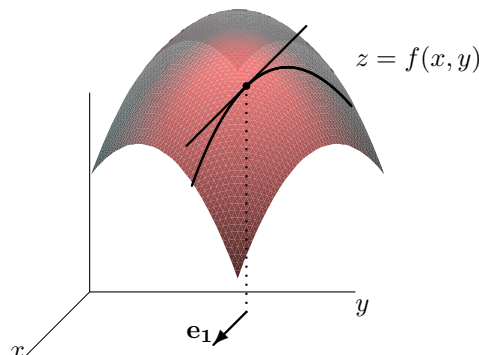
21.4.2 Partial Derivatives

There are some special unit vectors which come to mind immediately. These are the vectors, $\mathbf{e}_i$ where

$$\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)^T$$

and the 1 is in the i^{th} position.

Thus in case of a function of two variables, the directional derivative in the direction $\mathbf{i} = \mathbf{e}_1$ is the slope of the indicated straight line in the following picture.



As in the case of a general directional derivative, you fix y and take the derivative of the function, $x \rightarrow f(x, y)$. More generally, even in situations which cannot be drawn, the definition of a partial derivative is as follows.

Definition 21.4.3 Let U be an open subset of $\mathbb{R}^n$ and let $f : U \rightarrow \mathbb{R}$. Then letting $\mathbf{x} = (x_1, \dots, x_n)^T$ be a typical element of $\mathbb{R}^n$,

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) \equiv D_{\mathbf{e}_i} f(\mathbf{x}).$$

This is called the **partial derivative** of f . Thus,

$$\begin{aligned} \frac{\partial f}{\partial x_i}(\mathbf{x}) &\equiv \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{e}_i) - f(\mathbf{x})}{t} \\ &= \lim_{t \rightarrow 0} \frac{f(x_1, \dots, x_i + t, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{t}, \end{aligned}$$

and to find the partial derivative, differentiate with respect to the variable of interest and regard all the others as constants. Other notation for this partial derivative is f_{x_i} , $f_{,i}$, or $D_i f$. If $y = f(\mathbf{x})$, the partial derivative of f with respect to x_i may also be denoted by

$$\frac{\partial y}{\partial x_i} \text{ or } y_{x_i}.$$

Example 21.4.4 Find $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, and $\frac{\partial f}{\partial z}$ if $f(x, y) = y \sin x + x^2 y + z$.

From the definition above, $\frac{\partial f}{\partial x} = y \cos x + 2xy$, $\frac{\partial f}{\partial y} = \sin x + x^2$, and $\frac{\partial f}{\partial z} = 1$. Having taken one partial derivative, there is no reason to stop doing it. Thus, one could take the partial derivative with respect to y of the partial derivative with respect to x , denoted by $\frac{\partial^2 f}{\partial y \partial x}$ or f_{xy} . In the above example,

$$\frac{\partial^2 f}{\partial y \partial x} = f_{xy} = \cos x + 2x.$$

Also observe that

$$\frac{\partial^2 f}{\partial x \partial y} = f_{yx} = \cos x + 2x.$$

Higher order partial derivatives are defined by analogy to the above. Thus in the above example,

$$f_{yxx} = -\sin x + 2.$$

These partial derivatives, f_{xy} are called mixed partial derivatives.

There is an interesting relationship between the directional derivatives and the partial derivatives, provided the partial derivatives exist and are continuous.

Definition 21.4.5 Suppose $f : U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ where U is an open set and the partial derivatives of f all exist and are continuous on U . Under these conditions, define the **gradient** of f denoted $\nabla f(\mathbf{x})$ to be the vector

$$\nabla f(\mathbf{x}) = (f_{x_1}(\mathbf{x}), f_{x_2}(\mathbf{x}), \dots, f_{x_n}(\mathbf{x}))^T.$$

Proposition 21.4.6 In the situation of Definition 21.4.5 and for $\mathbf{v}$ a unit vector,

$$D_{\mathbf{v}}f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v}.$$

This proposition will be proved in a more general setting later. For now, you can use it to compute directional derivatives.

Example 21.4.7 Find the directional derivative of the function, $f(x, y) = \sin(2x^2 + y^3)$ at $(1, 1)$ in the direction $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T$.

First find the gradient.

$$\nabla f(x, y) = (4x \cos(2x^2 + y^3), 3y^2 \cos(2x^2 + y^3))^T.$$

Therefore,

$$\nabla f(1, 1) = (4 \cos(3), 3 \cos(3))^T$$

The directional derivative is therefore,

$$(4 \cos(3), 3 \cos(3))^T \cdot \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T = \frac{7}{2} (\cos 3) \sqrt{2}.$$

Another important observation is that the gradient gives the direction in which the function changes most rapidly. The following proposition will be proved later.

Proposition 21.4.8 In the situation of Definition 21.4.5, suppose $\nabla f(\mathbf{x}) \neq \mathbf{0}$. Then the direction in which f increases most rapidly, that is the direction in which the directional derivative is largest, is the direction of the gradient. Thus $\mathbf{v} = \nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$ is the unit vector which maximizes $D_{\mathbf{v}}f(\mathbf{x})$ and this maximum value is $|\nabla f(\mathbf{x})|$. Similarly, $\mathbf{v} = -\nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$ is the unit vector which minimizes $D_{\mathbf{v}}f(\mathbf{x})$ and this minimum value is $-|\nabla f(\mathbf{x})|$.

The concept of a **directional derivative for a vector valued function** is also easy to define although the geometric significance expressed in pictures is not.

Definition 21.4.9 Let $\mathbf{f} : U \rightarrow \mathbb{R}^p$ where U is an open set in $\mathbb{R}^n$ and let $\mathbf{v}$ be a unit vector. For $\mathbf{x} \in U$, define the directional derivative of $\mathbf{f}$ in the direction, $\mathbf{v}$, at the point $\mathbf{x}$ as

$$D_{\mathbf{v}}\mathbf{f}(\mathbf{x}) \equiv \lim_{t \rightarrow 0} \frac{\mathbf{f}(\mathbf{x} + t\mathbf{v}) - \mathbf{f}(\mathbf{x})}{t}.$$

Example 21.4.10 Let $\mathbf{f}(x, y) = (xy^2, yx)^T$. Find the directional derivative in the direction $(1, 2)^T$ at the point (x, y) .

First, a unit vector in this direction is $(1/\sqrt{5}, 2/\sqrt{5})^T$ and from the definition, the desired limit is

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{\left((x + t(1/\sqrt{5})) (y + t(2/\sqrt{5}))^2 - xy^2, (x + t(1/\sqrt{5})) (y + t(2/\sqrt{5})) - xy \right)}{t} \\ = \lim_{t \rightarrow 0} \left(\frac{4}{5}xy\sqrt{5} + \frac{4}{5}xt + \frac{1}{5}\sqrt{5}y^2 + \frac{4}{5}ty + \frac{4}{25}t^2\sqrt{5}, \frac{2}{5}x\sqrt{5} + \frac{1}{5}y\sqrt{5} + \frac{2}{5}t \right) \\ = \left(\frac{4}{5}xy\sqrt{5} + \frac{1}{5}\sqrt{5}y^2, \frac{2}{5}x\sqrt{5} + \frac{1}{5}y\sqrt{5} \right). \end{aligned}$$

You see from this example and the above definition that all you have to do is to form the vector which is obtained by replacing each component of the vector with its directional derivative. In particular, you can take partial derivatives of vector valued functions and use the same notation.

Example 21.4.11 Find the partial derivative with respect to x of the function

$$\mathbf{f}(x, y, z, w) = (xy^2, z \sin(xy), z^3x)^T.$$

From the above definition, $\mathbf{f}_x(x, y, z) = D_1\mathbf{f}(x, y, z) = (y^2, zy \cos(xy), z^3)^T$.

21.5 Exercises

- Find the directional derivative of $f(x, y, z) = x^2y + z^4$ in the direction of the vector, $(1, 3, -1)$ when $(x, y, z) = (1, 1, 1)$.
- Find the directional derivative of $f(x, y, z) = \sin(x + y^2) + z$ in the direction of the vector, $(1, 2, -1)$ when $(x, y, z) = (1, 1, 1)$.
- Find the directional derivative of $f(x, y, z) = \ln(x + y^2) + z^2$ in the direction of the vector, $(1, 1, -1)$ when $(x, y, z) = (1, 1, 1)$.
- Using the conclusion of Proposition 21.4.6, prove Proposition 21.4.8 from the geometric description of the dot product, the one which says the dot product is the product of the lengths of the vectors and the cosine of the included angle which is no larger than π .
- Find the largest value of the directional derivative of $f(x, y, z) = \ln(x + y^2) + z^2$ at the point $(1, 1, 1)$.
- Find the smallest value of the directional derivative of $f(x, y, z) = x \sin(4xy^2) + z^2$ at the point $(1, 1, 1)$.
- An ant falls to the top of a stove having temperature $T(x, y) = x^2 \sin(x + y)$ at the point $(2, 3)$. In what direction should the ant go to minimize the temperature? In what direction should he go to maximize the temperature?
- Find the partial derivative with respect to y of the function

$$\mathbf{f}(x, y, z, w) = (y^2, z^2 \sin(xy), z^3x)^T.$$

9. Find the partial derivative with respect to x of the function

$$\mathbf{f}(x, y, z, w) = (wx, zx \sin(xy), z^3x)^T.$$

10. Find $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$, and $\frac{\partial f}{\partial z}$ for $f =$

- (a) $x^2y + \cos(xy) + z^3y$
- (b) $e^{x^2+y^2}z \sin(x+y)$
- (c) $z^2 \sin^3(e^{x^2+y^3})$
- (d) $x^2 \cos(\sin(\tan(z^2 + y^2)))$
- (e) xy^{2+z}

11. Suppose

$$f(x, y) = \begin{cases} \frac{2xy+6x^3+12xy^2+18yx^2+36y^3+\sin(x^3)+\tan(3y^3)}{3x^2+6y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

Find $\frac{\partial f}{\partial x}(0, 0)$ and $\frac{\partial f}{\partial y}(0, 0)$.

12. Why must the vector in the definition of the directional derivative be a unit vector?

Hint: Suppose not. Would the directional derivative be a correct manifestation of steepness?

21.6 Mixed Partial Derivatives

Under certain conditions the **mixed partial derivatives** will always be equal. This astonishing fact may have been known to Euler in 1734.

Theorem 21.6.1 Suppose $f : U \subseteq \mathbb{R}^2 \rightarrow \mathbb{R}$ where U is an open set on which f_x, f_y, f_{xy} and f_{yx} exist. Then if f_{xy} and f_{yx} are continuous at the point $(x, y) \in U$, it follows

$$f_{xy}(x, y) = f_{yx}(x, y).$$

Proof: Since U is open, there exists $r > 0$ such that $B((x, y), r) \subseteq U$. Now let $|t|, |s| < r/2$ and consider

$$\Delta(s, t) \equiv \frac{1}{st} \{ \overbrace{f(x+t, y+s) - f(x+t, y)}^{h(t)} - \overbrace{(f(x, y+s) - f(x, y))}^{h(0)} \}. \quad (21.6)$$

Note that $(x+t, y+s) \in U$ because

$$\begin{aligned} |(x+t, y+s) - (x, y)| &= |(t, s)| = (t^2 + s^2)^{1/2} \\ &\leq \left(\frac{r^2}{4} + \frac{r^2}{4} \right)^{1/2} = \frac{r}{\sqrt{2}} < r. \end{aligned}$$

As implied above, $h(t) \equiv f(x+t, y+s) - f(x, y)$. Therefore, by the mean value theorem from calculus and the (one variable) chain rule,

$$\begin{aligned} \Delta(s, t) &= \frac{1}{st} (h(t) - h(0)) = \frac{1}{st} h'(\alpha t) t \\ &= \frac{1}{s} (f_x(x + \alpha t, y+s) - f_x(x, y)) \end{aligned}$$

for some $\alpha \in (0, 1)$. Applying the mean value theorem again,

$$\Delta(s, t) = f_{xy}(x + \alpha t, y + \beta s)$$

where $\alpha, \beta \in (0, 1)$.

If the terms $f(x + t, y)$ and $f(x, y + s)$ are interchanged in 21.6, $\Delta(s, t)$ is also unchanged and the above argument shows there exist $\gamma, \delta \in (0, 1)$ such that

$$\Delta(s, t) = f_{yx}(x + \gamma t, y + \delta s).$$

Letting $(s, t) \rightarrow (0, 0)$ and using the continuity of f_{xy} and f_{yx} at (x, y) ,

$$\lim_{(s,t) \rightarrow (0,0)} \Delta(s, t) = f_{xy}(x, y) = f_{yx}(x, y).$$

This proves the theorem.

The following is obtained from the above by simply fixing all the variables except for the two of interest.

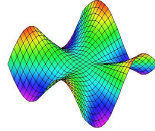
Corollary 21.6.2 Suppose U is an open subset of $\mathbb{R}^n$ and $f : U \rightarrow \mathbb{R}$ has the property that for two indices, k, l , f_{x_k} , f_{x_l} , $f_{x_l x_k}$, and $f_{x_k x_l}$ exist on U and $f_{x_k x_l}$ and $f_{x_l x_k}$ are both continuous at $\mathbf{x} \in U$. Then $f_{x_k x_l}(\mathbf{x}) = f_{x_l x_k}(\mathbf{x})$.

It is necessary to assume the mixed partial derivatives are continuous in order to assert they are equal. The following is a well known example [3].

Example 21.6.3 Let

$$f(x, y) = \begin{cases} \frac{xy(x^2 - y^2)}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

Here is a picture of the graph of this function. It looks innocuous but isn't.



From the definition of partial derivatives it follows immediately that $f_x(0, 0) = f_y(0, 0) = 0$. Using the standard rules of differentiation, for $(x, y) \neq (0, 0)$,

$$f_x = y \frac{x^4 - y^4 + 4x^2 y^2}{(x^2 + y^2)^2}, \quad f_y = x \frac{x^4 - y^4 - 4x^2 y^2}{(x^2 + y^2)^2}$$

Now

$$f_{xy}(0, 0) \equiv \lim_{y \rightarrow 0} \frac{f_x(0, y) - f_x(0, 0)}{y} = \lim_{y \rightarrow 0} \frac{-y^4}{(y^2)^2} = -1$$

while

$$f_{yx}(0, 0) \equiv \lim_{x \rightarrow 0} \frac{f_y(x, 0) - f_y(0, 0)}{x} = \lim_{x \rightarrow 0} \frac{x^4}{(x^2)^2} = 1$$

showing that although the mixed partial derivatives do exist at $(0, 0)$, they are not equal there.

21.7 Partial Differential Equations

Partial differential equations are equations which involve the partial derivatives of some function. The most famous partial differential equations involve the **Laplacian**, named after Laplace³.

Definition 21.7.1 Let u be a function of n variables. Then $\Delta u \equiv \sum_{k=1}^n u_{x_k x_k}$. This is also written as $\nabla^2 u$. The symbol, Δ or ∇^2 is called the Laplacian. When $\Delta u = 0$ the function, u is called **harmonic**. **Laplace's equation** is $\Delta u = 0$. The **heat equation** is $u_t - \Delta u = 0$ and the **wave equation** is $u_{tt} - \Delta u = 0$.

Example 21.7.2 Find the Laplacian of $u(x, y) = x^2 - y^2$.

$u_{xx} = 2$ while $u_{yy} = -2$. Therefore, $\Delta u = u_{xx} + u_{yy} = 2 - 2 = 0$. Thus this function is harmonic, $\Delta u = 0$.

Example 21.7.3 Find $u_t - \Delta u$ where $u(t, x, y) = e^{-t} \cos x$.

In this case, $u_t = -e^{-t} \cos x$ while $u_{yy} = 0$ and $u_{xx} = -e^{-t} \cos x$ therefore, $u_t - \Delta u = 0$ and so u solves the heat equation, $u_t - \Delta u = 0$.

Example 21.7.4 Let $u(t, x) = \sin t \cos x$. Find $u_{tt} - \Delta u$.

In this case, $u_{tt} = -\sin t \cos x$ while $\Delta u = -\sin t \cos x$. Therefore, u is a solution of the wave equation, $u_{tt} - \Delta u = 0$.

21.8 Exercises

- Find $f_x, f_y, f_z, f_{xy}, f_{yx}, f_{xz}, f_{zx}, f_{zy}, f_{yz}$ for the following. Verify the mixed partial derivatives are equal.
 - $x^2 y^3 z^4 + \sin(xyz)$
 - $\sin(xyz) + x^2 yz$
 - $z \ln |x^2 + y^2 + 1|$
 - $e^{x^2 + y^2 + z^2}$
 - $\tan(xyz)$
- Suppose f is a continuous function and $f : U \rightarrow \mathbb{R}$ where U is an open set and suppose that $\mathbf{x} \in U$ has the property that for all $\mathbf{y}$ near $\mathbf{x}$, $f(\mathbf{x}) \leq f(\mathbf{y})$. Prove that if f has all of its partial derivatives at $\mathbf{x}$, then $f_{x_i}(\mathbf{x}) = 0$ for each x_i . **Hint:** This is just a repeat of the usual one variable theorem seen in beginning calculus. You just do this one variable argument for each variable to get the conclusion.
- As an important application of Problem 2 consider the following. Experiments are done at n times, $t_1, t_2, \dots, t_n$ and at each time there results a collection of numerical outcomes. Denote by $\{(t_i, x_i)\}_{i=1}^n$ the set of all such pairs and try to find numbers a and b such that the line $x = at + b$ approximates these ordered pairs as well as possible in the sense that out of all choices of a and b , $\sum_{i=1}^n (at_i + b - x_i)^2$ is as small as possible. In other words, you want to minimize the function of two variables, $f(a, b) \equiv \sum_{i=1}^n (at_i + b - x_i)^2$. Find a formula for a and b in terms of the given ordered pairs. You will be finding the formula for the least squares regression line.

³Laplace was a great physicist of the 1700's. He made fundamental contributions to mechanics and astronomy.

4. Show that if $v(x, y) = u(\alpha x, \beta y)$, then $v_x = \alpha u_x$ and $v_y = \beta u_y$. State and prove a generalization to any number of variables.
5. Let f be a function which has continuous derivatives. Show that $u(t, x) = f(x - ct)$ solves the wave equation, $u_{tt} - c^2 \Delta u = 0$. What about $u(x, t) = f(x + ct)$?
6. D'Alembert found a formula for the solution to the wave equation, $u_{tt} = c^2 u_{xx}$ along with the initial conditions $u(x, 0) = f(x)$, $u_t(x, 0) = g(x)$. Here is how he did it. He looked for a solution of the form $u(x, t) = h(x + ct) + k(x - ct)$ and then found h and k in terms of the given functions f and g . He ended up with something like

$$u(x, t) = \frac{1}{2c} \int_{x-ct}^{x+ct} g(r) dr + \frac{1}{2} (f(x + ct) + f(x - ct)).$$

Fill in the details.

7. Determine which of the following functions satisfy Laplace's equation.
 - (a) $x^3 - 3xy^2$
 - (b) $3x^2y - y^3$
 - (c) $x^3 - 3xy^2 + 2x^2 - 2y^2$
 - (d) $3x^2y - y^3 + 4xy$
 - (e) $3x^2 - y^3 + 4xy$
 - (f) $3x^2y - y^3 + 4y$
 - (g) $x^3 - 3x^2y^2 + 2x^2 - 2y^2$
8. Show that $z = \sqrt{x^2 + y^2}$ is a solution to $x \frac{\partial z}{\partial x} + y \frac{\partial z}{\partial y} = 0$.
9. Show that if $\Delta u = \lambda u$ where u is a function of only x , then $e^{\lambda t} u$ solves the heat equation, $u_t - \Delta u = 0$.
10. Show that if a, b are scalars and u, v are functions which satisfy Laplace's equation then $au + bv$ also satisfies Laplace's equation. Verify a similar statement for the heat and wave equations.
11. Show that $u(x, t) = \frac{1}{\sqrt{t}} e^{-x^2/4c^2t}$ solves the heat equation, $u_t = c^2 u_{xx}$.

The Derivative Of A Function Of Many Variables

22.0.1 Outcomes

1. Define differentiability and explain what the derivative is for a function of n variables.
2. Describe the relation between existence of partial derivatives, continuity, and differentiability.
3. Give examples of functions which have partial derivatives but are not continuous, examples of functions which are differentiable but not C^1 , and examples of functions which are continuous without having partial derivatives.
4. Evaluate derivatives of composite functions using the chain rule.
5. Solve related rates problems using the chain rule.
6. Interpret the gradient of a function as a normal to a level curve or a level surface.
7. Find the normal line and tangent plane to a smooth surface at a given point.
8. Find the angles between curves and surfaces.

22.1 The Derivative Of Functions Of One Variable, o (v)

First consider the notion of the derivative of a function of one variable.

Observation 22.1.1 Suppose a function, f of one variable has a derivative at x . Then

$$\lim_{h \rightarrow 0} \frac{|f(x+h) - f(x) - f'(x)h|}{|h|} = 0.$$

This observation follows from the definition of the derivative of a function of one variable, namely

$$f'(x) \equiv \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Thus

$$\lim_{h \rightarrow 0} \frac{|f(x+h) - f(x) - f'(x)h|}{|h|} = \lim_{h \rightarrow 0} \left| \frac{f(x+h) - f(x)}{h} - f'(x) \right| = 0$$

Definition 22.1.2 A vector valued function of a vector, $\mathbf{v}$ is called $\mathbf{o}(\mathbf{v})$ if

$$\lim_{|\mathbf{v}| \rightarrow 0} \frac{\mathbf{o}(\mathbf{v})}{|\mathbf{v}|} = \mathbf{0}. \quad (22.1)$$

Thus for a function of one variable, the function $f(x+h) - f(x) - f'(x)h$ is $o(h)$. When we say a function is $o(h)$, it is used like an adjective. It is like saying the function is white or black or green or fat or thin. The term is used very imprecisely. Thus in general,

$$\mathbf{o}(\mathbf{v}) = \mathbf{o}(\mathbf{v}) + \mathbf{o}(\mathbf{v}), \mathbf{o}(\mathbf{v}) = 45\mathbf{o}(\mathbf{v}), \mathbf{o}(\mathbf{v}) = \mathbf{o}(\mathbf{v}) - \mathbf{o}(\mathbf{v}), \text{ etc.}$$

When you add two functions with the property of the above definition, you get another one having that same property. When you multiply by 45 the property is also retained as it is when you subtract two such functions. How could something so sloppy be useful? The notation is useful precisely because it prevents you from obsessing over things which are not relevant and should be ignored.

Theorem 22.1.3 Let $f : (a, b) \rightarrow \mathbb{R}$ be a function of one variable. Then $f'(x)$ exists if and only if

$$f(x+h) - f(x) = ph + o(h) \quad (22.2)$$

In this case, $p = f'(x)$.

Proof: From the above observation it follows that if $f'(x)$ does exist, then 22.2 holds.

Suppose then that 22.2 is true. Then

$$\frac{f(x+h) - f(x)}{h} - p = \frac{o(h)}{h}.$$

Taking a limit, you see that

$$p = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

and that in fact this limit exists which shows that $p = f'(x)$. This proves the theorem.

This theorem shows that one way to define $f'(x)$ is as the number p , if there is one, which has the property that

$$f(x+h) = f(x) + ph + o(h).$$

You should think of p as the linear transformation resulting from multiplication by the 1×1 matrix, (p) .

Example 22.1.4 Let $f(x) = x^3$. Find $f'(x)$.

$f(x+h) = (x+h)^3 = x^3 + 3x^2h + 3xh^2 + h^3 = f(x) + 3x^2h + (3xh + h^2)h$. Since $(3xh + h^2)h = o(h)$, it follows $f'(x) = 3x^2$.

Example 22.1.5 Let $f(x) = \sin(x)$. Find $f'(x)$.

$$\begin{aligned} f(x+h) - f(x) &= \sin(x+h) - \sin(x) = \sin(x)\cos(h) + \cos(x)\sin(h) - \sin(x) \\ &= \cos(x)\sin(h) + \sin(x)\frac{(\cos(h)-1)}{h}h \\ &= \cos(x)h + \cos(x)\frac{(\sin(h)-h)}{h}h + \sin(x)\frac{(\cos(h)-1)}{h}h. \end{aligned}$$

Now

$$\cos(x) \frac{(\sin(h) - h)}{h} h + \sin(x) \frac{(\cos(h) - 1)}{h} h = o(h). \quad (22.3)$$

Remember the fundamental limits which allowed you to find the derivative of $\sin(x)$ were

$$\lim_{h \rightarrow 0} \frac{\sin(h)}{h} = 1, \quad \lim_{h \rightarrow 0} \frac{\cos(h) - 1}{h} = 0. \quad (22.4)$$

These same limits are what is needed to verify 22.3.

22.2 Exercises

1. Determine which of the following functions are $o(h)$.
 - (a) h^2
 - (b) $h \sin(h)$
 - (c) $|h|^{3/2} \ln(|h|)$
 - (d) $h^2 x + y h^3$
 - (e) $\sin(h^2)$
 - (f) $\sin(h)$
 - (g) $x h \sin(\sqrt{|h|}) + x^5 h^2$
 - (h) $\exp(-1/|h|^2)$
2. Here are some scalar valued functions of several variables. Determine which of these functions are $o(\mathbf{v})$. Here $\mathbf{v}$ is a vector in $\mathbb{R}^n$, $\mathbf{v} = (v_1, \dots, v_n)$.
 - (a) $v_1 v_2$
 - (b) $v_2 \sin(v_1)$
 - (c) $v_1^2 + v_2$
 - (d) $v_2 \sin(v_1 + v_2)$
 - (e) $v_1(v_1 + v_2 + x v_3)$
3. Here are some vector valued functions of $\mathbf{v} \in \mathbb{R}^n$. Determine which ones are $o(\mathbf{v})$.
 - (a) $(e^{v_1} - 1 - v_1)$
 - (g) $(\mathbf{x} \cdot \mathbf{v}) |\mathbf{v}|$
 - (a) $(\mathbf{x} \cdot \mathbf{v}) \mathbf{v}$
 - (b) $\sin(v_1) \mathbf{v}$
 - (c) $\sqrt{|\mathbf{x} \cdot \mathbf{v}|} |\mathbf{v}|^{2/3}$
 - (d) $\sqrt{|\mathbf{x} \cdot \mathbf{v}|} |\mathbf{v}|^{1/2}$
 - (e) $\left(\sin\left(\sqrt{|\mathbf{x} \cdot \mathbf{v}|}\right) - \sqrt{|\mathbf{x} \cdot \mathbf{v}|} \right) |\mathbf{v}|^{-1/4}$
 - (f) $\exp(-1/|\mathbf{v}|^2)$
 - (g) $\mathbf{v}^T A \mathbf{v}$ where A is an $n \times n$ matrix.
4. Show that if $f(x) = o(x)$, then $f'(0) = 0$.
5. Show that if $\lim_{h \rightarrow 0} f(x) = 0$ then $x f(x) = o(x)$.
6. Show that if $f'(0)$ exists and $f(0) = 0$, then $f(|x|^p) = o(x)$ whenever $p > 1$.

22.3 The Derivative Of Functions Of Many Variables

This way of thinking about the derivative is exactly what is needed to define the derivative of a function of n variables. Recall the following definition.

Definition 22.3.1 A function, T which maps $\mathbb{R}^n$ to $\mathbb{R}^p$ is called a linear transformation if for every pair of scalars, a, b and vectors, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, it follows that $T(a\mathbf{x} + b\mathbf{y}) = aT(\mathbf{x}) + bT(\mathbf{y})$.

Recall that from the properties of matrix multiplication, it follows that if A is an $n \times p$ matrix, and if $\mathbf{x}, \mathbf{y}$ are vectors in $\mathbb{R}^n$, then $A(a\mathbf{x} + b\mathbf{y}) = aA(\mathbf{x}) + bA(\mathbf{y})$. Thus you can define a linear transformation by multiplying by a matrix. Of course the simplest example is that of a 1×1 matrix or number. You can think of the number 3 as a

linear transformation, T mapping $\mathbb{R}$ to $\mathbb{R}$ according to the rule $Tx = 3x$. It satisfies the properties needed for a linear transformation because $3(ax + by) = a3x + b3y = aTx + bTy$. The case of the derivative of a scalar valued function of one variable is of this sort. You get a number for the derivative. However, you can think of this number as a linear transformation. Of course it might not be worth the fuss to think of it this way for a function of one variable but this is the way you must think of it for a function of n variables.

Definition 22.3.2 Let $\mathbf{f} : U \rightarrow \mathbb{R}^p$ where U is an open set in $\mathbb{R}^n$ for $n, p \geq 1$ and let $\mathbf{x} \in U$ be given. Then $\mathbf{f}$ is defined to be **differentiable** at $\mathbf{x} \in U$ if and only if there exist column vectors, $\mathbf{v}_i$ such that for $\mathbf{h} = (h_1 \cdots, h_n)^T$,

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) = \mathbf{f}(\mathbf{x}) + \sum_{i=1}^n \mathbf{v}_i h_i + \mathbf{o}(\mathbf{h}). \quad (22.5)$$

The derivative of the function, $\mathbf{f}$, denoted by $D\mathbf{f}(\mathbf{x})$, is the linear transformation defined by multiplying by the matrix whose columns are the $p \times 1$ vectors, $\mathbf{v}_i$. Thus if $\mathbf{h}$ is a vector in $\mathbb{R}^n$,

$$D\mathbf{f}(\mathbf{x})\mathbf{h} \equiv \begin{pmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{pmatrix} \mathbf{h}$$

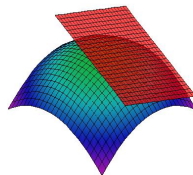
and one can write $\mathbf{f}$ is differentiable if and only if there is a matrix $D\mathbf{f}(\mathbf{x})$ such that

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) = \mathbf{f}(\mathbf{x}) + D\mathbf{f}(\mathbf{x})\mathbf{h} + \mathbf{o}(\mathbf{h}).$$

If $\mathbf{h} = \mathbf{x} - \mathbf{x}_0$, this takes the form

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0) + D\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \mathbf{o}(\mathbf{x} - \mathbf{x}_0)$$

If you deleted the $\mathbf{o}(\mathbf{x} - \mathbf{x}_0)$ term and considered the function of $\mathbf{x}$ given by what is left, this is called the linear approximation to the function at the point $\mathbf{x}_0$. In the case where $\mathbf{x} \in \mathbb{R}^2$ and f has values in $\mathbb{R}$ one can draw a picture to illustrate this.



The linear approximation is graphed as that plane which is tangent to the graph of the function. This is what is going on geometrically.

It is common to think of this matrix as the derivative but strictly speaking, this is incorrect. The derivative is a “linear transformation” determined by multiplication by this matrix, called the **standard matrix** because it is based on the standard basis vectors for $\mathbb{R}^n$. The subtle issues involved in a thorough exploration of this issue will be avoided for now. It will be fine to think of the above matrix as the derivative. Other notations which are often used for this matrix or the linear transformation are $\mathbf{f}'(\mathbf{x})$, $J(\mathbf{x})$, and even $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ or $\frac{d\mathbf{f}}{d\mathbf{x}}$.

The next theorem gives a description of this matrix and shows it is unique if it exists.

Theorem 22.3.3 Suppose $\mathbf{f}$ is differentiable, as given above in 22.5. Then

$$\mathbf{v}_k = \lim_{h \rightarrow 0} \frac{\mathbf{f}(\mathbf{x} + h\mathbf{e}_k) - \mathbf{f}(\mathbf{x})}{h} \equiv \frac{\partial \mathbf{f}}{\partial x_k}(\mathbf{x}),$$

the k^{th} partial derivative.

Proof: Let $\mathbf{h} = (0, \dots, h, 0, \dots, 0)^T = h\mathbf{e}_k$ where the h is in the k^{th} slot. Then 22.5 reduces to

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) = \mathbf{f}(\mathbf{x}) + \mathbf{v}_k h + \mathbf{o}(h).$$

Therefore, dividing by h

$$\frac{\mathbf{f}(\mathbf{x} + h\mathbf{e}_k) - \mathbf{f}(\mathbf{x})}{h} = \mathbf{v}_k + \frac{\mathbf{o}(h)}{h}$$

and taking the limit,

$$\lim_{h \rightarrow 0} \frac{\mathbf{f}(\mathbf{x} + h\mathbf{e}_k) - \mathbf{f}(\mathbf{x})}{h} = \lim_{h \rightarrow 0} \left(\mathbf{v}_k + \frac{\mathbf{o}(h)}{h} \right) = \mathbf{v}_k$$

and so, the above limit exists. This proves the theorem.

The above description of the derivative can be used to approximate a nonlinear function. Since the $\mathbf{o}(\mathbf{x} - \mathbf{x}_0)$ term is so small, it is a good approximation to write

$$\begin{aligned} \mathbf{f}(\mathbf{x}) &\approx \mathbf{f}(\mathbf{x}_0) + D\mathbf{f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \\ &= \mathbf{f}(\mathbf{x}_0) + \left(\frac{\partial \mathbf{f}}{\partial x_1}(\mathbf{x}_0) \quad \frac{\partial \mathbf{f}}{\partial x_2}(\mathbf{x}_0) \quad \cdots \quad \frac{\partial \mathbf{f}}{\partial x_n}(\mathbf{x}_0) \right) \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix} \\ &= \mathbf{f}(\mathbf{x}_0) + \sum_{i=1}^n \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0) h_i \end{aligned}$$

Here is an example of a scalar valued nonlinear function.

Example 22.3.4 Suppose $f(x, y) = \sqrt{xy}$. Find the approximate change in f if x goes from 1 to 1.01 and y goes from 4 to 3.99.

We can do this by noting that

$$\begin{aligned} f(1.01, 3.99) - f(1, 4) &\approx f_x(1, 2)(.01) + f_y(1, 2)(-.01) \\ &= 1(.01) + \frac{1}{4}(-.01) = 7.5 \times 10^{-3}. \end{aligned}$$

Of course the exact value is

$$\sqrt{(1.01)(3.99)} - \sqrt{4} = 7.4610831 \times 10^{-3}.$$

Notation 22.3.5 When f is a scalar valued function of n variables, the following is often written to express the idea that a small change in f due to small changes in the variables can be expressed in the form

$$df(\mathbf{x}) = f_{x_1}(\mathbf{x})dx_1 + \cdots + f_{x_n}(\mathbf{x})dx_n$$

where the small change in x_i is denoted as dx_i . As explained above, df is the approximate change in the function f . Sometimes df is referred to as the differential of f .

Let $\mathbf{f} : U \rightarrow \mathbb{R}^q$ where U is an open subset of $\mathbb{R}^p$ and $\mathbf{f}$ is differentiable. It was just shown

$$\mathbf{f}(\mathbf{x} + \mathbf{v}) = \mathbf{f}(\mathbf{x}) + \sum_{j=1}^p \frac{\partial \mathbf{f}(\mathbf{x})}{\partial x_j} v_j + \mathbf{o}(\mathbf{v}).$$

Taking the i^{th} coordinate of the above equation yields

$$f_i(\mathbf{x} + \mathbf{v}) = f_i(\mathbf{x}) + \sum_{j=1}^p \frac{\partial f_i(\mathbf{x})}{\partial x_j} v_j + o(\mathbf{v}),$$

and it follows that the term with a sum is nothing more than the i^{th} component of $J(\mathbf{x})\mathbf{v}$ where $J(\mathbf{x})$ is the $q \times p$ matrix,

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_p} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_q}{\partial x_1} & \frac{\partial f_q}{\partial x_2} & \cdots & \frac{\partial f_q}{\partial x_p} \end{pmatrix}.$$

Thus

$$\mathbf{f}(\mathbf{x} + \mathbf{v}) = \mathbf{f}(\mathbf{x}) + J(\mathbf{x})\mathbf{v} + \mathbf{o}(\mathbf{v}), \quad (22.6)$$

and to reiterate, the linear transformation which results by multiplication by this $q \times p$ matrix is known as the derivative.

Sometimes x, y, z is written instead of x_1, x_2 , and x_3 . This is to save on notation and is easier to write and to look at although it lacks generality. When this is done it is understood that $x = x_1, y = x_2$, and $z = x_3$. Thus the derivative is the linear transformation determined by

$$\begin{pmatrix} f_{1x} & f_{1y} & f_{1z} \\ f_{2x} & f_{2y} & f_{2z} \\ f_{3x} & f_{3y} & f_{3z} \end{pmatrix}.$$

Example 22.3.6 Let A be a constant $m \times n$ matrix and consider $\mathbf{f}(\mathbf{x}) = A\mathbf{x}$. Find $D\mathbf{f}(\mathbf{x})$ if it exists.

$$\mathbf{f}(\mathbf{x} + \mathbf{h}) - \mathbf{f}(\mathbf{x}) = A(\mathbf{x} + \mathbf{h}) - A\mathbf{x} = A\mathbf{h} = A\mathbf{h} + \mathbf{o}(\mathbf{h}).$$

In fact in this case, $\mathbf{o}(\mathbf{h}) = \mathbf{0}$. Therefore, $D\mathbf{f}(\mathbf{x}) = A$. Note that this looks the same as the case in one variable, $f(x) = ax$.

Example 22.3.7 Let $f(x, y, z) = xy + z^2x$. Find $Df(x, y, z)$.

Consider $f(x + h, y + k, z + l) - f(x, y, z)$. This is something which is easily computed from the definition of the function. It equals

$$(x + h)(y + k) + (z + l)^2(x + h) - (xy + z^2x)$$

Multiply everything together and collect the terms. This yields

$$(z^2 + y)h + xk + 2zxl + (hk + 2zlh + l^2x + l^2h)$$

It follows easily the last term in at the end is $o(h, k, l)$ and so the derivative of this function is the linear transformation coming from multiplication by the matrix $((z^2 + y), x, 2zx)$ and so this is the derivative. It follows from this and the description of the derivative in terms of partial derivatives that

$$\frac{\partial f}{\partial x}(x, y, z) = z^2 + y, \quad \frac{\partial f}{\partial y}(x, y, z) = x, \quad \frac{\partial f}{\partial z}(x, y, z) = 2xz.$$

Of course you could compute these partial derivatives directly.

Given a function of many variables, how can you tell if it is differentiable? In other words, when you make the linear approximation, how can you tell easily that what is left over is $\mathbf{o}(\mathbf{v})$. Sometimes you have to go directly to the definition and verify it is differentiable from the definition. For example, you may have seen the following important example in one variable calculus.

Example 22.3.8 Let $f(x) = \begin{cases} x^2 \sin(\frac{1}{x}) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$. Find $Df(0)$.

$f(h) - f(0) = 0h + h^2 \sin(\frac{1}{h}) = o(h)$ and so $Df(0) = 0$. If you find the derivative for $x \neq 0$, it is totally useless information if what you want is $Df(0)$. This is because the derivative, turns out to be discontinuous. Try it. Find the derivative for $x \neq 0$ and try to obtain $Df(0)$ from it. You see, in this example you had to revert to the definition to find the derivative.

It isn't really too hard to use the definition even for more ordinary examples.

Example 22.3.9 Let $\mathbf{f}(x, y) = \begin{pmatrix} x^2y + y^2 \\ y^3x \end{pmatrix}$. Find $D\mathbf{f}(1, 2)$.

First of all note that the thing you are after is a 2×2 matrix.

$$\mathbf{f}(1, 2) = \begin{pmatrix} 6 \\ 8 \end{pmatrix}.$$

Then

$$\begin{aligned} & \mathbf{f}(1 + h_1, 2 + h_2) - \mathbf{f}(1, 2) \\ &= \begin{pmatrix} (1 + h_1)^2(2 + h_2) + (2 + h_2)^2 \\ (2 + h_2)^3(1 + h_1) \end{pmatrix} - \begin{pmatrix} 6 \\ 8 \end{pmatrix} \\ &= \begin{pmatrix} 5h_2 + 4h_1 + 2h_1h_2 + 2h_1^2 + h_1^2h_2 + h_2^2 \\ 8h_1 + 12h_2 + 12h_1h_2 + 6h_2^2 + 6h_2^2h_1 + h_2^3 + h_2^3h_1 \end{pmatrix} \\ &= \begin{pmatrix} 4 & 5 \\ 8 & 12 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \begin{pmatrix} 2h_1h_2 + 2h_1^2 + h_1^2h_2 + h_2^2 \\ 12h_1h_2 + 6h_2^2 + 6h_2^2h_1 + h_2^3 + h_2^3h_1 \end{pmatrix} \\ &= \begin{pmatrix} 4 & 5 \\ 8 & 12 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} + \mathbf{o}(\mathbf{h}). \end{aligned}$$

Therefore, the standard matrix of the derivative is $\begin{pmatrix} 4 & 5 \\ 8 & 12 \end{pmatrix}$.

22.4 Exercises

1. Let $f(t) = t^2 + t$. Use the above definition of the derivative in terms of $o(h)$ to compute the derivative.
2. Let $f(x, y) = x \sin(y)$. Compute the derivative directly from the definition.
3. Let $f(x, y) = x^2 \sin(y)$. Compute the derivative directly from the definition.
4. Let $\mathbf{f}(x, y) = \begin{pmatrix} x^2y \\ x + y^2 \end{pmatrix}$. Compute the derivative directly from the definition.
5. Let $f(x, y) = x^\alpha y^\beta$. Show that $Df(x, y) = \begin{pmatrix} \alpha x^{\alpha-1} y^\beta & x^\alpha \beta y^{\beta-1} \end{pmatrix}$.

6. Let $f(x, y) = \sqrt{x} \sqrt[3]{y}$. Find the approximate change in f when (x, y) goes from $(4, 8)$ to $(4.01, 7.99)$.
7. Suppose f is differentiable and $\mathbf{g}$ is also differentiable, $\mathbf{g}$ having values in $\mathbb{R}^3$ and f having values in $\mathbb{R}$. Find $D(f\mathbf{g})$ directly from the definition. Assume both functions are defined on an open subset of $\mathbb{R}^n$.

22.5 C^1 Functions

Most of the time, there is an easier way to conclude a derivative exists and to find it. It involves the notion of a C^1 function.

Definition 22.5.1 When $\mathbf{f} : U \rightarrow \mathbb{R}^p$ for U an open subset of $\mathbb{R}^n$ and the vector valued functions, $\frac{\partial \mathbf{f}}{\partial x_i}$ are all continuous, (equivalently each $\frac{\partial f_i}{\partial x_j}$ is continuous), the function is said to be $C^1(U)$. If all the partial derivatives up to order k exist and are continuous, then the function is said to be C^k .

It turns out that for a C^1 function, all you have to do is write the matrix described in Theorem 22.3.3 and this will be the derivative. There is no question of existence for the derivative for such functions. This is the importance of the next theorem.

Theorem 22.5.2 Suppose $\mathbf{f} : U \rightarrow \mathbb{R}^p$ where U is an open set in $\mathbb{R}^n$. Suppose also that all partial derivatives of $\mathbf{f}$ exist on U and are continuous. Then $\mathbf{f}$ is differentiable at every point of U .

Proof: If you fix all the variables but one, you can apply the fundamental theorem of calculus as follows.

$$\mathbf{f}(\mathbf{x} + v_k \mathbf{e}_k) - \mathbf{f}(\mathbf{x}) = \int_0^1 \frac{\partial \mathbf{f}}{\partial x_k}(\mathbf{x} + tv_k \mathbf{e}_k) v_k dt. \quad (22.7)$$

Here is why. Let $\mathbf{h}(t) = \mathbf{f}(\mathbf{x} + tv_k \mathbf{e}_k)$. Then

$$\frac{\mathbf{h}(t+h) - \mathbf{h}(t)}{h} = \frac{\mathbf{f}(\mathbf{x} + tv_k \mathbf{e}_k + hv_k \mathbf{e}_k) - \mathbf{f}(\mathbf{x} + tv_k \mathbf{e}_k)}{hv_k} v_k$$

and so, taking the limit as $h \rightarrow 0$ yields

$$\mathbf{h}'(t) = \frac{\partial \mathbf{f}}{\partial x_k}(\mathbf{x} + tv_k \mathbf{e}_k) v_k$$

Therefore,

$$\mathbf{f}(\mathbf{x} + v_k \mathbf{e}_k) - \mathbf{f}(\mathbf{x}) = \mathbf{h}(1) - \mathbf{h}(0) = \int_0^1 \mathbf{h}'(t) dt = \int_0^1 \frac{\partial \mathbf{f}}{\partial x_k}(\mathbf{x} + tv_k \mathbf{e}_k) v_k dt.$$

Now I will use this observation to prove the theorem. Let $\mathbf{v} = (v_1, \dots, v_n)$ with $|\mathbf{v}|$ sufficiently small. Thus $\mathbf{v} = \sum_{k=1}^n v_k \mathbf{e}_k$. For the purposes of this argument, define

$$\sum_{k=n+1}^n v_k \mathbf{e}_k \equiv \mathbf{0}.$$

Then with this convention,

$$\begin{aligned} \mathbf{f}(\mathbf{x} + \mathbf{v}) - \mathbf{f}(\mathbf{x}) &= \sum_{i=1}^n \left(\mathbf{f} \left(\mathbf{x} + \sum_{k=i}^n v_k \mathbf{e}_k \right) - \mathbf{f} \left(\mathbf{x} + \sum_{k=i+1}^n v_k \mathbf{e}_k \right) \right) \\ &= \sum_{i=1}^n \int_0^1 \frac{\partial \mathbf{f}}{\partial x_i} \left(\mathbf{x} + \sum_{k=i+1}^n v_k \mathbf{e}_k + tv_i \mathbf{e}_i \right) v_i dt \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \int_0^1 \left(\frac{\partial \mathbf{f}}{\partial x_i} \left(\mathbf{x} + \sum_{k=i+1}^n v_k \mathbf{e}_k + tv_i \mathbf{e}_i \right) v_i - \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}) v_i \right) dt \\
&\quad + \sum_{i=1}^n \int_0^1 \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}) v_i dt = \sum_{i=1}^n \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}) v_i \\
&\quad + \int_0^1 \sum_{i=1}^n \left(\frac{\partial \mathbf{f}}{\partial x_i} \left(\mathbf{x} + \sum_{k=i+1}^n v_k \mathbf{e}_k + tv_i \mathbf{e}_i \right) - \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}) \right) v_i dt \\
&= \sum_{i=1}^n \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}) v_i + o(\mathbf{v})
\end{aligned}$$

and this shows $\mathbf{f}$ is differentiable at $\mathbf{x}$.

Some explanation of the step to the last line is in order. The messy thing at the end is $o(\mathbf{v})$ because of the continuity of the partial derivatives. In fact, from the Cauchy Schwarz inequality,

$$\begin{aligned}
&\int_0^1 \sum_{i=1}^n \left(\frac{\partial \mathbf{f}}{\partial x_i} \left(\mathbf{x} + \sum_{k=i+1}^n v_k \mathbf{e}_k + tv_i \mathbf{e}_i \right) - \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}) \right) v_i dt \\
&\leq \int_0^1 \left(\sum_{i=1}^n \left| \frac{\partial \mathbf{f}}{\partial x_i} \left(\mathbf{x} + \sum_{k=i+1}^n v_k \mathbf{e}_k + tv_i \mathbf{e}_i \right) - \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}) \right|^2 \right)^{1/2} dt \left(\sum_{i=1}^n v_i^2 \right)^{1/2} \\
&= \int_0^1 \left(\sum_{i=1}^n \left| \frac{\partial \mathbf{f}}{\partial x_i} \left(\mathbf{x} + \sum_{k=i+1}^n v_k \mathbf{e}_k + tv_i \mathbf{e}_i \right) - \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}) \right|^2 \right)^{1/2} dt |\mathbf{v}|
\end{aligned}$$

Thus, dividing by $|\mathbf{v}|$ and taking a limit as $|\mathbf{v}| \rightarrow 0$, the quotient is nothing but

$$\int_0^1 \left(\sum_{i=1}^n \left| \frac{\partial \mathbf{f}}{\partial x_i} \left(\mathbf{x} + \sum_{k=i+1}^n v_k \mathbf{e}_k + tv_i \mathbf{e}_i \right) - \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}) \right|^2 \right)^{1/2} dt$$

which converges to 0 due to continuity of the partial derivatives of $\mathbf{f}$. This proves the theorem.

Here is an example to illustrate.

Example 22.5.3 Let $\mathbf{f}(x, y) = \begin{pmatrix} x^2 y + y^2 \\ y^3 x \end{pmatrix}$. Find $D\mathbf{f}(x, y)$.

From Theorem 22.5.2 this function is differentiable because all possible partial derivatives are continuous. Thus

$$D\mathbf{f}(x, y) = \begin{pmatrix} 2xy & x^2 + 2y \\ y^3 & 3y^2 x \end{pmatrix}.$$

In particular,

$$D\mathbf{f}(1, 2) = \begin{pmatrix} 4 & 5 \\ 8 & 12 \end{pmatrix}.$$

Here is another example.

Example 22.5.4 Let $\mathbf{f}(x_1, x_2, x_3) = \begin{pmatrix} x_1^2 x_2 + x_2^2 \\ x_2 x_1 + x_3 \\ \sin(x_1 x_2 x_3) \end{pmatrix}$. Find $D\mathbf{f}(x_1, x_2, x_3)$.

All possible partial derivatives are continuous so the function is differentiable. The matrix for this derivative is therefore the following 3×3 matrix

$$\begin{pmatrix} 2x_1x_2 & x_1^2 + 2x_2 & 0 \\ x_2 & x_1 & 1 \\ x_2x_3 \cos(x_1x_2x_3) & x_1x_3 \cos(x_1x_2x_3) & x_1x_2 \cos(x_1x_2x_3) \end{pmatrix}$$

Example 22.5.5 Suppose $f(x, y, z) = xy + z^2$. Find $Df(1, 2, 3)$.

Taking the partial derivatives of f , $f_x = y$, $f_y = x$, $f_z = 2z$. These are all continuous. Therefore, the function has a derivative and $f_x(1, 2, 3) = 1$, $f_y(1, 2, 3) = 2$, and $f_z(1, 2, 3) = 6$. Therefore, $Df(1, 2, 3)$ is given by

$$Df(1, 2, 3) = (1, 2, 6).$$

Also, for (x, y, z) close to $(1, 2, 3)$,

$$\begin{aligned} f(x, y, z) &\approx f(1, 2, 3) + 1(x - 1) + 2(y - 2) + 6(z - 3) \\ &= 11 + 1(x - 1) + 2(y - 2) + 6(z - 3) = -12 + x + 2y + 6z \end{aligned}$$

When a function is differentiable at $\mathbf{x}_0$ it follows the function must be continuous there. This is the content of the following important lemma.

Lemma 22.5.6 Let $\mathbf{f} : U \rightarrow \mathbb{R}^q$ where U is an open subset of $\mathbb{R}^p$. If $\mathbf{f}$ is differentiable, then $\mathbf{f}$ is continuous at $\mathbf{x}_0$. Furthermore, if $C \geq \max \left\{ \left| \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0) \right|, i = 1, \dots, p \right\}$, then whenever $|\mathbf{x} - \mathbf{x}_0|$ is small enough,

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)| \leq (Cp + 1)|\mathbf{x} - \mathbf{x}_0| \quad (22.8)$$

Proof: Suppose $\mathbf{f}$ is differentiable. Since $\mathbf{o}(\mathbf{v})$ satisfies 22.1, there exists $\delta_1 > 0$ such that if $|\mathbf{x} - \mathbf{x}_0| < \delta_1$, then $|\mathbf{o}(\mathbf{x} - \mathbf{x}_0)| < |\mathbf{x} - \mathbf{x}_0|$. But also, by the triangle inequality, Corollary 11.5.5 on Page 302,

$$\left| \sum_{i=1}^p \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0)(x_i - x_{0i}) \right| \leq C \sum_{i=1}^p |x_i - x_{0i}| \leq Cp|\mathbf{x} - \mathbf{x}_0|$$

Therefore, if $|\mathbf{x} - \mathbf{x}_0| < \delta_1$,

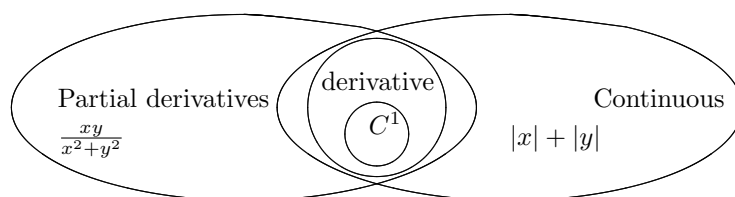
$$\begin{aligned} |\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)| &\leq \left| \sum_{i=1}^p \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{x}_0)(x_i - x_{0i}) \right| + |\mathbf{x} - \mathbf{x}_0| \\ &< (Cp + 1)|\mathbf{x} - \mathbf{x}_0| \end{aligned}$$

which verifies 22.8. Now letting $\varepsilon > 0$ be given, let $\delta = \min \left(\delta_1, \frac{\varepsilon}{Cp+1} \right)$. Then for $|\mathbf{x} - \mathbf{x}_0| < \delta$,

$$|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)| < (Cp + 1)|\mathbf{x} - \mathbf{x}_0| < (Cp + 1) \frac{\varepsilon}{Cp + 1} = \varepsilon$$

showing $\mathbf{f}$ is continuous at $\mathbf{x}_0$.

There have been quite a few terms defined. First there was the concept of continuity. Next the concept of partial or directional derivative. Next there was the concept of differentiability and the derivative being a linear transformation determined by a certain matrix. Finally, it was shown that if a function is C^1 , then it has a derivative. To give a rough idea of the relationships of these topics, here is a picture.



You might ask whether there are examples of functions which are differentiable but not C^1 . Of course there are. In fact, Example 22.3.8 is just such an example as explained earlier. Then you should verify that $f'(x)$ exists for all $x \in \mathbb{R}$ but f' fails to be continuous at $x = 0$. Thus the function is differentiable at every point of $\mathbb{R}$ but fails to be C^1 because the derivative is not continuous at 0.

Example 22.5.7 Find an example of a function which is not differentiable at $(0, 0)$ even though both partial derivatives exist at this point and the function is continuous at this point.

Here is a simple example.

$$f(x, y) \equiv \begin{cases} x \sin\left(\frac{1}{xy}\right) & \text{if } xy \neq 0 \\ 0 & \text{if } xy = 0 \end{cases}$$

To see this works, note that f is defined everywhere and

$$|f(x, y)| \leq |x|$$

so clearly f is continuous at $(0, 0)$.

$$\frac{f(x, 0) - f(0, 0)}{x} = \frac{0 - 0}{x} = 0$$

and so $f_x(0, 0) = 0$. Similarly,

$$\frac{f(0, y) - f(0, 0)}{y} = \frac{0 - 0}{y} = 0$$

and so $f_y(0, 0) = 0$. Thus the partial derivatives exist. However, the function is not differentiable at $(0, 0)$ because

$$\lim_{(x, y) \rightarrow (0, 0)} \frac{x \sin\left(\frac{1}{xy}\right)}{|(x, y)|}$$

does not even exist, much less equals 0. To see this, let $x = y$ and let $x \rightarrow 0$.

22.6 The Chain Rule

22.6.1 The Chain Rule For Functions Of One Variable

First recall the chain rule for a function of one variable. Consider the following picture.

$$I \xrightarrow{g} J \xrightarrow{f} \mathbb{R}$$

Here I and J are open intervals and it is assumed that $g(I) \subseteq J$. The chain rule says that if $f'(g(x))$ exists and $g'(x)$ exists for $x \in I$, then the composition, $f \circ g$ also has a derivative at x and

$$(f \circ g)'(x) = f'(g(x))g'(x).$$

Recall that $f \circ g$ is the name of the function defined by $f \circ g(x) \equiv f(g(x))$. In the notation of this chapter, the chain rule is written as

$$Df(g(x))Dg(x) = D(f \circ g)(x). \quad (22.9)$$

22.6.2 The Chain Rule For Functions Of Many Variables

Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^p$ be open sets and let $\mathbf{f}$ be a function defined on V having values in $\mathbb{R}^q$ while $\mathbf{g}$ is a function defined on U such that $\mathbf{g}(U) \subseteq V$ as in the following picture.

$$U \xrightarrow{\mathbf{g}} V \xrightarrow{\mathbf{f}} \mathbb{R}^q$$

The chain rule says that if the linear transformations (matrices) on the left in 22.9 both exist then the same formula holds in this more general case. Thus

$$D\mathbf{f}(\mathbf{g}(\mathbf{x}))D\mathbf{g}(\mathbf{x}) = D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})$$

Note this all makes sense because $D\mathbf{f}(\mathbf{g}(\mathbf{x}))$ is a $q \times p$ matrix and $D\mathbf{g}(\mathbf{x})$ is a $p \times n$ matrix. Remember it is all right to do $(q \times p)(p \times n)$. The middle numbers match. More precisely,

Theorem 22.6.1 (*Chain rule*) Let U be an open set in $\mathbb{R}^n$, let V be an open set in $\mathbb{R}^p$, let $\mathbf{g} : U \rightarrow \mathbb{R}^p$ be such that $\mathbf{g}(U) \subseteq V$, and let $\mathbf{f} : V \rightarrow \mathbb{R}^q$. Suppose $D\mathbf{g}(\mathbf{x})$ exists for some $\mathbf{x} \in U$ and that $D\mathbf{f}(\mathbf{g}(\mathbf{x}))$ exists. Then $D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})$ exists and furthermore,

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{x}) = D\mathbf{f}(\mathbf{g}(\mathbf{x}))D\mathbf{g}(\mathbf{x}). \quad (22.10)$$

In particular,

$$\frac{\partial(\mathbf{f} \circ \mathbf{g})(\mathbf{x})}{\partial x_j} = \sum_{i=1}^p \frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial y_i} \frac{\partial g_i(\mathbf{x})}{\partial x_j}. \quad (22.11)$$

There is an easy way to remember this in terms of the repeated index summation convention presented earlier. Let $\mathbf{y} = \mathbf{g}(\mathbf{x})$ and $\mathbf{z} = \mathbf{f}(\mathbf{y})$. Then the above says

$$\frac{\partial \mathbf{z}}{\partial y_i} \frac{\partial y_i}{\partial x_k} = \frac{\partial \mathbf{z}}{\partial x_k}. \quad (22.12)$$

Remember there is a sum on the repeated index. In particular, for each index, r ,

$$\frac{\partial z_r}{\partial y_i} \frac{\partial y_i}{\partial x_k} = \frac{\partial z_r}{\partial x_k}.$$

The proof of this major theorem will be given at the end of this section. It will include the chain rule for functions of one variable as a special case. First here are some examples.

Example 22.6.2 Let $f(u, v) = \sin(uv)$ and let $u(x, y, t) = t \sin x + \cos y$ and $v(x, y, t, s) = s \tan x + y^2 + ts$. Letting $z = f(u, v)$ where u, v are as just described, find $\frac{\partial z}{\partial t}$ and $\frac{\partial z}{\partial x}$.

From 22.12,

$$\begin{aligned}\frac{\partial z}{\partial t} &= \frac{\partial z}{\partial u} \frac{\partial u}{\partial t} + \frac{\partial z}{\partial v} \frac{\partial v}{\partial t} \\ &= v \cos(uv) \sin(x) + us \cos(uv).\end{aligned}$$

Here $y_1 = u, y_2 = v, t = x_k$. Also,

$$\begin{aligned}\frac{\partial z}{\partial x} &= \frac{\partial z}{\partial u} \frac{\partial u}{\partial x} + \frac{\partial z}{\partial v} \frac{\partial v}{\partial x} = v \cos(uv) t \cos(x) + \\ &\quad + us \sec^2(x) \cos(uv).\end{aligned}$$

Clearly you can continue in this way taking partial derivatives with respect to any of the other variables.

Example 22.6.3 Let $w = f(u_1, u_2) = u_2 \sin(u_1)$ and $u_1 = x^2y + z, u_2 = \sin(xy)$. Find $\frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}$, and $\frac{\partial w}{\partial z}$.

The derivative of f is of the form (w_x, w_y, w_z) and so it suffices to find the derivative of f using the chain rule. You need to find $Df(u_1, u_2) D\mathbf{g}(x, y, z)$ where

$$\mathbf{g}(x, y) = \begin{pmatrix} x^2y + z \\ \sin(xy) \end{pmatrix}.$$

Then

$$D\mathbf{g}(x, y, z) = \begin{pmatrix} 2xy & x^2 & 1 \\ y \cos(xy) & x \cos(xy) & 0 \end{pmatrix}.$$

Also $Df(u_1, u_2) = (u_2 \cos(u_1), \sin(u_1))$. Therefore, the derivative is

$$\begin{aligned}& Df(u_1, u_2) D\mathbf{g}(x, y, z) \\ &= (u_2 \cos(u_1), \sin(u_1)) \begin{pmatrix} 2xy & x^2 & 1 \\ y \cos(xy) & x \cos(xy) & 0 \end{pmatrix} \\ &= (2u_2 (\cos u_1) xy + (\sin u_1) y \cos xy, u_2 (\cos u_1) x^2 \\ &\quad + (\sin u_1) x \cos xy, u_2 \cos u_1) \\ &= (w_x, w_y, w_z)\end{aligned}$$

Thus

$$\begin{aligned}\frac{\partial w}{\partial x} &= 2u_2 (\cos u_1) xy + (\sin u_1) y \cos xy \\ &= 2(\sin(xy)) (\cos(x^2y + z)) xy \\ &\quad + (\sin(x^2y + z)) y \cos xy.\end{aligned}$$

Similarly, you can find the other partial derivatives of w in terms of substituting in for u_1 and u_2 in the above. Note

$$\frac{\partial w}{\partial x} = \frac{\partial w}{\partial u_1} \frac{\partial u_1}{\partial x} + \frac{\partial w}{\partial u_2} \frac{\partial u_2}{\partial x}.$$

In fact, in general if you have $w = f(u_1, u_2)$ and $\mathbf{g}(x, y, z) = \begin{pmatrix} u_1(x, y, z) \\ u_2(x, y, z) \end{pmatrix}$, then $D(f \circ \mathbf{g})(x, y, z)$ is of the form

$$\begin{aligned}& \begin{pmatrix} w_{u_1} & w_{u_2} \end{pmatrix} \begin{pmatrix} u_{1x} & u_{1y} & u_{1z} \\ u_{2x} & u_{2y} & u_{2z} \end{pmatrix} \\ &= \begin{pmatrix} w_{u_1} u_x + w_{u_2} u_{2x} & w_{u_1} u_y + w_{u_2} u_{2y} & w_{u_1} u_z + w_{u_2} u_{2z} \end{pmatrix}.\end{aligned}$$

Example 22.6.4 Let $w = f(u_1, u_2, u_3) = u_1^2 + u_3 + u_2$ and $\mathbf{g}(x, y, z) = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} x + 2yz \\ x^2 + y \\ z^2 + x \end{pmatrix}$. Find $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial z}$.

By the chain rule,

$$\begin{aligned} & (w_x, w_y, w_z) \\ &= \begin{pmatrix} w_{u_1} & w_{u_2} & w_{u_3} \end{pmatrix} \begin{pmatrix} u_{1x} & u_{1y} & u_{1z} \\ u_{2x} & u_{2y} & u_{2z} \\ u_{3x} & u_{3y} & u_{3z} \end{pmatrix} = \\ & (w_{u_1}u_{1x} + w_{u_2}u_{2x} + w_{u_3}u_{3x}, w_{u_1}u_{1y} + w_{u_2}u_{2y} + w_{u_3}u_{3y}, w_{u_1}u_{1z} + w_{u_2}u_{2z} + w_{u_3}u_{3z}) \end{aligned}$$

Note the pattern,

$$\begin{aligned} w_x &= w_{u_1}u_{1x} + w_{u_2}u_{2x} + w_{u_3}u_{3x}, \\ w_y &= w_{u_1}u_{1y} + w_{u_2}u_{2y} + w_{u_3}u_{3y}, \\ w_z &= w_{u_1}u_{1z} + w_{u_2}u_{2z} + w_{u_3}u_{3z}. \end{aligned}$$

Therefore,

$$w_x = 2u_1(1) + 1(2x) + 1(1) = 2(x + 2yz) + 2x + 1 = 4x + 4yz + 1$$

and

$$w_z = 2u_1(2y) + 1(0) + 1(2z) = 4(x + 2yz)y + 2z = 4yx + 8y^2z + 2z.$$

Of course to find all the partial derivatives at once, you just use the chain rule. Thus you would get

$$\begin{aligned} & (w_x \quad w_y \quad w_z) \\ &= \begin{pmatrix} 2u_1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2z & 2y \\ 2x & 1 & 0 \\ 1 & 0 & 2z \end{pmatrix} \\ &= \begin{pmatrix} 2u_1 + 2x + 1 & 4u_1z + 1 & 4u_1y + 2z \end{pmatrix} \\ &= \begin{pmatrix} 4x + 4yz + 1 & 4zx + 8yz^2 + 1 & 4yx + 8y^2z + 2z \end{pmatrix} \end{aligned}$$

Example 22.6.5 Let $\mathbf{f}(u_1, u_2) = \begin{pmatrix} u_1^2 + u_2 \\ \sin(u_2) + u_1 \end{pmatrix}$ and

$$\mathbf{g}(x_1, x_2, x_3) = \begin{pmatrix} u_1(x_1, x_2, x_3) \\ u_2(x_1, x_2, x_3) \end{pmatrix} = \begin{pmatrix} x_1x_2 + x_3 \\ x_2^2 + x_1 \end{pmatrix}.$$

Find $D(\mathbf{f} \circ \mathbf{g})(x_1, x_2, x_3)$.

To do this,

$$D\mathbf{f}(u_1, u_2) = \begin{pmatrix} 2u_1 & 1 \\ 1 & \cos u_2 \end{pmatrix}, D\mathbf{g}(x_1, x_2, x_3) = \begin{pmatrix} x_2 & x_1 & 1 \\ 1 & 2x_2 & 0 \end{pmatrix}.$$

Then

$$D\mathbf{f}(\mathbf{g}(x_1, x_2, x_3)) = \begin{pmatrix} 2(x_1x_2 + x_3) & 1 \\ 1 & \cos(x_2^2 + x_1) \end{pmatrix}$$

and so by the chain rule,

$$\begin{aligned}
 & D(\mathbf{f} \circ \mathbf{g})(x_1, x_2, x_3) \\
 &= \overbrace{\begin{pmatrix} 2(x_1x_2 + x_3) & 1 \\ 1 & \cos(x_2^2 + x_1) \end{pmatrix}}^{D\mathbf{f}(\mathbf{g}(\mathbf{x}))} \overbrace{\begin{pmatrix} x_2 & x_1 & 1 \\ 1 & 2x_2 & 0 \end{pmatrix}}^{D\mathbf{g}(\mathbf{x})} \\
 &= \begin{pmatrix} (2x_1x_2 + 2x_3)x_2 + 1 & (2x_1x_2 + 2x_3)x_1 + 2x_2 & 2x_1x_2 + 2x_3 \\ x_2 + \cos(x_2^2 + x_1) & x_1 + 2x_2(\cos(x_2^2 + x_1)) & 1 \end{pmatrix}
 \end{aligned}$$

Therefore, in particular,

$$\frac{\partial f_1 \circ \mathbf{g}}{\partial x_1}(x_1, x_2, x_3) = (2x_1x_2 + 2x_3)x_2 + 1,$$

$$\frac{\partial f_2 \circ \mathbf{g}}{\partial x_3}(x_1, x_2, x_3) = 1, \quad \frac{\partial f_2 \circ \mathbf{g}}{\partial x_2}(x_1, x_2, x_3) = x_1 + 2x_2(\cos(x_2^2 + x_1)).$$

etc.

In different notation, let $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \mathbf{f}(u_1, u_2) = \begin{pmatrix} u_1^2 + u_2 \\ \sin(u_2) + u_1 \end{pmatrix}$. Then

$$\begin{aligned}
 \frac{\partial z_1}{\partial x_1} &= \frac{\partial z_1}{\partial u_1} \frac{\partial u_1}{\partial x_1} + \frac{\partial z_1}{\partial u_2} \frac{\partial u_2}{\partial x_1} \\
 &= 2u_1x_2 + 1 = 2(x_1x_2 + x_3)x_2 + 1.
 \end{aligned}$$

Example 22.6.6 Let

$$\mathbf{f}(u_1, u_2, u_3) = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} u_1^2 + u_2u_3 \\ u_1^2 + u_3^2 \\ \ln(1 + u_3^2) \end{pmatrix}$$

and let

$$\begin{aligned}
 & \mathbf{g}(x_1, x_2, x_3, x_4) \\
 &= \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_2^2 + \sin(x_3) + \cos(x_4) \\ x_4^2 - x_1 \\ x_3^2 + x_4 \end{pmatrix}.
 \end{aligned}$$

Find $(\mathbf{f} \circ \mathbf{g})'(\mathbf{x})$.

$$D\mathbf{f}(\mathbf{u}) = \begin{pmatrix} 2u_1 & u_3 & u_2 \\ 2u_1 & 3u_2^2 & 0 \\ 0 & 0 & \frac{2u_3}{(1+u_3^2)} \end{pmatrix}$$

Similarly,

$$D\mathbf{g}(\mathbf{x}) = \begin{pmatrix} 1 & 2x_2 & \cos(x_3) & -\sin(x_4) \\ -1 & 0 & 0 & 2x_4 \\ 0 & 0 & 2x_3 & 1 \end{pmatrix}.$$

Then by the chain rule, $D(\mathbf{f} \circ \mathbf{g})(\mathbf{x}) = D\mathbf{f}(\mathbf{u})D\mathbf{g}(\mathbf{x})$ where $\mathbf{u} = \mathbf{g}(\mathbf{x})$ as described above. Thus $D(\mathbf{f} \circ \mathbf{g})(\mathbf{x}) =$

$$\begin{pmatrix} 2u_1 & u_3 & u_2 \\ 2u_1 & 3u_2^2 & 0 \\ 0 & 0 & \frac{2u_3}{(1+u_3^2)} \end{pmatrix} \begin{pmatrix} 1 & 2x_2 & \cos(x_3) & -\sin(x_4) \\ -1 & 0 & 0 & 2x_4 \\ 0 & 0 & 2x_3 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 2u_1 - u_3 & 4u_1x_2 & 2u_1 \cos x_3 + 2u_2x_3 & -2u_1 \sin x_4 + 2u_3x_4 + u_2 \\ 2u_1 - 3u_2^2 & 4u_1x_2 & 2u_1 \cos x_3 & -2u_1 \sin x_4 + 6u_2^2x_4 \\ 0 & 0 & 4\frac{u_3}{1+u_3^2}x_3 & 2\frac{u_3}{1+u_3^2} \end{pmatrix} \quad (22.13)$$

where each u_i is given by the above formulas. Thus $\frac{\partial z_1}{\partial x_1}$ equals

$$\begin{aligned} 2u_1 - u_3 &= 2(x_1 + x_2^2 + \sin(x_3) + \cos(x_4)) - (x_3^2 + x_4) \\ &= 2x_1 + 2x_2^2 + 2\sin x_3 + 2\cos x_4 - x_3^2 - x_4. \end{aligned}$$

while $\frac{\partial z_2}{\partial x_4}$ equals

$$\begin{aligned} &-2u_1 \sin x_4 + 6u_2^2x_4 \\ &= -2(x_1 + x_2^2 + \sin(x_3) + \cos(x_4)) \sin(x_4) + 6(x_4^2 - x_1)^2 x_4. \end{aligned}$$

If you wanted $\frac{\partial \mathbf{z}}{\partial x_2}$ it would be the second column of the above matrix in 22.13. Thus $\frac{\partial \mathbf{z}}{\partial x_2}$ equals

$$\begin{aligned} \begin{pmatrix} \frac{\partial z_1}{\partial x_2} \\ \frac{\partial z_2}{\partial x_2} \\ \frac{\partial z_3}{\partial x_2} \end{pmatrix} &= \begin{pmatrix} 4u_1x_2 \\ 4u_1x_2 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 4(x_1 + x_2^2 + \sin(x_3) + \cos(x_4))x_2 \\ 4(x_1 + x_2^2 + \sin(x_3) + \cos(x_4))x_2 \\ 0 \end{pmatrix}. \end{aligned}$$

I hope that by now it is clear that all the information you could desire about various partial derivatives is available and it all reduces to matrix multiplication and the consideration of entries of the matrix obtained by multiplying the two derivatives.

22.7 Exercises

1. Let $z = f(\mathbf{y}) = (y_1^2 + \sin y_2 + \tan y_3)$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + x_2 \\ x_2^2 - x_1 + x_2 \\ x_2^2 + x_1 + \sin x_2 \end{pmatrix}$.

Find $D(f \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z}{\partial x_i}$ for $i = 1, 2$.

2. Let $z = f(\mathbf{y}) = (y_1^2 + \cot y_2 + \sin y_3)$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + x_4 + x_3 \\ x_2^2 - x_1 + x_2 \\ x_2^2 + x_1 + \sin x_4 \end{pmatrix}$.

Find $D(f \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z}{\partial x_i}$ for $i = 1, 2, 3, 4$.

3. Let $z = f(\mathbf{y}) = (y_1^2 + y_2^2 + \sin y_3 + y_4)$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + x_4 + x_3 \\ x_2^2 - x_1 + x_2 \\ x_2^2 + x_1 + \sin x_4 \\ x_4 + x_2 \end{pmatrix}$.

Find $D(f \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z}{\partial x_i}$ for $i = 1, 2, 3, 4$.

4. Let $\mathbf{z} = \mathbf{f}(\mathbf{y}) = \begin{pmatrix} y_1^2 + \sin y_2 + \tan y_3 \\ y_1^2 y_2 + y_3 \end{pmatrix}$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + x_2 \\ x_2^2 - x_1 + x_2 \\ x_2^2 + x_1 + \sin x_2 \end{pmatrix}$.

Find $D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z_k}{\partial x_i}$ for $i = 1, 2$ and $k = 1, 2$.

5. Let $\mathbf{z} = \mathbf{f}(\mathbf{y}) = \begin{pmatrix} y_1^2 + \sin y_2 + \tan y_3 \\ y_1^2 y_2 + y_3 \\ \cos(y_1^2) + y_2^3 y_3 \end{pmatrix}$ and $\mathbf{y} = \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} x_1 + x_4 \\ x_2^2 - x_1 + x_3 \\ x_3^2 + x_1 + \sin x_2 \end{pmatrix}$.

Find $D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})$. Use to write $\frac{\partial z_k}{\partial x_i}$ for $i = 1, 2, 3, 4$ and $k = 1, 2, 3$.

6. Give a version of the chain rule which involves three functions $\mathbf{f}, \mathbf{g}, \mathbf{h}$. If $\mathbf{f}: U \rightarrow V$ and $\mathbf{f}^{-1}: V \rightarrow U$ for U, V open sets such that $\mathbf{f}, \mathbf{f}^{-1}$ are both differentiable, show that

$$\det(D\mathbf{f}(\mathbf{f}^{-1}(\mathbf{y}))) \det(D\mathbf{f}^{-1}(\mathbf{y})) = 1.$$

22.7.1 Related Rates Problems

Sometimes several variables are related and, given information about how one variable is changing, you want to find how the others are changing. The following law is discussed later in the book, on Page 606.

Example 22.7.1 *Bernoulli's law states that in an incompressible fluid,*

$$\frac{v^2}{2g} + z + \frac{P}{\gamma} = C$$

where C is a constant. Here v is the speed, P is the pressure, and z is the height above some reference point. The constants, g and γ are the acceleration of gravity and the weight density of the fluid. Suppose measurements indicate that $\frac{dv}{dt} = -3$, and $\frac{dz}{dt} = 2$. Find $\frac{dP}{dt}$ when $v = 7$ and $z = 8$ in terms of g and γ .

This is just an exercise in using the chain rule. Differentiate the two sides with respect to t .

$$\frac{1}{g} v \frac{dv}{dt} + \frac{dz}{dt} + \frac{1}{\gamma} \frac{dP}{dt} = 0.$$

Then when $v = 7$ and $z = 8$, finding $\frac{dP}{dt}$ involves nothing more than solving the following for $\frac{dP}{dt}$.

$$\frac{7}{g}(-3) + 2 + \frac{1}{\gamma} \frac{dP}{dt} = 0$$

Thus

$$\frac{dP}{dt} = \gamma \left(\frac{21}{g} - 2 \right)$$

at this instant in time.

Example 22.7.2 *In Bernoulli's law above, each of v, z , and P are functions of (x, y, z) , the position of a point in the fluid. Find a formula for $\frac{\partial P}{\partial x}$ in terms of the partial derivatives of the other variables.*

This is an example of the chain rule. Differentiate both sides with respect to x .

$$\frac{v}{g} v_x + z_x + \frac{1}{\gamma} P_x = 0$$

and so

$$P_x = - \left(\frac{v v_x + z_x g}{g} \right) \gamma$$

Example 22.7.3 *Suppose a level curve is of the form $f(x, y) = C$ and that near a point on this level curve, y is a differentiable function of x . Find $\frac{dy}{dx}$.*

This is an example of the chain rule. Differentiate both sides with respect to x . This gives

$$f_x + f_y \frac{dy}{dx} = 0.$$

Solving for $\frac{dy}{dx}$ gives

$$\frac{dy}{dx} = \frac{-f_x(x, y)}{f_y(x, y)}.$$

Example 22.7.4 Suppose a level surface is of the form $f(x, y, z) = C$. and that near a point, (x, y, z) on this level surface, z is a C^1 function of x and y . Find a formula for z_x .

This is an example of the use of the chain rule. Differentiate both sides of the equation with respect to x . Since $y_x = 0$,

$$f_x + f_z z_x = 0.$$

Then solving for z_x ,

$$z_x = \frac{-f_x(x, y, z)}{f_z(x, y, z)}$$

Example 22.7.5 Polar coordinates are

$$x = r \cos \theta, \quad y = r \sin \theta. \quad (22.14)$$

Thus if f is a C^1 scalar valued function you could ask to express f_x in terms of the variables, r and θ . Do so.

This is an example of the chain rule. Abusing notation slightly, regard f as a function of position in the plane. This position can be described with any set of coordinates. Thus $f(x, y) = f(r, \theta)$ and so

$$f_x = f_r r_x + f_\theta \theta_x.$$

This will be done if you can find r_x and θ_x . However you must find these in terms of r and θ , not in terms of x and y . Using the chain rule on the two equations for the transformation in 22.14,

$$\begin{aligned} 1 &= r_x \cos \theta - (r \sin \theta) \theta_x \\ 0 &= r_x \sin \theta + (r \cos \theta) \theta_x \end{aligned}$$

Solving these using Cramer's rule,

$$r_x = \cos(\theta), \quad \theta_x = \frac{-\sin(\theta)}{r}$$

Hence f_x in polar coordinates is

$$f_x = f_r(r, \theta) \cos(\theta) - f_\theta(r, \theta) \left(\frac{\sin(\theta)}{r} \right)$$

22.7.2 The Derivative Of The Inverse Function

Example 22.7.6 Let $\mathbf{f} : U \rightarrow V$ where U and V are open sets in $\mathbb{R}^n$ and $\mathbf{f}$ is one to one and onto. Suppose also that $\mathbf{f}$ and $\mathbf{f}^{-1}$ are both differentiable. How are $D\mathbf{f}^{-1}$ and $D\mathbf{f}$ related?

This can be done as follows. From the assumptions, $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{f}(\mathbf{x}))$. Let $I\mathbf{x} = \mathbf{x}$. Then by Example 22.3.6 on Page 500 $DI = I$. By the chain rule,

$$\begin{aligned} I &= DI = D\mathbf{f}^{-1}(\mathbf{f}(\mathbf{x}))(D\mathbf{f}(\mathbf{x})) \\ I &= DI = D\mathbf{f}(\mathbf{f}^{-1}(\mathbf{y}))(D\mathbf{f}^{-1}(\mathbf{y})) \end{aligned}$$

Letting $\mathbf{y} = \mathbf{f}(\mathbf{x})$, the second yields

$$I = D\mathbf{f}(\mathbf{x}) D\mathbf{f}^{-1}(\mathbf{f}(\mathbf{x})).$$

Therefore,

$$D\mathbf{f}(\mathbf{x})^{-1} = D\mathbf{f}^{-1}(\mathbf{f}(\mathbf{x})).$$

This is equivalent to

$$D\mathbf{f}(\mathbf{f}^{-1}(\mathbf{y}))^{-1} = D\mathbf{f}^{-1}(\mathbf{y})$$

or

$$D\mathbf{f}(\mathbf{x})^{-1} = D\mathbf{f}^{-1}(\mathbf{y}), \mathbf{y} = \mathbf{f}(\mathbf{x}).$$

This is just like a similar situation for functions of one variable. Remember

$$(f^{-1})'(f(x)) = 1/f'(x).$$

In terms of the repeated index summation convention, suppose $\mathbf{y} = \mathbf{f}(\mathbf{x})$ so that $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{y})$. Then the above can be written as

$$\delta_{ij} = \frac{\partial x_i}{\partial y_k}(\mathbf{f}(\mathbf{x})) \frac{\partial y_k}{\partial x_j}(\mathbf{x}).$$

22.7.3 Proof Of The Chain Rule

As in the case of a function of one variable, it is important to consider the derivative of a composition of two functions. As in the case of a function of one variable, this rule is called the chain rule. Its proof depends on the following fundamental lemma. This proof will include the one dimensional case.

Lemma 22.7.7 Let $\mathbf{g} : U \rightarrow \mathbb{R}^p$ where U is an open set in $\mathbb{R}^n$ and suppose $\mathbf{g}$ has a derivative at $\mathbf{x} \in U$. Then $\mathbf{o}(\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x})) = \mathbf{o}(\mathbf{v})$.

Proof: It is necessary to show that

$$\lim_{\mathbf{v} \rightarrow \mathbf{0}} \frac{|\mathbf{o}(\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x}))|}{|\mathbf{v}|} = 0. \quad (22.15)$$

From Lemma 22.5.6, there exists $\delta > 0$ such that if $|\mathbf{v}| < \delta$, then

$$|\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x})| \leq (Cn + 1) |\mathbf{v}|. \quad (22.16)$$

Now let $\varepsilon > 0$ be given. There exists $\eta > 0$ such that if $|\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x})| < \eta$, then

$$|\mathbf{o}(\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x}))| < \left(\frac{\varepsilon}{Cn + 1} \right) |\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x})| \quad (22.17)$$

Let $|\mathbf{v}| < \min\left(\delta, \frac{\eta}{Cn+1}\right)$. For such $\mathbf{v}$, $|\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x})| \leq \eta$, which implies

$$\begin{aligned} |\mathbf{o}(\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x}))| &< \left(\frac{\varepsilon}{Cn+1}\right) |\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x})| \\ &< \left(\frac{\varepsilon}{Cn+1}\right) (Cn+1) |\mathbf{v}| \end{aligned}$$

and so

$$\frac{|\mathbf{o}(\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x}))|}{|\mathbf{v}|} < \varepsilon$$

which establishes 22.15. This proves the lemma.

Recall the notation $\mathbf{f} \circ \mathbf{g}(\mathbf{x}) \equiv \mathbf{f}(\mathbf{g}(\mathbf{x}))$. Thus $\mathbf{f} \circ \mathbf{g}$ is the name of a function and this function is defined by what was just written. The following theorem is known as the **chain rule**.

Theorem 22.7.8 (*Chain rule*) Let U be an open set in $\mathbb{R}^n$, let V be an open set in $\mathbb{R}^p$, let $\mathbf{g} : U \rightarrow \mathbb{R}^p$ be such that $\mathbf{g}(U) \subseteq V$, and let $\mathbf{f} : V \rightarrow \mathbb{R}^q$. Suppose $D\mathbf{g}(\mathbf{x})$ exists for some $\mathbf{x} \in U$ and that $D\mathbf{f}(\mathbf{g}(\mathbf{x}))$ exists. Then $D(\mathbf{f} \circ \mathbf{g})(\mathbf{x})$ exists and furthermore,

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{x}) = D\mathbf{f}(\mathbf{g}(\mathbf{x})) D\mathbf{g}(\mathbf{x}). \quad (22.18)$$

In particular, If $\mathbf{y} = \mathbf{g}(\mathbf{x})$ so $y_i = g_i(\mathbf{x})$,

$$\frac{\partial(\mathbf{f} \circ \mathbf{g})(\mathbf{x})}{\partial x_j} = \sum_{i=1}^p \frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{x}))}{\partial y_i} \frac{\partial g_i(\mathbf{x})}{\partial x_j}. \quad (22.19)$$

Proof: From the assumption that $D\mathbf{f}(\mathbf{g}(\mathbf{x}))$ exists,

$$\begin{aligned} \mathbf{f}(\mathbf{g}(\mathbf{x} + \mathbf{v})) &= \mathbf{f}(\mathbf{g}(\mathbf{x})) + D\mathbf{f}(\mathbf{g}(\mathbf{x}))(\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x})) + \mathbf{o}(\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x})) \\ &= \mathbf{f}(\mathbf{g}(\mathbf{x})) + D\mathbf{f}(\mathbf{g}(\mathbf{x}))(D\mathbf{g}(\mathbf{x})\mathbf{v} + \mathbf{o}(\mathbf{v})) + \mathbf{o}(\mathbf{g}(\mathbf{x} + \mathbf{v}) - \mathbf{g}(\mathbf{x})) \end{aligned}$$

which by Lemma 22.7.7 equals

$$\begin{aligned} &= \mathbf{f}(\mathbf{g}(\mathbf{x})) + D\mathbf{f}(\mathbf{g}(\mathbf{x})) D\mathbf{g}(\mathbf{x})\mathbf{v} + D\mathbf{f}(\mathbf{g}(\mathbf{x}))\mathbf{o}(\mathbf{v}) + \mathbf{o}(\mathbf{v}) \\ &= \mathbf{f}(\mathbf{g}(\mathbf{x})) + D\mathbf{f}(\mathbf{g}(\mathbf{x})) D\mathbf{g}(\mathbf{x})\mathbf{v} + \mathbf{o}(\mathbf{v}) \end{aligned}$$

and this shows

$$D(\mathbf{f} \circ \mathbf{g})(\mathbf{x}) = D\mathbf{f}(\mathbf{g}(\mathbf{x})) D\mathbf{g}(\mathbf{x})$$

from the definition of the derivative and its uniqueness established in Theorem 22.3.3 on Page 498. This proves the theorem.

22.8 Exercises

1. Suppose $\mathbf{f} : U \rightarrow \mathbb{R}^q$ and let $\mathbf{x} \in U$ and $\mathbf{v}$ be a unit vector. Show that $D_{\mathbf{v}}\mathbf{f}(\mathbf{x}) = D\mathbf{f}(\mathbf{x})\mathbf{v}$. Recall that

$$D_{\mathbf{v}}\mathbf{f}(\mathbf{x}) \equiv \lim_{t \rightarrow 0} \frac{\mathbf{f}(\mathbf{x} + t\mathbf{v}) - \mathbf{f}(\mathbf{x})}{t}.$$

2. Let $f(x, y) = \begin{cases} xy \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$. Find where f is differentiable and compute the derivative at all these points.

3. Let

$$f(x, y) = \begin{cases} x & \text{if } |y| > |x| \\ -x & \text{if } |y| \leq |x| \end{cases}.$$

Show that f is continuous at $(0, 0)$ and that the partial derivatives exist at $(0, 0)$ but the function is not differentiable at $(0, 0)$.

4. Let

$$\mathbf{f}(x, y, z) = \begin{pmatrix} x^2 \sin y + z^3 \\ \sin(x + y) + z^3 \cos x \end{pmatrix}.$$

Find $D\mathbf{f}(1, 2, 3)$.

5. Let

$$\mathbf{f}(x, y, z) = \begin{pmatrix} x \tan y + z^3 \\ \cos(x + y) + z^3 \cos x \end{pmatrix}.$$

Find $D\mathbf{f}(x, y, z)$.

6. Let

$$\mathbf{f}(x, y, z) = \begin{pmatrix} x \sin y + z^3 \\ \sin(x + y) + z^3 \cos x \\ x^5 + y^2 \end{pmatrix}.$$

Find $D\mathbf{f}(x, y, z)$.

7. Let

$$f(x, y) = \begin{cases} \frac{(x^2 - y^4)^2}{(x^2 + y^4)^2} & \text{if } (x, y) \neq (0, 0) \\ 1 & \text{if } (x, y) = (0, 0) \end{cases}.$$

Show that all directional derivatives of f exist at $(0, 0)$, and are all equal to zero but the function is not even continuous at $(0, 0)$. Therefore, it is not differentiable. Why?

8. In the example of Problem 7 show that the partial derivatives exist but are not continuous.
9. A certain building is shaped like the top half of the ellipsoid, $\frac{x^2}{900} + \frac{y^2}{900} + \frac{z^2}{400} = 1$ determined by letting $z \geq 0$. Here dimensions are measured in feet. The building needs to be painted. The paint, when applied is about .005 feet thick. About how many cubic feet of paint will be needed. **Hint:** This is going to replace the numbers, 900 and 400 with slightly larger numbers when the ellipsoid is fattened slightly by the paint. The volume of the top half of the ellipsoid, $x^2/a^2 + y^2/b^2 + z^2/c^2 \leq 1, z \geq 0$ is $(2/3)\pi abc$.
10. Suppose $\mathbf{r}_1(t) = (\cos t, \sin t, t)$, $\mathbf{r}_2(t) = (t, 2t, 1)$, and $\mathbf{r}_3(t) = (1, t, 1)$. Find the rate of change with respect to t of the volume of the parallelepiped determined by these three vectors when $t = 1$.
11. A trash compacter is compacting a rectangular block of trash. The width is changing at the rate of -1 inches per second, the length is changing at the rate of -2 inches per second and the height is changing at the rate of -3 inches per second. How fast is the volume changing when the length is 20, the height is 10, and the width is 10?
12. A trash compacter is compacting a rectangular block of trash. The width is changing at the rate of -2 inches per second, the length is changing at the rate of -1 inches per second and the height is changing at the rate of -4 inches per second. How fast is the surface area changing when the length is 20, the height is 10, and the width is 10?

13. The ideal gas law is $PV = kT$ where k is a constant which depends on the number of moles and on the gas being considered. If V is changing at the rate of 2 cubic cm. per second and T is changing at the rate of 3 degrees Kelvin per second, how fast is the pressure changing when $T = 300$ and V equals 400 cubic cm.?
14. Let S denote a level surface of the form $f(x_1, x_2, x_3) = C$. Show that any smooth curve in the level surface is perpendicular to the gradient.
15. Suppose $\mathbf{f}$ is a C^1 function which maps U , an open subset of $\mathbb{R}^n$ one to one and onto V , an open set in $\mathbb{R}^m$ such that the inverse map, $\mathbf{f}^{-1}$ is also C^1 . What must be true of m and n ? Why? **Hint:** Consider Example 22.7.6 on Page 513. Also you can use the fact that if A is an $m \times n$ matrix which maps $\mathbb{R}^n$ onto $\mathbb{R}^m$, then $m \leq n$.
16. Finish Example 22.7.5 by finding f_y in terms of θ, r . Show that $f_y = \sin(\theta) f_r + \frac{\cos(\theta)}{r} f_\theta$.
17. *Think of ∂_x as a differential operator which takes functions and differentiates them with respect to x . Thus $\partial_x f \equiv f_x$. In the context of Example 22.7.5, which is on polar coordinates, and Problem 16, explain how

$$\begin{aligned}\partial_x &= \cos(\theta) \partial_r - \frac{\sin(\theta)}{r} \partial_\theta \\ \partial_y &= \sin(\theta) \partial_r + \frac{\cos(\theta)}{r} \partial_\theta\end{aligned}$$

The Laplacian of a function u is defined as $\Delta u = u_{xx} + u_{yy}$. Use the above observation to give a formula Δu in terms of r and θ . You should get $u_{rr} + \frac{1}{r} u_r + \frac{1}{r^2} u_{\theta\theta}$.

22.9 The Gradient

Here we review the concept of the gradient. This has already been considered in the special case of a C^1 function. However, you do not need so much to define the gradient.

Let $f : U \rightarrow \mathbb{R}$ where U is an open subset of $\mathbb{R}^n$ and suppose f is differentiable on U . Thus if $\mathbf{x} \in U$,

$$f(\mathbf{x} + \mathbf{v}) = f(\mathbf{x}) + \sum_{j=1}^n \frac{\partial f(\mathbf{x})}{\partial x_j} v_j + o(\mathbf{v}). \quad (22.20)$$

Recall Proposition 21.4.6, a more general version of which is stated here for convenience. It is more general because we only assume f is differentiable, not C^1 .

Proposition 22.9.1 *If f is differentiable at $\mathbf{x}$ and for $\mathbf{v}$ a unit vector,*

$$D_{\mathbf{v}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v}.$$

Proof:

$$\begin{aligned}\frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t} &= \frac{1}{t} \left(f(\mathbf{x}) + \sum_{j=1}^n \frac{\partial f(\mathbf{x})}{\partial x_j} t v_j + o(t\mathbf{v}) - f(\mathbf{x}) \right) \\ &= \frac{1}{t} \left(\sum_{j=1}^n \frac{\partial f(\mathbf{x})}{\partial x_j} t v_j + o(t\mathbf{v}) \right) \\ &= \sum_{j=1}^n \frac{\partial f(\mathbf{x})}{\partial x_j} v_j + \frac{o(t\mathbf{v})}{t}.\end{aligned}$$

Now $\lim_{t \rightarrow 0} \frac{o(t\mathbf{v})}{t} = 0$ and so

$$D_{\mathbf{v}}f(\mathbf{x}) = \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t} = \sum_{j=1}^n \frac{\partial f(\mathbf{x})}{\partial x_j} v_j = \nabla f(\mathbf{x}) \cdot \mathbf{v}$$

as claimed.

Definition 22.9.2 When f is differentiable, define

$$\nabla f(\mathbf{x}) \equiv \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right)^T$$

just as was done in the special case where f is C^1 . As before, this vector is called the **gradient vector**.

This defines the gradient for a differentiable scalar valued function. There are ways to define the gradient for vector valued functions but this will not be attempted in this book.

It follows immediately from 22.20 that

$$f(\mathbf{x} + \mathbf{v}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{v} + o(\mathbf{v}) \quad (22.21)$$

As mentioned above, an important aspect of the gradient is its relation with the directional derivative. A repeat of the above argument gives the following. From 22.21, for $\mathbf{v}$ a unit vector,

$$\begin{aligned} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t} &= \nabla f(\mathbf{x}) \cdot \mathbf{v} + \frac{o(t\mathbf{v})}{t} \\ &= \nabla f(\mathbf{x}) \cdot \mathbf{v} + \frac{o(t)}{t}. \end{aligned}$$

Therefore, taking $t \rightarrow 0$,

$$D_{\mathbf{v}}f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v}. \quad (22.22)$$

Example 22.9.3 Let $f(x, y, z) = x^2 + \sin(xy) + z$. Find $D_{\mathbf{v}}f(1, 0, 1)$ where

$$\mathbf{v} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right).$$

Note this vector which is given is already a unit vector. Therefore, from the above, it is only necessary to find $\nabla f(1, 0, 1)$ and take the dot product.

$$\nabla f(x, y, z) = (2x + (\cos xy)y, (\cos xy)x, 1).$$

Therefore, $\nabla f(1, 0, 1) = (2, 1, 1)$. Therefore, the directional derivative is

$$(2, 1, 1) \cdot \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right) = \frac{4}{3}\sqrt{3}.$$

Because of 22.22 it is easy to find the largest possible directional derivative and the smallest possible directional derivative. That which follows is a more algebraic treatment of an earlier result with the trigonometry removed.

Proposition 22.9.4 Let $f : U \rightarrow \mathbb{R}$ be a differentiable function and let $\mathbf{x} \in U$. Then

$$\max \{D_{\mathbf{v}}f(\mathbf{x}) : |\mathbf{v}| = 1\} = |\nabla f(\mathbf{x})| \quad (22.23)$$

and

$$\min \{D_{\mathbf{v}}f(\mathbf{x}) : |\mathbf{v}| = 1\} = -|\nabla f(\mathbf{x})|. \quad (22.24)$$

Furthermore, the maximum in 22.23 occurs when $\mathbf{v} = \nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$ and the minimum in 22.24 occurs when $\mathbf{v} = -\nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$.

Proof: From 22.22 and the Cauchy Schwarz inequality,

$$|D_{\mathbf{v}}f(\mathbf{x})| \leq |\nabla f(\mathbf{x})|$$

and so for any choice of $\mathbf{v}$ with $|\mathbf{v}| = 1$,

$$-|\nabla f(\mathbf{x})| \leq D_{\mathbf{v}}f(\mathbf{x}) \leq |\nabla f(\mathbf{x})|.$$

The proposition is proved by noting that if $\mathbf{v} = -\nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$, then

$$\begin{aligned} D_{\mathbf{v}}f(\mathbf{x}) &= \nabla f(\mathbf{x}) \cdot (-\nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|) \\ &= -|\nabla f(\mathbf{x})|^2 / |\nabla f(\mathbf{x})| = -|\nabla f(\mathbf{x})| \end{aligned}$$

while if $\mathbf{v} = \nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|$, then

$$\begin{aligned} D_{\mathbf{v}}f(\mathbf{x}) &= \nabla f(\mathbf{x}) \cdot (\nabla f(\mathbf{x}) / |\nabla f(\mathbf{x})|) \\ &= |\nabla f(\mathbf{x})|^2 / |\nabla f(\mathbf{x})| = |\nabla f(\mathbf{x})|. \end{aligned}$$

This proves the proposition.

For a different approach to the proposition, see Problem 6 which follows.

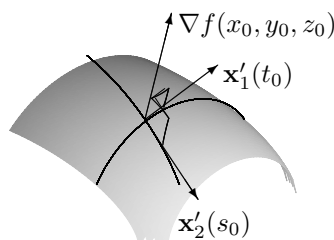
The conclusion of the above proposition is important in many physical models. For example, consider some material which is at various temperatures depending on location. Because it has cool places and hot places, it is expected that the heat will flow from the hot places to the cool places. Consider a small surface having a unit normal, $\mathbf{n}$. Thus $\mathbf{n}$ is a normal to this surface and has unit length. If it is desired to find the rate in calories per second at which heat crosses this little surface in the direction of $\mathbf{n}$ it is defined as $\mathbf{J} \cdot \mathbf{n}A$ where A is the area of the surface and $\mathbf{J}$ is called the heat flux. It is reasonable to suppose the rate at which heat flows across this surface will be largest when $\mathbf{n}$ is in the direction of greatest rate of decrease of the temperature. In other words, heat flows most readily in the direction which involves the maximum rate of decrease in temperature. This expectation will be realized by taking $\mathbf{J} = -K\nabla u$ where K is a positive scalar function which can depend on a variety of things. The above relation between the heat flux and ∇u is usually called the Fourier heat conduction law and the constant, K is known as the coefficient of thermal conductivity. It is a material property, different for iron than for aluminum. In most applications, K is considered to be a constant but this is wrong. Experiments show that this scalar should depend on temperature. Nevertheless, things get very difficult if this dependence is allowed. The constant can depend on position in the material or even on time.

An identical relationship is usually postulated for the flow of a diffusing species. In this problem, something like a pollutant diffuses. It may be an insecticide in ground water for example. Like heat, it tries to move from areas of high concentration toward areas of low concentration. In this case $\mathbf{J} = -K\nabla c$ where c is the concentration of the diffusing species. When applied to diffusion, this relationship is known as Fick's law. Mathematically, it is indistinguishable from the problem of heat flow.

Note the importance of the gradient in formulating these models.

22.10 The Gradient And Tangent Planes

The gradient has fundamental geometric significance illustrated by the following picture.



In this picture, the surface is a piece of a level surface of a function of three variables, $f(x, y, z)$. Thus the surface is defined by $f(x, y, z) = c$ or more completely as $\{(x, y, z) : f(x, y, z) = c\}$. For example, if $f(x, y, z) = x^2 + y^2 + z^2$, this would be a piece of a sphere. There are two smooth curves in this picture which lie in the surface having parameterizations, $\mathbf{x}_1(t) = (x_1(t), y_1(t), z_1(t))$ and $\mathbf{x}_2(s) = (x_2(s), y_2(s), z_2(s))$ which intersect at the point, (x_0, y_0, z_0) on this surface¹. This intersection occurs when $t = t_0$ and $s = s_0$. Since the points, $\mathbf{x}_1(t)$ for t in an interval lie in the level surface, it follows

$$f(x_1(t), y_1(t), z_1(t)) = c$$

for all t in some interval. Therefore, taking the derivative of both sides and using the chain rule on the left,

$$\begin{aligned} & \frac{\partial f}{\partial x}(x_1(t), y_1(t), z_1(t)) x_1'(t) + \\ & \frac{\partial f}{\partial y}(x_1(t), y_1(t), z_1(t)) y_1'(t) + \frac{\partial f}{\partial z}(x_1(t), y_1(t), z_1(t)) z_1'(t) = 0. \end{aligned}$$

In terms of the gradient, this merely states

$$\nabla f(x_1(t), y_1(t), z_1(t)) \cdot \mathbf{x}_1'(t) = 0.$$

Similarly,

$$\nabla f(x_2(s), y_2(s), z_2(s)) \cdot \mathbf{x}_2'(s) = 0.$$

Letting $s = s_0$ and $t = t_0$, it follows

$$\nabla f(x_0, y_0, z_0) \cdot \mathbf{x}_1'(t_0) = 0, \quad \nabla f(x_0, y_0, z_0) \cdot \mathbf{x}_2'(s_0) = 0.$$

It follows $\nabla f(x_0, y_0, z_0)$ is perpendicular to both the direction vectors of the two indicated curves shown. Surely if things are as they should be, these two direction vectors would determine a plane which deserves to be called the tangent plane to the level surface of f at the point (x_0, y_0, z_0) and that $\nabla f(x_0, y_0, z_0)$ is perpendicular to this tangent plane at the point, (x_0, y_0, z_0) .

Example 22.10.1 Find the equation of the tangent plane to the level surface, $f(x, y, z) = 6$ of the function, $f(x, y, z) = x^2 + 2y^2 + 3z^2$ at the point $(1, 1, 1)$.

First note that $(1, 1, 1)$ is a point on this level surface. To find the desired plane it suffices to find the normal vector to the proposed plane. But $\nabla f(x, y, z) = (2x, 4y, 6z)$ and so $\nabla f(1, 1, 1) = (2, 4, 6)$. Therefore, from this problem, the equation of the plane is

$$(2, 4, 6) \cdot (x - 1, y - 1, z - 1) = 0$$

¹Do there exist any smooth curves which lie in the level surface of f and pass through the point (x_0, y_0, z_0) ? It turns out there do if $\nabla f(x_0, y_0, z_0) \neq \mathbf{0}$ and if the function, f , is C^1 . However, this is a consequence of the implicit function theorem, one of the greatest theorems in all mathematics and a topic for an advanced calculus class. See the appendix on the subject found in this book for a relatively elementary presentation of this theorem.

or in other words,

$$2x - 12 + 4y + 6z = 0.$$

Example 22.10.2 The point, $(\sqrt{3}, 1, 4)$ is on both the surfaces, $z = x^2 + y^2$ and $z = 8 - (x^2 + y^2)$. Find the cosine of the angle between the two tangent planes at this point.

Recall this is the same as the angle between two normal vectors. Of course there is some ambiguity here because if $\mathbf{n}$ is a normal vector, then so is $-\mathbf{n}$ and replacing $\mathbf{n}$ with $-\mathbf{n}$ in the formula for the cosine of the angle will change the sign. We agree to look for the acute angle and its cosine rather than the obtuse angle. The normals are $(2\sqrt{3}, 2, -1)$ and $(2\sqrt{3}, 2, 1)$. Therefore, the cosine of the angle desired is

$$\frac{(2\sqrt{3})^2 + 4 - 1}{17} = \frac{15}{17}.$$

Example 22.10.3 The point, $(1, \sqrt{3}, 4)$ is on the surface, $z = x^2 + y^2$. Find the line perpendicular to the surface at this point.

All that is needed is the direction vector of this line. The surface is the level surface, $x^2 + y^2 - z = 0$. The normal to this surface is given by the gradient at this point. Thus the desired line is

$$(1, \sqrt{3}, 4) + t(2, 2\sqrt{3}, -1).$$

22.11 Exercises

- Find the gradient of $f =$
 - $x^2y + z^3$ at $(1, 1, 2)$
 - $z \sin(x^2y) + 2^{x+y}$ at $(1, 1, 0)$
 - $u \ln(x + y + z^2 + w)$ at $(x, y, z, w, u) = (1, 1, 1, 1, 2)$
- Find the directional derivatives of f at the indicated point in the direction, $(\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}})$.
 - $x^2y + z^3$ at $(1, 1, 1)$
 - $z \sin(x^2y) + 2^{x+y}$ at $(1, 1, 0)$
 - $xy + z^2 + 1$ at $(1, 2, 3)$
- Find the tangent plane to the indicated level surface at the indicated point.
 - $x^2y + z^3 = 2$ at $(1, 1, 1)$
 - $z \sin(x^2y) + 2^{x+y} = 2 \sin 1 + 4$ at $(1, 1, 2)$
 - $\cos(x) + z \sin(x + y) = 1$ at $(-\pi, \frac{3\pi}{2}, 2)$
- The point $(1, 1, \sqrt{2})$ is a point on the level surface, $x^2 + y^2 + z^2 = 4$. Find the line perpendicular to the surface at this point.
- The level surfaces $x^2 + y^2 + z^2 = 4$ and $z + x^2 + y^2 = 4$ have the point $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 1)$ in the curve formed by the intersection of these surfaces. Find a direction vector for this curve at this point. **Hint:** Recall the gradients of the two surfaces are perpendicular to the corresponding surfaces at this point. A direction vector for the desired curve should be perpendicular to both of these gradients.

6. For $\mathbf{v}$ a unit vector, recall that $D_{\mathbf{v}}f(\mathbf{x}) = \nabla f(\mathbf{x}) \cdot \mathbf{v}$. It was shown above that the largest directional derivative is in the direction of the gradient and the smallest in the direction of $-\nabla f$. Establish the same result using the geometric description of the dot product, the one which says the dot product is the product of the lengths of the vectors times the cosine of the included angle.
7. The point $(1, 1, \sqrt{2})$ is a point on the level surface, $x^2 + y^2 + z^2 = 4$ and the level surface, $y^2 + 2z^2 = 5$. Find an equation for the line tangent to the curve of intersection of these two surfaces at this point.
8. *In a slightly more general setting, suppose $f_1(x, y, z) = 0$ and $f_2(x, y, z) = 0$ are two level surfaces which intersect in a curve which has parametrization, $(x(t), y(t), z(t))$. Find a system of differential equations for $(x(t), y(t), z(t))$ where as t varies, the point determined by $(x(t), y(t), z(t))$ moves over the curve.

Optimization

23.0.1 Outcomes

1. Define what is meant by a local extreme point.
2. Find candidates for local extrema using the gradient.
3. Find the local extreme values and saddle points of a C^2 function.
4. Use the second derivative test to identify the nature of a singular point.
5. Find the extreme values of a function defined on a closed and bounded region.
6. Solve word problems involving maximum and minimum values.
7. Use the method of Lagrange to determine the extreme values of a function subject to a constraint.
8. Solve word problems using the method of Lagrange multipliers.

Suppose $f : D(f) \rightarrow \mathbb{R}$ where $D(f) \subseteq \mathbb{R}^n$.

23.1 Local Extrema

Definition 23.1.1 A point $\mathbf{x} \in D(f) \subseteq \mathbb{R}^n$ is called a **local minimum** if $f(\mathbf{x}) \leq f(\mathbf{y})$ for all $\mathbf{y} \in D(f)$ sufficiently close to $\mathbf{x}$. A point $\mathbf{x} \in D(f)$ is called a **local maximum** if $f(\mathbf{x}) \geq f(\mathbf{y})$ for all $\mathbf{y} \in D(f)$ sufficiently close to $\mathbf{x}$. A **local extremum** is a point of $D(f)$ which is either a local minimum or a local maximum. The plural for extremum is extrema. The plural for minimum is **minima** and the plural for maximum is **maxima**.

Procedure 23.1.2 To find candidates for local extrema which are interior points of $D(f)$ where f is a differentiable function, you simply identify those points where ∇f equals the zero vector. To justify this, note that the graph of f is the level surface

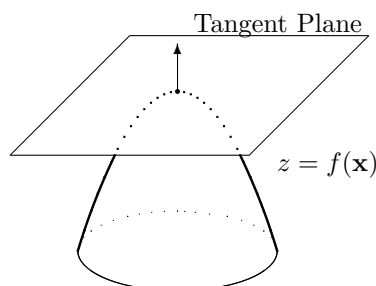
$$F(\mathbf{x}, z) \equiv z - f(\mathbf{x}) = 0$$

and the local extrema at such interior points must have horizontal tangent planes. Therefore, a normal vector at such points must be a multiple of $(0, \dots, 0, 1)$. Thus ∇F at such points must be a multiple of this vector. That is, if $\mathbf{x}$ is such a point,

$$k(0, \dots, 0, 1) = (-f_{x_1}(\mathbf{x}), \dots, -f_{x_n}(\mathbf{x}), 1).$$

Thus $\nabla f(\mathbf{x}) = \mathbf{0}$.

This is illustrated in the following picture.



A more rigorous explanation is as follows. Let $\mathbf{v}$ be any vector in $\mathbb{R}^n$ and suppose $\mathbf{x}$ is a local maximum (minimum) for $\mathbf{f}$. Then consider the real valued function of one variable, $h(t) \equiv f(\mathbf{x} + t\mathbf{v})$ for small $|t|$. Since $\mathbf{f}$ has a local maximum (minimum), it follows that h is a differentiable function of the single variable t for small t which has a local maximum (minimum) when $t = 0$. Therefore, $h'(0) = 0$. But $h'(t) = Df(\mathbf{x} + t\mathbf{v})\mathbf{v}$ by the chain rule. Therefore,

$$h'(0) = Df(\mathbf{x})\mathbf{v} = 0$$

and since $\mathbf{v}$ is arbitrary, it follows $Df(\mathbf{x}) = 0$. However,

$$Df(\mathbf{x}) = \begin{pmatrix} f_{x_1}(\mathbf{x}) & \cdots & f_{x_n}(\mathbf{x}) \end{pmatrix}$$

and so $\nabla f(\mathbf{x}) = \mathbf{0}$. This proves the following theorem.

Theorem 23.1.3 Suppose U is an open set contained in $D(f)$ such that f is differentiable on U and suppose $\mathbf{x} \in U$ is a local minimum or local maximum for f . Then $\nabla f(\mathbf{x}) = \mathbf{0}$.

Definition 23.1.4 A **singular point** for f is a point $\mathbf{x}$ where $\nabla f(\mathbf{x}) = \mathbf{0}$. This is also called a **critical point**. By analogy with the one variable case, a point where the gradient does not exist will also be called a critical point.

Example 23.1.5 Find the critical points for the function, $f(x, y) \equiv xy - x - y$ for $x, y > 0$.

Note that here $D(f)$ is an open set and so every point is an interior point. Where is the gradient equal to zero?

$$f_x = y - 1 = 0, \quad f_y = x - 1 = 0$$

and so there is exactly one critical point, $(1, 1)$.

Example 23.1.6 Find the volume of the smallest tetrahedron made up of the coordinate planes in the first octant and a plane which is tangent to the sphere $x^2 + y^2 + z^2 = 4$.

The normal to the sphere at a point, (x_0, y_0, z_0) on a point of the sphere is

$$\left(x_0, y_0, \sqrt{4 - x_0^2 - y_0^2} \right)$$

and so the equation of the tangent plane at this point is

$$x_0(x - x_0) + y_0(y - y_0) + \sqrt{4 - x_0^2 - y_0^2} \left(z - \sqrt{4 - x_0^2 - y_0^2} \right) = 0$$

When $x = y = 0$,

$$z = \frac{4}{\sqrt{(4 - x_0^2 - y_0^2)}}$$

When $z = 0 = y$,

$$x = \frac{4}{x_0},$$

and when $z = x = 0$,

$$y = \frac{4}{y_0}.$$

Therefore, the function to minimize is

$$f(x, y) = \frac{1}{6} \frac{64}{xy\sqrt{(4 - x^2 - y^2)}}$$

This is because in beginning calculus it was shown that the volume of a pyramid is $1/3$ the area of the base times the height. Therefore, you simply need to find the gradient of this and set it equal to zero. Thus upon taking the partial derivatives, you need to have

$$\frac{-4 + 2x^2 + y^2}{x^2y(-4 + x^2 + y^2)\sqrt{(4 - x^2 - y^2)}} = 0,$$

and

$$\frac{-4 + x^2 + 2y^2}{xy^2(-4 + x^2 + y^2)\sqrt{(4 - x^2 - y^2)}} = 0.$$

Therefore, $x^2 + 2y^2 = 4$ and $2x^2 + y^2 = 4$. Thus $x = y$ and so $x = y = \frac{2}{\sqrt{3}}$. It follows from the equation for z that $z = \frac{2}{\sqrt{3}}$ also. How do you know this is not the largest tetrahedron.

Example 23.1.7 *An open box is to contain 32 cubic feet. Find the dimensions which will result in the least surface area.*

Let the height of the box be z and the length and width be x and y respectively. Then $xyz = 32$ and so $z = 32/xy$. The total area is $xy + 2xz + 2yz$ and so in terms of the two variables, x and y , the area is

$$A = xy + \frac{64}{y} + \frac{64}{x}$$

To find best dimensions you note these must result in a local minimum.

$$A_x = \frac{yx^2 - 64}{x^2} = 0, \quad A_y = \frac{xy^2 - 64}{y^2}.$$

Therefore, $yx^2 - 64 = 0$ and $xy^2 - 64 = 0$ so $xy^2 = yx^2$. For sure the answer excludes the case where any of the variables equals zero. Therefore, $x = y$ and so $x = 4 = y$. Then $z = 2$ from the requirement that $xyz = 32$. How do you know this gives the least surface area? Why doesn't this give the largest surface area?

23.2 Exercises

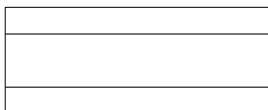
1. Find the points where possible local minima or local maxima occur in the following functions.

- (a) $x^2 - 2x + 5 + y^2 - 4y$
- (b) $-xy + y^2 - y + x$
- (c) $3x^2 - 4xy + 2y^2 - 2y + 2x$
- (d) $\cos(x) + \sin(2y)$
- (e) $x^4 - 4x^3y + 6x^2y^2 - 4xy^3 + y^4 + x^2 - 2x$
- (f) $y^2x^2 - 2xy^2 + y^2$

2. Find the volume of the largest box which can be inscribed in a sphere of radius a .
3. Find in terms of a, b, c the volume of the largest box which can be inscribed in the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

4. Find three numbers which add to 36 whose product is as large as possible.
5. Find three numbers x, y, z such that $x^2 + y^2 + z^2 = 1$ and $x + y + z$ is as large as possible.
6. Find three numbers x, y, z such that $x^2 + y^2 + z^2 = 4$ and xyz is as large as possible.
7. A feeding trough in the form of a trapezoid with equal base angles is made from a long rectangular piece of metal of width 24 inches by bending up equal strips along both sides. Find the base angles and the width of these strips which will maximize the volume of the feeding trough.



8. An open box (no top) is to contain 40 cubic feet. The material for the bottom costs twice as much as the material for the sides. Find the dimensions of the box which is cheapest.
9. The function $f(x, y) = 2x^2 + y^2$ is defined on the disk $x^2 + y^2 \leq 1$. Find its maximum value.
10. Find the point on the surface $z = x^2 + y + 1$ which is closest to $(0, 0, 0)$.
11. Let $L_1 = (t, 2t, 3 - t)$ and $L_2 = (2s, s + 2, 4 - s)$ be two lines. Find a pair of points, one on the first line and the other on the second such that these two points are closer together than any other pair of points on the two lines.
12. *Let

$$f(x, y) = \begin{cases} -1 & \text{if } y = x^2, x \neq 0 \\ (y - x^2)^2 & \text{if } y \neq x^2 \\ 0 & \text{if } (x, y) = (0, 0) \end{cases}$$

Show that $\nabla f(0, 0) = \mathbf{0}$. Now show that if (a, b) is any nonzero unit vector, the function $t \rightarrow f(ta, tb)$ has a local minimum of 0 when $t = 0$. Thus in every direction, this function has a local minimum at $(0, 0)$ but the function f does not have a local minimum at $(0, 0)$.

23.3 The Second Derivative Test

There is a version of the second derivative test in the case that the function and its first and second partial derivatives are all continuous. The proof of this theorem is dependent on fundamental results in linear algebra which are in an appendix. You can skip the proof if you like. It is given later.

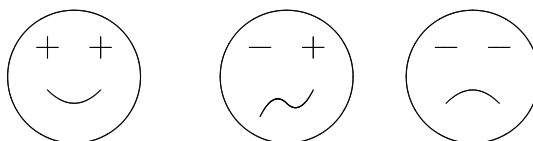
Definition 23.3.1 The matrix, $H(\mathbf{x})$ whose ij^{th} entry at the point $\mathbf{x}$ is

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})$$

is called the **Hessian matrix**. The eigenvalues of $H(\mathbf{x})$ are the solutions λ to the equation

$$\det(\lambda I - H(\mathbf{x})) = 0$$

The following theorem says that if all the eigenvalues of the Hessian matrix at a critical point are positive, then the critical point is a local minimum. If all the eigenvalues of the Hessian matrix at a critical point are negative, then the critical point is a local maximum. Finally, if some of the eigenvalues of the Hessian matrix at the critical point are positive and some are negative then the critical point is a saddle point. The following picture illustrates the situation.



Theorem 23.3.2 Let $f : U \rightarrow \mathbb{R}$ for U an open set in $\mathbb{R}^n$ and let f be a C^2 function and suppose that at some $\mathbf{x} \in U$, $\nabla f(\mathbf{x}) = \mathbf{0}$. Also let μ and λ be respectively, the largest and smallest eigenvalues of the matrix, $H(\mathbf{x})$. If $\lambda > 0$ then f has a local minimum at $\mathbf{x}$. If $\mu < 0$ then f has a local maximum at $\mathbf{x}$. If either λ or μ equals zero, the test fails. If $\lambda < 0$ and $\mu > 0$ there exists a direction in which when f is evaluated on the line through the critical point having this direction, the resulting function of one variable has a local minimum and there exists a direction in which when f is evaluated on the line through the critical point having this direction, the resulting function of one variable has a local maximum. This last case is called a **saddle point**.

Here is an example.

Example 23.3.3 Let $f(x, y) = 10xy + y^2$. Find the critical points and determine whether they are local minima, local maxima or saddle points.

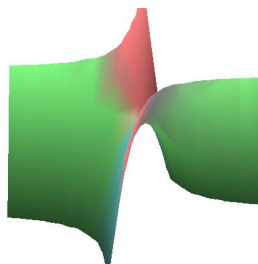
First

$$\nabla(10xy + y^2) = (10y, 10x + 2y)$$

and so there is one critical point at the point $(0, 0)$. What is it? The Hessian matrix is

$$\begin{pmatrix} 0 & 10 \\ 10 & 2 \end{pmatrix}$$

and the eigenvalues are of different signs. Therefore, the critical point $(0, 0)$ is a saddle point. Here is a graph drawn by Maple.



Here is another example.

Example 23.3.4 Let $f(x, y) = 2x^4 - 4x^3 + 14x^2 + 12yx^2 - 12yx - 12x + 2y^2 + 4y + 2$. Find the critical points and determine whether they are local minima, local maxima, or saddle points.

$f_x(x, y) = 8x^3 - 12x^2 + 28x + 24yx - 12y - 12$ and $f_y(x, y) = 12x^2 - 12x + 4y + 4$. The points at which both f_x and f_y equal zero are $(\frac{1}{2}, -\frac{1}{4})$, $(0, -1)$, and $(1, -1)$.

The Hessian matrix is

$$\begin{pmatrix} 24x^2 + 28 + 24y - 24x & 24x - 12 \\ 24x - 12 & 4 \end{pmatrix}.$$

and the thing to determine is the sign of its eigenvalues evaluated at the critical points.

First consider the point $(\frac{1}{2}, -\frac{1}{4})$. This matrix is $\begin{pmatrix} 16 & 0 \\ 0 & 4 \end{pmatrix}$ and its eigenvalues are 16, 4 showing that this is a local minimum.

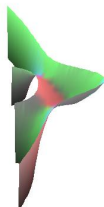
Next consider $(0, -1)$ at this point the Hessian matrix is $\begin{pmatrix} 4 & -12 \\ -12 & 4 \end{pmatrix}$ and the eigenvalues are 16, -8 . Therefore, this point is a saddle point. To determine this, find the eigenvalues.

$$\begin{aligned} & \det\left(\lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 4 & -12 \\ -12 & 4 \end{pmatrix}\right) \\ &= \lambda^2 - 8\lambda - 128 = (\lambda + 8)(\lambda - 16) \end{aligned}$$

so the eigenvalues are -8 and 16 as claimed.

Finally consider the point $(1, -1)$. At this point the Hessian is $\begin{pmatrix} 4 & 12 \\ 12 & 4 \end{pmatrix}$ and the eigenvalues are 16, -8 so this point is also a saddle point.

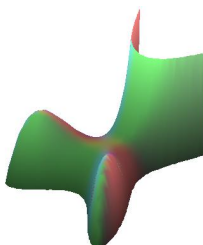
Below is a graph of this function which illustrates the behavior near saddle points.



Of course sometimes the second derivative test is inadequate to determine what is going on. This should be no surprise since this was the case even for a function of one variable. For a function of two variables, a nice example is the Monkey saddle.

Example 23.3.5 Suppose $f(x, y) = 6xy^2 - 2x^3 - 3y^4$. Show that $(0, 0)$ is a critical point for which the second derivative test gives no information.

Before doing anything it might be interesting to look at the graph of this function of two variables plotted using Maple.



This picture should indicate why this is called a monkey saddle. It is because the monkey can sit in the saddle and have a place for his tail. Now to see $(0, 0)$ is a critical point, note that

$$f(x, y) = 6xy^2 - 2x^3 - 3y^4$$

and so $f_x(0, 0) = f_y(0, 0) = 0$.

$$f_x(x, y) = 6y^2 - 6x^2, \quad f_y(x, y) = 12xy - 12y^3$$

and clearly $(0, 0)$ is a critical point. So are $(1, 1)$ and $(1, -1)$. Now $f_{xx}(0, 0) = 0$ and so are $f_{xy}(0, 0)$ and $f_{yy}(0, 0)$. Therefore, the Hessian matrix is the zero matrix and clearly has only the zero eigenvalue. Therefore, the second derivative test is totally useless at this point.

However, suppose you took $x = t$ and $y = t$ and evaluated this function on this line. This reduces to $h(t) = f(t, t) = 4t^3 - 3t^4$, which is strictly increasing near $t = 0$. This shows the critical point, $(0, 0)$ of f is neither a local max. nor a local min. Next let $x = 0$ and $y = t$. Then $p(t) \equiv f(0, t) = -3t^4$. Therefore, along the line, $(0, t)$, f has a local maximum at $(0, 0)$.

Example 23.3.6 Find the critical points of the following function of three variables and classify them as local minimums, local maximums or saddle points.

$$f(x, y, z) = \frac{5}{6}x^2 + 4x + 16 - \frac{7}{3}xy - 4y - \frac{4}{3}xz + 12z + \frac{5}{6}y^2 - \frac{4}{3}zy + \frac{1}{3}z^2$$

First you need to locate the critical points. This involves taking the gradient.

$$\begin{aligned} & \nabla \left(\frac{5}{6}x^2 + 4x + 16 - \frac{7}{3}xy - 4y - \frac{4}{3}xz + 12z + \frac{5}{6}y^2 - \frac{4}{3}zy + \frac{1}{3}z^2 \right) \\ &= \left(\frac{5}{3}x + 4 - \frac{7}{3}y - \frac{4}{3}z, -\frac{7}{3}x - 4 + \frac{5}{3}y - \frac{4}{3}z, -\frac{4}{3}x + 12 - \frac{4}{3}y + \frac{2}{3}z \right) \end{aligned}$$

Next you need to set the gradient equal to zero and solve the equations. This yields $y = 5, x = 3, z = -2$. Now to use the second derivative test, you assemble the Hessian matrix which is

$$\begin{pmatrix} \frac{5}{3} & -\frac{7}{3} & -\frac{4}{3} \\ -\frac{7}{3} & \frac{5}{3} & -\frac{4}{3} \\ -\frac{4}{3} & -\frac{4}{3} & \frac{2}{3} \end{pmatrix}.$$

Note that in this simple example, the Hessian matrix is constant and so all that is left is to consider the eigenvalues. Writing the characteristic equation and solving yields the eigenvalues are $2, -2, 4$. Thus the given point is a saddle point.

23.4 Exercises

1. Use the second derivative test on the critical points $(1, 1)$, and $(1, -1)$ for Example 23.3.5.

The function is $\arctan(6xy^2 - 2x^3 - 3x^4)$.

2. If $H = H^T$ and $H\mathbf{x} = \lambda\mathbf{x}$ while $H\mathbf{x} = \mu\mathbf{x}$ for $\lambda \neq \mu$, show that $\mathbf{x} \cdot \mathbf{y} = 0$.
3. Show that the points $(\frac{1}{2}, -\frac{21}{4})$, $(0, -4)$, and $(1, -4)$ are critical points of the following function of two variables and classify them as local minima, local maxima or saddle points.
4. Show that the points $(\frac{1}{2}, -\frac{53}{12})$, $(0, -4)$, and $(1, -4)$ are critical points of the following function of two variables and classify them according to whether they are local minima, local maxima or saddle points.
5. Show that the points $(\frac{1}{2}, \frac{37}{20})$, $(0, 2)$, and $(1, 2)$ are critical points of the following function of two variables and classify them according to whether they are local minima, local maxima or saddle points.
6. Show that the points $(\frac{1}{2}, -\frac{17}{8})$, $(0, -2)$, and $(1, -2)$ are critical points of the following function of two variables and classify them according to whether they are local minima, local maxima or saddle points.
7. Find the critical points of the following function of three variables and classify them according to whether they are local minima, local maxima or saddle points.
8. Find the critical points of the following function of three variables and classify them according to whether they are local minima, local maxima or saddle points.
9. Find the critical points of the following function of three variables and classify them according to whether they are local minima, local maxima or saddle points.
10. Find the critical points of the following function of three variables and classify them according to whether they are local minima, local maxima or saddle points.
11. *Show that if f has a critical point and some eigenvalue of the Hessian matrix is positive, then there exists a direction in which when f is evaluated on the line through the critical point having this direction, the resulting function of one variable has a local minimum. State and prove a similar result in the case where some eigenvalue of the Hessian matrix is negative.
12. Suppose $\mu = 0$ but there are negative eigenvalues of the Hessian at a critical point. Show by giving examples that the second derivative tests fails.
13. Show that the points $(\frac{1}{2}, -\frac{9}{2})$, $(0, -5)$, and $(1, -5)$ are critical points of the following function of two variables and classify them as local minima, local maxima or saddle points.

$$f(x, y) = 2x^4 - 4x^3 + 42x^2 + 8yx^2 - 8yx - 40x + 2y^2 + 20y + 50.$$
14. Show that the points $(1, -\frac{11}{2})$, $(0, -5)$, and $(2, -5)$ are critical points of the following function of two variables and classify them as local minima, local maxima or saddle points.

$$f(x, y) = 4x^4 - 16x^3 - 4x^2 - 4yx^2 + 8yx + 40x + 4y^2 + 40y + 100.$$

15. Show that the points $(\frac{3}{2}, \frac{27}{20})$, $(0, 0)$, and $(3, 0)$ are critical points of the following function of two variables and classify them as local minima, local maxima or saddle points.

$$f(x, y) = 5x^4 - 30x^3 + 45x^2 + 6yx^2 - 18yx + 5y^2.$$

16. Find the critical points of the following function of three variables and classify them as local minima, local maxima or saddle points.

$$f(x, y, z) = \frac{10}{3}x^2 - \frac{44}{3}x + \frac{64}{3} - \frac{10}{3}yx + \frac{16}{3}y + \frac{2}{3}zx - \frac{20}{3}z + \frac{10}{3}y^2 + \frac{2}{3}zy + \frac{4}{3}z^2.$$

17. Find the critical points of the following function of three variables and classify them as local minima, local maxima or saddle points.

$$f(x, y, z) = -\frac{7}{3}x^2 - \frac{146}{3}x + \frac{83}{3} + \frac{16}{3}yx + \frac{4}{3}y - \frac{14}{3}zx + \frac{94}{3}z - \frac{7}{3}y^2 - \frac{14}{3}zy + \frac{8}{3}z^2.$$

18. Find the critical points of the following function of three variables and classify them as local minima, local maxima or saddle points.

$$f(x, y, z) = \frac{2}{3}x^2 + 4x + 75 - \frac{14}{3}yx - 38y - \frac{8}{3}zx - 2z + \frac{2}{3}y^2 - \frac{8}{3}zy - \frac{1}{3}z^2.$$

19. Find the critical points of the following function of three variables and classify them as local minima, local maxima or saddle points.

$$f(x, y, z) = 4x^2 - 30x + 510 - 2yx + 60y - 2zx - 70z + 4y^2 - 2zy + 4z^2.$$

20. Show that the critical points of the following function are points of the form, $(x, y, z) = (t, 2t^2 - 10t, -t^2 + 5t)$ for $t \in \mathbb{R}$ and classify them as local minima, local maxima or saddle points.

$$f(x, y, z) = -\frac{1}{6}x^4 + \frac{5}{3}x^3 - \frac{25}{6}x^2 + \frac{10}{3}yx^2 - \frac{50}{3}yx + \frac{19}{3}zx^2 - \frac{95}{3}zx - \frac{5}{3}y^2 - \frac{10}{3}zy - \frac{1}{6}z^2.$$

21. Show that the critical points of the following function are

$$(0, -3, 0), (2, -3, 0), \text{ and } \left(1, -3, -\frac{1}{3}\right)$$

and classify them as local minima, local maxima or saddle points.

$$f(x, y, z) = -\frac{3}{2}x^4 + 6x^3 - 6x^2 + zx^2 - 2zx - 2y^2 - 12y - 18 - \frac{3}{2}z^2.$$

22. Show that the critical points of the function,

$$f(x, y, z) = -2yx^2 - 6yx - 4zx^2 - 12zx + y^2 + 2yz.$$

are points of the form,

$$(x, y, z) = (t, 2t^2 + 6t, -t^2 - 3t)$$

for $t \in \mathbb{R}$ and classify them as local minima, local maxima or saddle points.

23. Show that the critical points of the function

$$f(x, y, z) = \frac{1}{2}x^4 - 4x^3 + 8x^2 - 3zx^2 + 12zx + 2y^2 + 4y + 2 + \frac{1}{2}z^2.$$

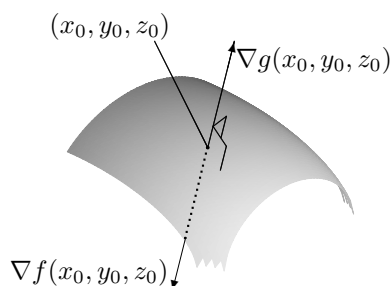
are $(0, -1, 0)$, $(4, -1, 0)$, and $(2, -1, -12)$ and classify them as local minima, local maxima or saddle points.

24. Suppose $f(x, y)$, a function of two variables defined on all $\mathbb{R}^n$ has all directional derivatives at $(0, 0)$ and they are all equal to 0 there. Suppose also that for $h(t) \equiv f(tu, tv)$ and (u, v) a unit vector, it follows that $h''(0) > 0$. By the one variable second derivative test, this implies that along every straight line through $(0, 0)$ the function restricted to this line has a local minimum at $(0, 0)$. Can it be concluded that f has a local minimum at $(0, 0)$. In other words, can you conclude a point is a local minimum if it appears to be so along every straight line through the point? **Hint:** Consider $f(x, y) = x^2 + y^2$ for (x, y) not on the curve $y = x^2$ for $x \neq 0$ and on this curve, let $f = -1$.

23.5 Lagrange Multipliers

Lagrange multipliers are used to solve extremum problems for a function defined on a level set of another function. For example, suppose you want to maximize xy given that $x + y = 4$. This is not too hard to do using methods developed earlier. Solve for one of the variables, say y , in the constraint equation, $x + y = 4$ to find $y = 4 - x$. Then the function to maximize is $f(x) = x(4 - x)$ and the answer is clearly $x = 2$. Thus the two numbers are $x = y = 2$. This was easy because you could easily solve the constraint equation for one of the variables in terms of the other. Now what if you wanted to maximize $f(x, y, z) = xyz$ subject to the constraint that $x^2 + y^2 + z^2 = 4$? It is still possible to do this using similar techniques. Solve for one of the variables in the constraint equation, say z , substitute it into f , and then find where the partial derivatives equal zero to find candidates for the extremum. However, it seems you might encounter many cases and it does look a little fussy. However, sometimes you can't solve the constraint equation for one variable in terms of the others. Also, what if you had many constraints. What if you wanted to maximize $f(x, y, z)$ subject to the constraints $x^2 + y^2 = 4$ and $z = 2x + 3y^2$. Things are clearly getting more involved and messy. It turns out that at an extremum, there is a simple relationship between the gradient of the function to be maximized and the gradient of the constraint function.

This relation can be seen geometrically in the following picture.



In the picture, the surface represents a piece of the level surface of $g(x, y, z) = 0$ and $f(x, y, z)$ is the function of three variables which is being maximized or minimized on the level surface and suppose the extremum of f occurs at the point (x_0, y_0, z_0) . As shown above, $\nabla g(x_0, y_0, z_0)$ is perpendicular to the surface or more precisely to the tangent plane. However, if $\mathbf{x}(t) = (x(t), y(t), z(t))$ is a point on a smooth curve which passes through (x_0, y_0, z_0) when $t = t_0$, then the function, $h(t) = f(x(t), y(t), z(t))$ must have either a maximum or a minimum at the point, $t = t_0$. Therefore, $h'(t_0) = 0$. But this means

$$\begin{aligned} 0 &= h'(t_0) = \nabla f(x(t_0), y(t_0), z(t_0)) \cdot \mathbf{x}'(t_0) \\ &= \nabla f(x_0, y_0, z_0) \cdot \mathbf{x}'(t_0) \end{aligned}$$

and since this holds for any such smooth curve, $\nabla f(x_0, y_0, z_0)$ is also perpendicular to the surface. This picture represents a situation in three dimensions and you can see that it is intuitively clear that this implies $\nabla f(x_0, y_0, z_0)$ is some scalar multiple of $\nabla g(x_0, y_0, z_0)$. Thus

$$\nabla f(x_0, y_0, z_0) = \lambda \nabla g(x_0, y_0, z_0)$$

This λ is called a **Lagrange multiplier** after Lagrange who considered such problems in the 1700's.

Of course the above argument is at best only heuristic. It does not deal with the question of existence of smooth curves lying in the constraint surface passing through (x_0, y_0, z_0) . Nor does it consider all cases, being essentially confined to three dimensions. In addition to this, it fails to consider the situation in which there are many constraints. However, I think it is likely a geometric notion like that presented above which led Lagrange to formulate the method.

Example 23.5.1 Maximize xyz subject to $x^2 + y^2 + z^2 = 27$.

Here $f(x, y, z) = xyz$ while $g(x, y, z) = x^2 + y^2 + z^2 - 27$. Then $\nabla g(x, y, z) = (2x, 2y, 2z)$ and $\nabla f(x, y, z) = (yz, xz, xy)$. Then at the point which maximizes this function¹,

$$(yz, xz, xy) = \lambda(2x, 2y, 2z).$$

Therefore, each of $2\lambda x^2, 2\lambda y^2, 2\lambda z^2$ equals xyz . It follows that at any point which maximizes xyz , $|x| = |y| = |z|$. Therefore, the only candidates for the point where the maximum occurs are $(3, 3, 3), (-3, -3, 3), (-3, 3, 3)$, etc. The maximum occurs at $(3, 3, 3)$ which can be verified by plugging in to the function which is being maximized.

The method of Lagrange multipliers allows you to consider maximization of functions defined on closed and bounded sets. Recall that any continuous function defined on a closed and bounded set has a maximum and a minimum on the set. Candidates for the extremum on the interior of the set can be located by setting the gradient equal to zero. The consideration of the boundary can then sometimes be handled with the method of Lagrange multipliers.

Example 23.5.2 Maximize $f(x, y) = xy + y$ subject to the constraint, $x^2 + y^2 \leq 1$.

Here I know there is a maximum because the set is the closed circle, a closed and bounded set. Therefore, it is just a matter of finding it. Look for singular points on the interior of the circle. $\nabla f(x, y) = (y, x + 1) = (0, 0)$. There are no points on the interior of the circle where the gradient equals zero. Therefore, the maximum occurs on the boundary of the circle. That is the problem reduces to maximizing $xy + y$ subject to $x^2 + y^2 = 1$. From the above,

$$(y, x + 1) - \lambda(2x, 2y) = 0.$$

Hence $y^2 - 2\lambda xy = 0$ and $x(x + 1) - 2\lambda xy = 0$ so $y^2 = x(x + 1)$. Therefore from the constraint, $x^2 + x(x + 1) = 1$ and the solution is $x = -1, x = \frac{1}{2}$. Then the candidates for a solution are $(-1, 0), (\frac{1}{2}, \frac{\sqrt{3}}{2}), (\frac{1}{2}, -\frac{\sqrt{3}}{2})$. Then

$$f(-1, 0) = 0, f\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = \frac{3\sqrt{3}}{4}, f\left(\frac{1}{2}, -\frac{\sqrt{3}}{2}\right) = -\frac{3\sqrt{3}}{4}.$$

It follows the maximum value of this function is $\frac{3\sqrt{3}}{4}$ and it occurs at $(\frac{1}{2}, \frac{\sqrt{3}}{2})$. The minimum value is $-\frac{3\sqrt{3}}{4}$ and it occurs at $(\frac{1}{2}, -\frac{\sqrt{3}}{2})$.

¹There exists such a point because the sphere is closed and bounded.

Example 23.5.3 Find candidates for the maximum and minimum values of the function, $f(x, y) = xy - x^2$ on the set, $\{(x, y) : x^2 + 2xy + y^2 \leq 4\}$.

First, the only point where ∇f equals zero is $(x, y) = (0, 0)$ and this is in the desired set. In fact it is an interior point of this set. This takes care of the interior points. What about those on the boundary $x^2 + 2xy + y^2 = 4$? The problem is to maximize $xy - x^2$ subject to the constraint, $x^2 + 2xy + y^2 = 4$. The Lagrangian is $xy - x^2 - \lambda(x^2 + 2xy + y^2 - 4)$ and this yields the following system.

$$\begin{aligned}y - 2x - \lambda(2x + 2y) &= 0 \\x - 2\lambda(x + y) &= 0 \\x^2 + 2xy + y^2 &= 4\end{aligned}$$

From the first two equations,

$$\begin{aligned}(2 + 2\lambda)x - (1 - 2\lambda)y &= 0 \\(1 - 2\lambda)x - 2\lambda y &= 0\end{aligned}$$

Since not both x and y equal zero, it follows

$$\det \begin{pmatrix} 2 + 2\lambda & 2\lambda - 1 \\ 1 - 2\lambda & -2\lambda \end{pmatrix} = 0$$

which yields

$$\lambda = 1/8$$

Therefore,

$$y = 3x \tag{23.1}$$

From the constraint equation,

$$x^2 + 2x(3x) + (3x)^2 = 4$$

and so

$$x = \frac{1}{2} \text{ or } -\frac{1}{2}$$

Now from 23.1, the points of interest on the boundary of this set are

$$\left(\frac{1}{2}, \frac{3}{2}\right), \text{ and } \left(-\frac{1}{2}, -\frac{3}{2}\right). \tag{23.2}$$

$$\begin{aligned}f\left(\frac{1}{2}, \frac{3}{2}\right) &= \left(\frac{1}{2}\right)\left(\frac{3}{2}\right) - \left(\frac{1}{2}\right)^2 \\&= \frac{1}{2}\end{aligned}$$

$$\begin{aligned}f\left(-\frac{1}{2}, -\frac{3}{2}\right) &= \left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right) - \left(-\frac{1}{2}\right)^2 \\&= \frac{1}{2}\end{aligned}$$

It follows the candidates for maximum and minimum are $\left(\frac{1}{2}, \frac{3}{2}\right)$, $(0, 0)$, and $\left(-\frac{1}{2}, -\frac{3}{2}\right)$. Therefore it appears that $(0, 0)$ yields a minimum and either $\left(\frac{1}{2}, \frac{3}{2}\right)$ or $\left(-\frac{1}{2}, -\frac{3}{2}\right)$ yields a maximum. However, this is a little misleading. How do you even know a maximum or a minimum exists? The set, $x^2 + 2xy + y^2 \leq 4$ is an unbounded set which lies between

the two lines $x + y = 2$ and $x + y = -2$. In fact there is no minimum. For example, take $x = 100, y = -98$. Then $xy - x^2 = x(y - x) = 100(-98 - 100)$ which is a large negative number much less than 0, the answer for the point $(0, 0)$.

There are no magic bullets here. It was still required to solve a system of nonlinear equations to get the answer. However, it does often help to do it this way.

A nice observation in the case that the function f , which you are trying to maximize, and the function g , which defines the constraint, are functions of two or three variables is the following.

At points of interest,

$$\nabla f \times \nabla g = \mathbf{0}$$

This follows from the above because at these points,

$$\nabla f = \lambda \nabla g$$

so the angle between the two vectors ∇f and ∇g is either 0 or π . Therefore, the sine of this angle equals 0. By the geometric description of the cross product, this implies the cross product equals 0. Here is an example.

Example 23.5.4 Minimize $f(x, y) = xy - x^2$ on the set, $\{(x, y) : x^2 + 2xy + y^2 = 4\}$

Using the observation about the cross product, and letting $f(x, y, z) = f(x, y)$ with a similar convention for g , $\nabla f = (y - 2x, x, 0)$, $\nabla g = (2x + 2y, 2x + 2y, 0)$ and so

$$\begin{aligned} & (y - 2x, x, 0) \times (2x + 2y, 2x + 2y, 0) \\ &= (0, 0, (y - 2x)(2x + 2y) - x(2x + 2y)) = 0 \end{aligned}$$

Thus there are two equations, $x^2 + 2xy + y^2 = 4$ and $4xy - 2y^2 + 6x^2 = 0$. Solving these two yields the points of interest $(-\frac{1}{2}, -\frac{3}{2}), (\frac{1}{2}, \frac{3}{2})$. Now one of these gives the minimum and the other gives the maximum.

The above generalizes to a general procedure which is described in the following major Theorem. All correct proofs of this theorem will involve some appeal to the implicit function theorem or to fundamental existence theorems from differential equations. A complete proof is very fascinating but it will not come cheap. Good advanced calculus books will usually give a correct proof. If you are interested, there is a complete proof in an appendix to this book. First here is a simple definition explaining one of the terms in the statement of this theorem.

Definition 23.5.5 Let A be an $m \times n$ matrix. A submatrix is any matrix which can be obtained from A by deleting some rows and some columns.

Theorem 23.5.6 Let U be an open subset of $\mathbb{R}^n$ and let $f : U \rightarrow \mathbb{R}$ be a C^1 function. Then if $\mathbf{x}_0 \in U$ is either a local maximum or local minimum of f subject to the constraints

$$g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \quad (23.3)$$

and if some $m \times m$ submatrix of

$$D\mathbf{g}(\mathbf{x}_0) \equiv \begin{pmatrix} g_{1x_1}(\mathbf{x}_0) & g_{1x_2}(\mathbf{x}_0) & \cdots & g_{1x_n}(\mathbf{x}_0) \\ \vdots & \vdots & & \vdots \\ g_{mx_1}(\mathbf{x}_0) & g_{mx_2}(\mathbf{x}_0) & \cdots & g_{mx_n}(\mathbf{x}_0) \end{pmatrix}$$

has nonzero determinant, then there exist scalars, $\lambda_1, \dots, \lambda_m$ such that

$$\begin{pmatrix} f_{x_1}(\mathbf{x}_0) \\ \vdots \\ f_{x_n}(\mathbf{x}_0) \end{pmatrix} = \lambda_1 \begin{pmatrix} g_{1x_1}(\mathbf{x}_0) \\ \vdots \\ g_{1x_n}(\mathbf{x}_0) \end{pmatrix} + \cdots + \lambda_m \begin{pmatrix} g_{mx_1}(\mathbf{x}_0) \\ \vdots \\ g_{mx_n}(\mathbf{x}_0) \end{pmatrix} \quad (23.4)$$

holds.

To help remember how to use 23.4 it may be helpful to do the following. First write the Lagrangian,

$$L = f(\mathbf{x}) - \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

and then proceed to take derivatives with respect to each of the components of $\mathbf{x}$ and also derivatives with respect to each λ_i and set all of these equations equal to 0. The formula 23.4 is what results from taking the derivatives of L with respect to the components of $\mathbf{x}$. When you take the derivatives with respect to the Lagrange multipliers, and set what results equal to 0, you just pick up the constraint equations. This yields $n + m$ equations for the $n + m$ unknowns, $x_1, \dots, x_n, \lambda_1, \dots, \lambda_m$. Then you proceed to look for solutions to these equations. Of course these might be impossible to find using methods of algebra, but you just do your best and hope it will work out.

Example 23.5.7 Minimize xyz subject to the constraints $x^2 + y^2 + z^2 = 4$ and $x - 2y = 0$.

Form the Lagrangian,

$$L = xyz - \lambda(x^2 + y^2 + z^2 - 4) - \mu(x - 2y)$$

and proceed to take derivatives with respect to every possible variable, leading to the following system of equations.

$$\begin{aligned} yz - 2\lambda x - \mu &= 0 \\ xz - 2\lambda y + 2\mu &= 0 \\ xy - 2\lambda z &= 0 \\ x^2 + y^2 + z^2 &= 4 \\ x - 2y &= 0 \end{aligned}$$

Now you have to find the solutions to this system of equations. In general, this could be very hard or even impossible. If $\lambda = 0$, then from the third equation, either x or y must equal 0. Therefore, from the first two equations, $\mu = 0$ also. If $\mu = 0$ and $\lambda \neq 0$, then from the first two equations, $xyz = 2\lambda x^2$ and $xyz = 2\lambda y^2$ and so either $x = y$ or $x = -y$, which requires that both x and y equal zero thanks to the last equation. But then from the fourth equation, $z = \pm 2$ and now this contradicts the third equation. Thus μ and λ are either both equal to zero or neither one is and the expression, xyz equals zero in this case. However, I know this is not the best value for a minimizer because I can take $x = 2\sqrt{\frac{3}{5}}, y = \sqrt{\frac{3}{5}}$, and $z = -1$. This satisfies the constraints and the product of these numbers equals a negative number. Therefore, both μ and λ must be non zero. Now use the last equation eliminate x and write the following system.

$$\begin{aligned} 5y^2 + z^2 &= 4 \\ y^2 - \lambda z &= 0 \\ yz - \lambda y + \mu &= 0 \\ yz - 4\lambda y - \mu &= 0 \end{aligned}$$

From the last equation, $\mu = (yz - 4\lambda y)$. Substitute this into the third and get

$$\begin{aligned} 5y^2 + z^2 &= 4 \\ y^2 - \lambda z &= 0 \\ yz - \lambda y + yz - 4\lambda y &= 0 \end{aligned}$$

$y = 0$ will not yield the minimum value from the above example. Therefore, divide the last equation by y and solve for λ to get $\lambda = (2/5)z$. Now put this in the second equation to conclude

$$\begin{aligned} 5y^2 + z^2 &= 4 \\ y^2 - (2/5)z^2 &= 0 \end{aligned}$$

a system which is easy to solve. Thus $y^2 = 8/15$ and $z^2 = 4/3$. Therefore, candidates for minima are $\left(2\sqrt{\frac{8}{15}}, \sqrt{\frac{8}{15}}, \pm\sqrt{\frac{4}{3}}\right)$, and $\left(-2\sqrt{\frac{8}{15}}, -\sqrt{\frac{8}{15}}, \pm\sqrt{\frac{4}{3}}\right)$, a choice of 4 points to check. Clearly the one which gives the smallest value is

$$\left(2\sqrt{\frac{8}{15}}, \sqrt{\frac{8}{15}}, -\sqrt{\frac{4}{3}}\right)$$

or $\left(-2\sqrt{\frac{8}{15}}, -\sqrt{\frac{8}{15}}, -\sqrt{\frac{4}{3}}\right)$ and the minimum value of the function subject to the constraints is $-\frac{2}{5}\sqrt{30} - \frac{2}{3}\sqrt{3}$.

You should rework this problem first solving the second easy constraint for x and then producing a simpler problem involving only the variables y and z .

23.6 Exercises

1. Maximize $x + y + z$ subject to the constraint $x^2 + y^2 + z^2 = 3$.
2. Minimize $2x - y + z$ subject to the constraint $2x^2 + y^2 + z^2 = 36$.
3. Minimize $x + 3y - z$ subject to the constraint $2x^2 + y^2 - 2z^2 = 36$ if possible. Note there is no guaranty this function has either a maximum or a minimum. Determine whether there exists a minimum also.
4. Find the dimensions of the largest rectangle which can be inscribed in a circle of radius r .
5. Maximize $2x + y$ subject to the condition that $\frac{x^2}{4} + \frac{y^2}{9} \leq 1$.
6. Maximize $x + 2y$ subject to the condition that $x^2 + \frac{y^2}{9} \leq 1$.
7. Maximize $x + y$ subject to the condition that $x^2 + \frac{y^2}{9} + z^2 \leq 1$.
8. Minimize $x + y + z$ subject to the condition that $x^2 + \frac{y^2}{9} + z^2 \leq 1$.
9. Find the points on $y^2x = 16$ which are closest to $(0, 0)$.
10. Find the points on $\sqrt{2}y^2x = 1$ which are closest to $(0, 0)$.
11. Find points on $xy = 4$ farthest from $(0, 0)$ if any exist. If none exist, tell why. What does this say about the method of Lagrange multipliers?
12. A can is supposed to have a volume of 36π cubic centimeters. Find the dimensions of the can which minimizes the surface area.
13. A can is supposed to have a volume of 36π cubic centimeters. The top and bottom of the can are made of tin costing 4 cents per square centimeter and the sides of the can are made of aluminum costing 5 cents per square centimeter. Find the dimensions of the can which minimizes the cost.
14. Minimize and maximize $\sum_{j=1}^n x_j$ subject to the constraint $\sum_{j=1}^n x_j^2 = a^2$. Your answer should be some function of a which you may assume is a positive number.
15. Find the point, (x, y, z) on the level surface, $4x^2 + y^2 - z^2 = 1$ which is closest to $(0, 0, 0)$.
16. A curve is formed from the intersection of the plane, $2x + y + z = 3$ and the cylinder $x^2 + y^2 = 4$. Find the point on this curve which is closest to $(0, 0, 0)$.

17. A curve is formed from the intersection of the plane, $2x + 3y + z = 3$ and the sphere $x^2 + y^2 + z^2 = 16$. Find the point on this curve which is closest to $(0, 0, 0)$.
18. Find the point on the plane, $2x + 3y + z = 4$ which is closest to the point $(1, 2, 3)$.
19. Let $A = (A_{ij})$ be an $n \times n$ matrix which is symmetric. Thus $A_{ij} = A_{ji}$ and recall $(A\mathbf{x})_i = A_{ij}x_j$ where as usual, sum over the repeated index. Show that $\frac{\partial}{\partial x_k}(A_{ij}x_jx_i) = 2A_{ik}x_i$. Show that when you use the method of Lagrange multipliers to maximize the function, $A_{ij}x_jx_i$ subject to the constraint, $\sum_{j=1}^n x_j^2 = 1$, the value of λ which corresponds to the maximum value of this functions is such that $A_{ij}x_j = \lambda x_i$. Thus $A\mathbf{x} = \lambda\mathbf{x}$. Thus λ is an eigenvalue of the matrix, A .
20. Here are two lines.

$$\mathbf{x} = (1 + 2t, 2 + t, 3 + t)^T$$
and $\mathbf{x} = (2 + s, 1 + 2s, 1 + 3s)^T$. Find points $\mathbf{p}_1$ on the first line and $\mathbf{p}_2$ on the second with the property that $|\mathbf{p}_1 - \mathbf{p}_2|$ is at least as small as the distance between any other pair of points, one chosen on one line and the other on the other line.
21. * Find points on the circle of radius r for the largest triangle which can be inscribed in it.
22. Find the point on the intersection of $z = x^2 + y^2$ and $x + y + z = 1$ which is closest to $(0, 0, 0)$.
23. Minimize xyz subject to the constraints $x^2 + y^2 + z^2 = r^2$ and $x - y = 0$.
24. Let n be a positive integer. Find n numbers whose sum is $8n$ and the sum of the squares is as small as possible.
25. Find the point on the level surface, $2x^2 + xy + z^2 = 16$ which is closest to $(0, 0, 0)$.
26. Find the point on $x^2 + y^2 + z^2 = 1$ closest to the plane $x + y + z = 10$.
27. Find the point on $\frac{x^2}{4} + \frac{y^2}{9} + z^2 = 1$ closest to the plane $x + y + z = 10$.
28. Let $x_1, \dots, x_5$ be 5 positive numbers. Maximize their product subject to the constraint that

$$x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 300.$$
29. Let $f(x_1, \dots, x_n) = x_1^n x_2^{n-1} \dots x_n^1$. Then f achieves a maximum on the set, $S \equiv$

$$\left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n ix_i = 1, \text{ each } x_i \geq 0 \right\}$$
If $\mathbf{x} \in S$ is the point where this maximum is achieved, find x_1/x_n .
30. * Let (x, y) be a point on the ellipse, $x^2/a^2 + y^2/b^2 = 1$ which is in the first quadrant. Extend the tangent line through (x, y) till it intersects the x and y axes and let $A(x, y)$ denote the area of the triangle formed by this line and the two coordinate axes. Find the minimum value of the area of this triangle as a function of a and b .
31. Maximize $\prod_{i=1}^n x_i^2$

$$(\equiv x_1^2 \times x_2^2 \times x_3^2 \times \dots \times x_n^2)$$
subject to the constraint, $\sum_{i=1}^n x_i^2 = r^2$. Show that the maximum is $(r^2/n)^n$. Now show from this that

$$\left(\prod_{i=1}^n x_i^2 \right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n x_i^2$$
and finally, conclude that if each number $x_i \geq 0$, then

$$\left(\prod_{i=1}^n x_i \right)^{1/n} \leq \frac{1}{n} \sum_{i=1}^n x_i$$
and there exist values of the x_i for which equality holds. This says the "geometric mean" is always smaller than the arithmetic mean.

32. Maximize x^2y^2 subject to the constraint

$$\frac{x^{2p}}{p} + \frac{y^{2q}}{q} = r^2$$

where p, q are real numbers larger than 1 which have the property that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

show that the maximum is achieved when $x^{2p} = y^{2q}$ and equals r^2 . Now

conclude that if $x, y > 0$, then

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}$$

and there are values of x and y where this inequality is an equation.

33. The area of the ellipse $x^2/a^2 + y^2/b^2 \leq 1$ is πab which is given to equal π . The length of the ellipse is $\int_0^{2\pi} \sqrt{a^2 \sin^2(t) + b^2 \cos^2(t)} dt$. Find a, b such that the ellipse having this volume is as short as possible.

23.7 Proof Of The Second Derivative Test*

Definition 23.7.1 The matrix, $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})\right)$ is called the Hessian matrix, denoted by $H(\mathbf{x})$.

Now recall the Taylor formula with the Lagrange form of the remainder. See Theorem 10.1.1 on page 253 for a proof.

Theorem 23.7.2 Let $h : (-\delta, 1 + \delta) \rightarrow \mathbb{R}$ have $m + 1$ derivatives. Then there exists $t \in (0, 1)$ such that

$$h(1) = h(0) + \sum_{k=1}^m \frac{h^{(k)}(0)}{k!} + \frac{h^{(m+1)}(t)}{(m+1)!}.$$

Now let $f : U \rightarrow \mathbb{R}$ where U is an open subset of $\mathbb{R}^n$. Suppose $f \in C^2(U)$. Let $\mathbf{x} \in U$ and let $r > 0$ be such that

$$B(\mathbf{x}, r) \subseteq U.$$

Then for $\|\mathbf{v}\| < r$ consider

$$f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x}) \equiv h(t)$$

for $t \in [0, 1]$. Then from Taylor's theorem for the case where $m = 2$ and the chain rule, using the repeated index summation convention and the chain rule,

$$h'(t) = \frac{\partial f}{\partial x_i}(\mathbf{x} + t\mathbf{v}) v_i, \quad h''(t) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x} + t\mathbf{v}) v_i v_j.$$

Thus

$$h''(t) = \mathbf{v}^T H(\mathbf{x} + t\mathbf{v}) \mathbf{v}.$$

From Theorem 23.7.2 there exists $t \in (0, 1)$ such that

$$f(\mathbf{x} + \mathbf{v}) = f(\mathbf{x}) + \frac{\partial f}{\partial x_i}(\mathbf{x}) v_i + \frac{1}{2} \mathbf{v}^T H(\mathbf{x} + t\mathbf{v}) \mathbf{v}$$

By the continuity of the second partial derivative

$$f(\mathbf{x} + \mathbf{v}) = f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{v} + \frac{1}{2} \mathbf{v}^T H(\mathbf{x}) \mathbf{v} +$$

$$\frac{1}{2} (\mathbf{v}^T (H(\mathbf{x}+t\mathbf{v}) - H(\mathbf{x})) \mathbf{v}) \quad (23.5)$$

where the last term satisfies

$$\lim_{|\mathbf{v}| \rightarrow 0} \frac{1}{2} \frac{(\mathbf{v}^T (H(\mathbf{x}+t\mathbf{v}) - H(\mathbf{x})) \mathbf{v})}{|\mathbf{v}|^2} = 0 \quad (23.6)$$

because of the continuity of the entries of $H(\mathbf{x})$.

Theorem 23.7.3 *Suppose $\mathbf{x}$ is a critical point for f . That is, suppose $\frac{\partial f}{\partial x_i}(\mathbf{x}) = 0$ for each i . Then if $H(\mathbf{x})$ has all positive eigenvalues, $\mathbf{x}$ is a local minimum. If $H(\mathbf{x})$ has all negative eigenvalues, then $\mathbf{x}$ is a local maximum. If $H(\mathbf{x})$ has a positive eigenvalue, then there exists a direction in which f has a local minimum at $\mathbf{x}$, while if $H(\mathbf{x})$ has a negative eigenvalue, there exists a direction in which f has a local maximum at $\mathbf{x}$.*

Proof: Since $\nabla f(\mathbf{x}) = \mathbf{0}$, formula 23.5 implies

$$f(\mathbf{x} + \mathbf{v}) = f(\mathbf{x}) + \frac{1}{2} \mathbf{v}^T H(\mathbf{x}) \mathbf{v} + \frac{1}{2} (\mathbf{v}^T (H(\mathbf{x}+t\mathbf{v}) - H(\mathbf{x})) \mathbf{v}) \quad (23.7)$$

and by continuity of the second derivatives, these mixed second derivatives are equal and so $H(\mathbf{x})$ is a symmetric matrix. Thus, by Lemma C.2.27 on Page 659 in the appendix, $H(\mathbf{x})$ has all real eigenvalues. Suppose first that $H(\mathbf{x})$ has all positive eigenvalues and that all are larger than $\delta^2 > 0$. Then by Theorem C.2.29 on Page 659 of the appendix,

$$\mathbf{u}^T H(\mathbf{x}) \mathbf{u} \geq \delta^2 |\mathbf{u}|^2$$

By continuity of H , if $\mathbf{v}$ is small enough,

$$f(\mathbf{x} + \mathbf{v}) \geq f(\mathbf{x}) + \frac{1}{2} \delta^2 |\mathbf{v}|^2 - \frac{1}{4} \delta^2 |\mathbf{v}|^2 = f(\mathbf{x}) + \frac{\delta^2}{4} |\mathbf{v}|^2.$$

This shows the first claim of the theorem. The second claim follows from similar reasoning or applying the above to $-f$.

Suppose $H(\mathbf{x})$ has a positive eigenvalue λ^2 . Then let $\mathbf{v}$ be an eigenvector for this eigenvalue. Then from 23.7, replacing $\mathbf{v}$ with $s\mathbf{v}$ and letting t depend on s ,

$$\begin{aligned} f(\mathbf{x} + s\mathbf{v}) &= f(\mathbf{x}) + \frac{1}{2} s^2 \mathbf{v}^T H(\mathbf{x}) \mathbf{v} + \\ &\quad \frac{1}{2} s^2 (\mathbf{v}^T (H(\mathbf{x} + ts\mathbf{v}) - H(\mathbf{x})) \mathbf{v}) \end{aligned}$$

which implies

$$\begin{aligned} f(\mathbf{x} + s\mathbf{v}) &= f(\mathbf{x}) + \frac{1}{2} s^2 \lambda^2 |\mathbf{v}|^2 + \frac{1}{2} s^2 (\mathbf{v}^T (H(\mathbf{x} + ts\mathbf{v}) - H(\mathbf{x})) \mathbf{v}) \\ &\geq f(\mathbf{x}) + \frac{1}{4} s^2 \lambda^2 |\mathbf{v}|^2 \end{aligned}$$

whenever s is small enough. Thus in the direction $\mathbf{v}$ the function has a local minimum at $\mathbf{x}$. The assertion about the local maximum in some direction follows similarly. This proves the theorem.

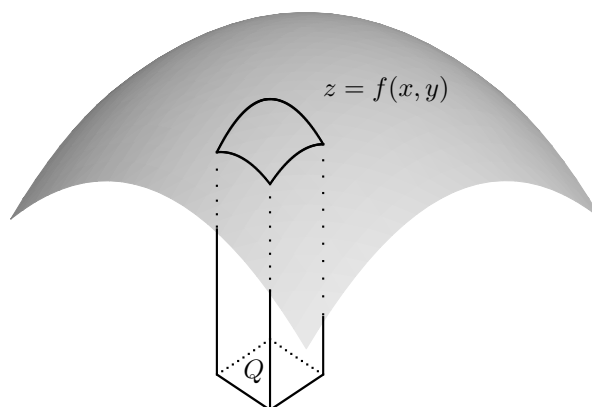
The Riemann Integral On $\mathbb{R}^n$

24.0.1 Outcomes

1. Recall and define the Riemann integral.
2. Recall the relation between iterated integrals and the Riemann integral.
3. Evaluate double integrals over simple regions.
4. Evaluate multiple integrals over simple regions.
5. Use multiple integrals to calculate the volume and mass.

24.1 Methods For Double Integrals

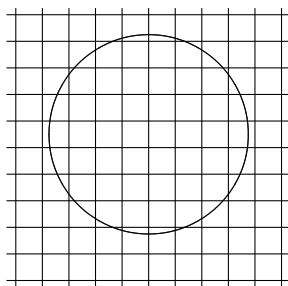
This chapter is on the Riemann integral for a function of n variables. It begins by introducing the basic concepts and applications of the integral. The proofs of the theorems involved are difficult and are left till the end. To begin with consider the problem of finding the volume under a surface of the form $z = f(x, y)$ where $f(x, y) \geq 0$ and $f(x, y) = 0$ for all (x, y) outside of some bounded set. To solve this problem, consider the following picture.



In this picture, the volume of the little prism which lies above the rectangle Q and below the graph of the function would lie between $M_Q(f) v(Q)$ and $m_Q(f) v(Q)$ where

$$M_Q(f) \equiv \sup \{f(\mathbf{x}) : \mathbf{x} \in Q\}, \quad m_Q(f) \equiv \inf \{f(\mathbf{x}) : \mathbf{x} \in Q\}, \quad (24.1)$$

and $v(Q)$ is defined as the area of Q . Now consider the following picture.

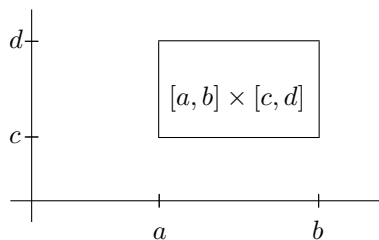


In this picture, it is assumed f equals zero outside the circle and f is a bounded nonnegative function. Then each of those little squares are the base of a prism of the sort in the previous picture and the sum of the volumes of those prisms should be the volume under the surface, $z = f(x, y)$. Therefore, the desired volume must lie between the two numbers,

$$\sum_Q M_Q(f) v(Q) \quad \text{and} \quad \sum_Q m_Q(f) v(Q)$$

where the notation, $\sum_Q M_Q(f) v(Q)$, means for each Q , take $M_Q(f)$, multiply it by the area of Q , $v(Q)$, and then add all these numbers together. Thus in $\sum_Q M_Q(f) v(Q)$, adds numbers which are at least as large as what is desired while in $\sum_Q m_Q(f) v(Q)$ numbers are added which are at least as small as what is desired. Note this is a finite sum because by assumption, $f = 0$ except for finitely many Q , namely those which intersect the circle. The sum, $\sum_Q M_Q(f) v(Q)$ is called an upper sum, $\sum_Q m_Q(f) v(Q)$ is a lower sum, and the desired volume is caught between these upper and lower sums.

None of this depends in any way on the function being nonnegative. It also does not depend in any essential way on the function being defined on $\mathbb{R}^2$, although it is impossible to draw meaningful pictures in higher dimensional cases. To define the Riemann integral, it is necessary to first give a description of something called a **grid**. First you must understand that something like $[a, b] \times [c, d]$ is a rectangle in $\mathbb{R}^2$, having sides parallel to the axes. The situation is illustrated in the following picture.



$(x, y) \in [a, b] \times [c, d]$, means $x \in [a, b]$ and also $y \in [c, d]$ and the points which do this comprise the rectangle just as shown in the picture.

Definition 24.1.1 For $i = 1, 2$, let $\{\alpha_k^i\}_{k=-\infty}^{\infty}$ be points on $\mathbb{R}$ which satisfy

$$\lim_{k \rightarrow \infty} \alpha_k^i = \infty, \quad \lim_{k \rightarrow -\infty} \alpha_k^i = -\infty, \quad \alpha_k^i < \alpha_{k+1}^i. \quad (24.2)$$

For such sequences, define a **grid** on $\mathbb{R}^2$ denoted by $\mathcal{G}$ or $\mathcal{F}$ as the collection of rectangles of the form

$$Q = [\alpha_k^1, \alpha_{k+1}^1] \times [\alpha_l^2, \alpha_{l+1}^2]. \quad (24.3)$$

If $\mathcal{G}$ is a grid, another grid, $\mathcal{F}$ is a **refinement** of $\mathcal{G}$ if every box of $\mathcal{G}$ is the union of boxes of $\mathcal{F}$.

For $\mathcal{G}$ a grid, the expression,

$$\sum_{Q \in \mathcal{G}} M_Q(f) v(Q)$$

is called the upper sum associated with the grid, $\mathcal{G}$ as described above in the discussion of the volume under a surface. Again, this means to take a rectangle from $\mathcal{G}$ multiply $M_Q(f)$ defined in 24.1 by its area, $v(Q)$ and sum all these products for every $Q \in \mathcal{G}$. The symbol,

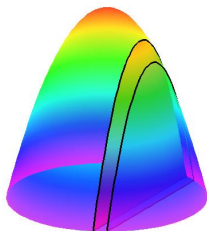
$$\sum_{Q \in \mathcal{G}} m_Q(f) v(Q),$$

called a lower sum, is defined similarly. With this preparation it is time to give a definition of the **Riemann integral** of a function of two variables.

Definition 24.1.2 *Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a bounded function which equals zero for all (x, y) outside some bounded set. Then $\int f dV$ is defined to be the unique number which lies between all upper sums and all lower sums. In the case of $\mathbb{R}^2$, it is common to replace the V with A and write this symbol as $\int f dA$ where A stands for area.*

This definition begs a difficult question. For which functions does there exist a unique number between all the upper and lower sums? This interesting and fundamental question is discussed in any advanced calculus book and may be seen in the appendix on the theory of the Riemann integral. It is a hard problem which was only solved in the first part of the twentieth century. When it was solved, it was also realized that the Riemann integral was not the right integral to use.

Consider the question: How can the Riemann integral be computed? Consider the following picture where f is assumed to be 0 outside the base of the solid which is contained in some rectangle $[a, b] \times [c, d]$.



It depicts a slice taken from the solid defined by $\{(x, y) : 0 \leq y \leq f(x, y)\}$. You see these when you look at a loaf of bread. If you wanted to find the volume of the loaf of bread, and you knew the volume of each slice of bread, you could find the volume of the whole loaf by adding the volumes of individual slices. It is the same here. If you could find the volume of the slice represented in this picture, you could add these up and get the volume of the solid. The slice in the picture corresponds to y and $y + h$ and is assumed to be very thin, having thickness equal to h . Denote the volume of the solid under the graph of $z = f(x, y)$ on $[a, b] \times [c, y]$ by $V(y)$. Then

$$V(y + h) - V(y) \approx h \int_a^b f(x, y) dx$$

where the integral is obtained by fixing y and integrating with respect to x and is the area of the cross section corresponding to y . It is hoped that the approximation would be increasingly good as h gets smaller. Thus, dividing by h and taking a limit, it is expected that

$$V'(y) = \int_a^b f(x, y) dx, \quad V(c) = 0.$$

Therefore, as in the method of cross sections, the volume of the solid under the graph of $z = f(x, y)$ is obtained by doing $\int_c^d$ to both sides,

$$\int_c^d \left(\int_a^b f(x, y) dx \right) dy \quad (24.4)$$

but this volume was also the result of $\int f dV$. Therefore, it is expected that this is a way to evaluate $\int f dV$.

Note what has been gained here. A hard problem, finding $\int f dV$, is reduced to a sequence of easier problems. First do

$$\int_a^b f(x, y) dx$$

getting a function of y , say $F(y)$ and then do

$$\int_c^d \left(\int_a^b f(x, y) dx \right) dy = \int_c^d F(y) dy.$$

Of course there is nothing special about fixing y first. The same thing should be obtained from the integral,

$$\int_a^b \left(\int_c^d f(x, y) dy \right) dx \quad (24.5)$$

These expressions in 24.4 and 24.5 are called **iterated integrals**. They are tools for evaluating $\int f dV$ which would be hard to find otherwise. In practice, the parenthesis is usually omitted in these expressions. Thus

$$\int_a^b \left(\int_c^d f(x, y) dy \right) dx = \int_a^b \int_c^d f(x, y) dy dx$$

and it is understood that you are to do the inside integral first and then when you have done it, obtaining a function of x , you integrate this function of x . Note that this is nothing more than using an integral to compute the area of a cross section and then using this method to find a volume.

However, there is no difference in the general case where f is not necessarily non-negative as can be seen by applying the method to the nonnegative functions f^+ , f^- given by

$$f^+ \equiv \frac{|f| + f}{2}, \quad f^- \equiv \frac{|f| - f}{2}$$

and then noting that $f = f^+ - f^-$ and the integral is linear. Thus

$$\begin{aligned} \int f dV &= \int f^+ - f^- dV = \int f^+ dV - \int f^- dV \\ &= \int_a^b \int_c^d f^+(x, y) dy dx - \int_a^b \int_c^d f^-(x, y) dy dx \\ &= \int_a^b \int_c^d f(x, y) dy dx \end{aligned}$$

A careful presentation which is not for the faint of heart is in an appendix.

Another aspect of this is the notion of integrating a function which is defined on some set, not on all $\mathbb{R}^2$. For example, suppose f is defined on the set, $S \subseteq \mathbb{R}^2$. What is meant by $\int_S f dV$?

Definition 24.1.3 Let $f : S \rightarrow \mathbb{R}$ where S is a subset of $\mathbb{R}^2$. Then denote by f_1 the function defined by

$$f_1(x, y) \equiv \begin{cases} f(x, y) & \text{if } (x, y) \in S \\ 0 & \text{if } (x, y) \notin S \end{cases}.$$

Then

$$\int_S f dV \equiv \int f_1 dV.$$

Example 24.1.4 Let $f(x, y) = x^2y + yx$ for $(x, y) \in [0, 1] \times [0, 2] \equiv R$. Find $\int_R f dV$.

This is done using iterated integrals like those defined above. Thus

$$\int_R f dV = \int_0^1 \int_0^2 (x^2y + yx) dy dx.$$

The inside integral yields

$$\int_0^2 (x^2y + yx) dy = 2x^2 + 2x$$

and now the process is completed by doing $\int_0^1$ to what was just obtained. Thus

$$\int_0^1 \int_0^2 (x^2y + yx) dy dx = \int_0^1 (2x^2 + 2x) dx = \frac{5}{3}.$$

If the integration is done in the opposite order, the same answer should be obtained.

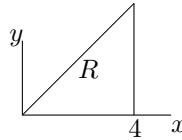
$$\begin{aligned} \int_0^2 \int_0^1 (x^2y + yx) dx dy \\ \int_0^1 (x^2y + yx) dx = \frac{5}{6}y \end{aligned}$$

Now

$$\int_0^2 \int_0^1 (x^2y + yx) dx dy = \int_0^2 \left(\frac{5}{6}y\right) dy = \frac{5}{3}.$$

If a different answer had been obtained it would have been a sign that a mistake had been made.

Example 24.1.5 Let $f(x, y) = x^2y + yx$ for $(x, y) \in R$ where R is the triangular region defined to be in the first quadrant, below the line $y = x$ and to the left of the line $x = 4$. Find $\int_R f dV$.



Now from the above discussion,

$$\int_R f dV = \int_0^4 \int_0^x (x^2y + yx) dy dx$$

The reason for this is that x goes from 0 to 4 and for each fixed x between 0 and 4, y goes from 0 to the slanted line, $y = x$, the function being defined to be 0 for larger

y . Thus y goes from 0 to x . This explains the inside integral. Now $\int_0^x (x^2y + yx) dy = \frac{1}{2}x^4 + \frac{1}{2}x^3$ and so

$$\int_R f dV = \int_0^4 \left(\frac{1}{2}x^4 + \frac{1}{2}x^3 \right) dx = \frac{672}{5}.$$

What of integration in a different order? Lets put the integral with respect to y on the outside and the integral with respect to x on the inside. Then

$$\int_R f dV = \int_0^4 \int_y^4 (x^2y + yx) dx dy$$

For each y between 0 and 4, the variable x , goes from y to 4.

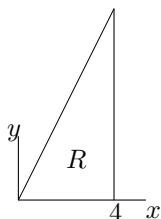
$$\int_y^4 (x^2y + yx) dx = \frac{88}{3}y - \frac{1}{3}y^4 - \frac{1}{2}y^3$$

Now

$$\int_R f dV = \int_0^4 \left(\frac{88}{3}y - \frac{1}{3}y^4 - \frac{1}{2}y^3 \right) dy = \frac{672}{5}.$$

Here is a similar example.

Example 24.1.6 Let $f(x, y) = x^2y$ for $(x, y) \in R$ where R is the triangular region defined to be in the first quadrant, below the line $y = 2x$ and to the left of the line $x = 4$. Find $\int_R f dV$.



Put the integral with respect to x on the outside first. Then

$$\int_R f dV = \int_0^4 \int_0^{2x} (x^2y) dy dx$$

because for each $x \in [0, 4]$, y goes from 0 to $2x$. Then

$$\int_0^{2x} (x^2y) dy = 2x^4$$

and so

$$\int_R f dV = \int_0^4 (2x^4) dx = \frac{2048}{5}$$

Now do the integral in the other order. Here the integral with respect to y will be on the outside. What are the limits of this integral? Look at the triangle and note that x goes from 0 to 4 and so $2x = y$ goes from 0 to 8. Now for fixed y between 0 and 8, where does x go? It goes from the x coordinate on the line $y = 2x$ which corresponds to this y to 4. What is the x coordinate on this line which goes with y ? It is $x = y/2$. Therefore, the iterated integral is

$$\int_0^8 \int_{y/2}^4 (x^2y) dx dy.$$

Now

$$\int_{y/2}^4 (x^2 y) \, dx = \frac{64}{3} y - \frac{1}{24} y^4$$

and so

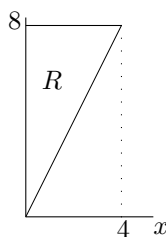
$$\int_R f \, dV = \int_0^8 \left(\frac{64}{3} y - \frac{1}{24} y^4 \right) dy = \frac{2048}{5}$$

the same answer.

A few observations are in order here. In finding $\int_S f \, dV$ there is no problem in setting things up if S is a rectangle. However, if S is not a rectangle, the procedure **always** is agonizing. A good rule of thumb is that if what you do is easy it will be wrong. There are no shortcuts! There are no quick fixes which require no thought! Pain and suffering is inevitable and you must not expect it to be otherwise. Always draw a picture and then begin **agonizing** over the correct limits. Even when you are careful you will make lots of mistakes until you get used to the process.

Sometimes an integral can be evaluated in one order but not in another.

Example 24.1.7 For R as shown below, find $\int_R \sin(y^2) \, dV$.



Setting this up to have the integral with respect to y on the inside yields

$$\int_0^4 \int_{2x}^8 \sin(y^2) \, dy \, dx.$$

Unfortunately, there is no antiderivative in terms of elementary functions for $\sin(y^2)$ so there is an immediate problem in evaluating the inside integral. It doesn't work out so the next step is to do the integration in another order and see if some progress can be made. This yields

$$\int_0^8 \int_0^{y/2} \sin(y^2) \, dx \, dy = \int_0^8 \frac{y}{2} \sin(y^2) \, dy$$

and $\int_0^8 \frac{y}{2} \sin(y^2) \, dy = -\frac{1}{4} \cos 64 + \frac{1}{4}$ which you can verify by making the substitution, $u = y^2$. Thus

$$\int_R \sin(y^2) \, dy = -\frac{1}{4} \cos 64 + \frac{1}{4}.$$

This illustrates an important idea. The integral $\int_R \sin(y^2) \, dV$ is defined as a number. It is the unique number between all the upper sums and all the lower sums. Finding it is another matter. In this case it was possible to find it using one order of integration but not the other. The iterated integral in this other order also is defined as a number but it can't be found directly without interchanging the order of integration. Of course sometimes nothing you try will work out.

24.1.1 Density And Mass

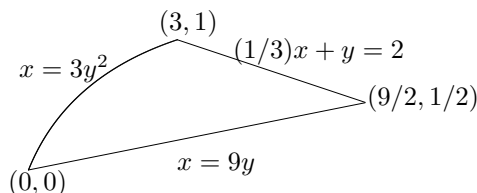
Consider a two dimensional material. Of course there is no such thing but a flat plate might be modeled as one. The density ρ is a function of position and is defined as follows. Consider a small chunk of area, dV located at the point whose Cartesian coordinates are (x, y) . Then the mass of this small chunk of material is given by $\rho(x, y) dV$. Thus if the material occupies a region in two dimensional space, U , the total mass of this material would be

$$\int_U \rho dV$$

In other words you integrate the density to get the mass. Now by letting ρ depend on position, you can include the case where the material is not homogeneous. Here is an example.

Example 24.1.8 Let $\rho(x, y)$ denote the density of the plane region determined by the curves $\frac{1}{3}x + y = 2$, $x = 3y^2$, and $x = 9y$. Find the total mass if $\rho(x, y) = y$.

You need to first draw a picture of the region, R . A rough sketch follows.



This region is in two pieces, one having the graph of $x = 9y$ on the bottom and the graph of $x = 3y^2$ on the top and another piece having the graph of $x = 9y$ on the bottom and the graph of $\frac{1}{3}x + y = 2$ on the top. Therefore, in setting up the integrals, with the integral with respect to x on the outside, the double integral equals the following sum of iterated integrals.

$$\overbrace{\int_0^3 \int_{x/9}^{\sqrt{x/3}} y dy dx}^{\text{has } x=3y^2 \text{ on top}} + \overbrace{\int_3^{9/2} \int_{x/9}^{2-\frac{1}{3}x} y dy dx}^{\text{has } \frac{1}{3}x+y=2 \text{ on top}}$$

You notice it is not necessary to have a perfect picture, just one which is good enough to figure out what the limits should be. The dividing line between the two cases is $x = 3$ and this was shown in the picture. Now it is only a matter of evaluating the iterated integrals which in this case is routine and gives 1.

24.2 Exercises

1. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_0^4 \int_0^{3y} x dx dy$.
2. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_0^3 \int_0^{3y} y dx dy$.
3. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_0^2 \int_0^{2y} (x+1) dx dy$.

4. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_0^3 \int_0^y \sin(x) \, dx \, dy$.
5. Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_0^1 \int_0^y \exp(y) \, dx \, dy$.
6. Let $\rho(x, y)$ denote the density of the plane region closest to $(0, 0)$ which is between the curves $x + 2y = 3$, $x = y^2$, and $x = 0$. Find the total mass if $\rho(x, y) = y$. Set up the integral in terms of $dx dy$ and in terms of $dy dx$.
7. Let $\rho(x, y)$ denote the density of the plane region determined by the curves $x + 2y = 3$, $x = y^2$, and $x = 4y$. Find the total mass if $\rho(x, y) = x$. Set up the integral in terms of $dx dy$ and $dy dx$.
8. Let $\rho(x, y)$ denote the density of the plane region determined by the curves $y = 2x$, $y = x$, $x + y = 3$. Find the total mass if $\rho(x, y) = y + 1$. Set up the integrals in terms of $dx dy$ and $dy dx$.
9. Let $\rho(x, y)$ denote the density of the plane region determined by the curves $y = 3x$, $y = x$, $2x + y = 4$. Find the total mass if $\rho(x, y) = 1$.
10. Let $\rho(x, y)$ denote the density of the plane region determined by the curves $y = 3x$, $y = x$, $x + y = 2$. Find the total mass if $\rho(x, y) = x + 1$. Set up the integrals in terms of $dx dy$ and $dy dx$.
11. Let $\rho(x, y)$ denote the density of the plane region determined by the curves $y = 5x$, $y = x$, $5x + 2y = 10$. Find the total mass if $\rho(x, y) = 1$. Set up the integrals in terms of $dx dy$ and $dy dx$.
12. Find $\int_0^4 \int_{y/2}^2 \frac{1}{x} e^{2\frac{y}{x}} \, dx \, dy$. You might need to interchange the order of integration.
13. Find $\int_0^8 \int_{y/2}^4 \frac{1}{x} e^{3\frac{y}{x}} \, dx \, dy$.
14. Find $\int_0^{\frac{1}{3}\pi} \int_x^{\frac{1}{3}\pi} \frac{\sin y}{y} \, dy \, dx$.
15. Find $\int_0^{\frac{1}{2}\pi} \int_x^{\frac{1}{2}\pi} \frac{\sin y}{y} \, dy \, dx$.
16. Find $\int_0^\pi \int_x^\pi \frac{\sin y}{y} \, dy \, dx$.
17. * Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_{-3}^3 \int_{-x}^x x^2 \, dy \, dx$

Your answer for the iterated integral should be $\int_3^0 \int_{-3}^{-y} x^2 \, dx \, dy + \int_0^{-3} \int_{-3}^y x^2 \, dx \, dy + \int_0^3 \int_y^3 x^2 \, dx \, dy + \int_{-3}^0 \int_{-y}^3 x^2 \, dx \, dy$. This is a very interesting example which shows that iterated integrals have a life of their own, not just as a method for evaluating double integrals.
18. * Evaluate the iterated integral and then write the iterated integral with the order of integration reversed. $\int_{-2}^2 \int_{-x}^x x^2 \, dy \, dx$.

24.3 Methods For Triple Integrals

24.3.1 Definition Of The Integral

The integral of a function of three variables is similar to the integral of a function of two variables.

Definition 24.3.1 For $i = 1, 2, 3$ let $\{\alpha_k^i\}_{k=-\infty}^{\infty}$ be points on $\mathbb{R}$ which satisfy

$$\lim_{k \rightarrow \infty} \alpha_k^i = \infty, \quad \lim_{k \rightarrow -\infty} \alpha_k^i = -\infty, \quad \alpha_k^i < \alpha_{k+1}^i. \quad (24.6)$$

For such sequences, define a **grid** on $\mathbb{R}^3$ denoted by $\mathcal{G}$ or $\mathcal{F}$ as the collection of boxes of the form

$$Q = [\alpha_k^1, \alpha_{k+1}^1] \times [\alpha_l^2, \alpha_{l+1}^2] \times [\alpha_p^3, \alpha_{p+1}^3]. \quad (24.7)$$

If $\mathcal{G}$ is a grid, $\mathcal{F}$ is called a **refinement** of $\mathcal{G}$ if every box of $\mathcal{G}$ is the union of boxes of $\mathcal{F}$.

For $\mathcal{G}$ a grid,

$$\sum_{Q \in \mathcal{G}} M_Q(f) v(Q)$$

is the upper sum associated with the grid, $\mathcal{G}$ where

$$M_Q(f) \equiv \sup \{f(\mathbf{x}) : \mathbf{x} \in Q\}$$

and if $Q = [a, b] \times [c, d] \times [e, f]$, then $v(Q)$ is the volume of Q given by $(b-a)(d-c)(f-e)$. Letting

$$m_Q(f) \equiv \inf \{f(\mathbf{x}) : \mathbf{x} \in Q\}$$

the lower sum associated with this partition is

$$\sum_{Q \in \mathcal{G}} m_Q(f) v(Q),$$

With this preparation it is time to give a definition of the **Riemann integral** of a function of three variables. This definition is just like the one for a function of two variables.

Definition 24.3.2 Let $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ be a bounded function which equals zero outside some bounded subset of $\mathbb{R}^3$. $\int f dV$ is defined as the unique number between all the upper sums and lower sums.

As in the case of a function of two variables there are all sorts of mathematical questions which are dealt with later.

The way to think of integrals is as follows. Located at a point $\mathbf{x}$, there is an “infinitesimal” chunk of volume, dV . The integral involves taking this little chunk of volume, dV , multiplying it by $f(\mathbf{x})$ and then adding up all such products. Upper sums are too large and lower sums are too small but the unique number between all the lower and upper sums is just right and corresponds to the notion of adding up all the $f(\mathbf{x}) dV$. Even the notation is suggestive of this concept of sum. It is a long thin S denoting sum. This is the fundamental concept for the integral in any number of dimensions and all the definitions and technicalities are designed to give precision and mathematical respectability to this notion.

Integrals of functions of three variables are also evaluated by using iterated integrals. Imagine a sum of the form $\sum_{ijk} a_{ijk}$ where there are only finitely many choices for i, j ,

and k and the symbol means you simply add up all the a_{ijk} . By the commutative law of addition, these may be added systematically in the form, $\sum_k \sum_j \sum_i a_{ijk}$. A similar process is used to evaluate triple integrals and since integrals are like sums, you might expect it to be valid. Specifically,

$$\int f dV = \int_{\gamma}^{\gamma} \int_{\gamma}^{\gamma} \int_{\gamma}^{\gamma} f(x, y, z) dx dy dz.$$

In words, sum with respect to x and then sum what you get with respect to y and finally, with respect to z . Of course this should hold in any other order such as

$$\int f dV = \int_{\gamma}^{\gamma} \int_{\gamma}^{\gamma} \int_{\gamma}^{\gamma} f(x, y, z) dz dy dx.$$

This is proved in an appendix¹.

Having discussed double and triple integrals, the definition of the integral of a function of n variables is accomplished in the same way.

Definition 24.3.3 For $i = 1, \dots, n$, let $\{\alpha_k^i\}_{k=-\infty}^{\infty}$ be points on $\mathbb{R}$ which satisfy

$$\lim_{k \rightarrow \infty} \alpha_k^i = \infty, \quad \lim_{k \rightarrow -\infty} \alpha_k^i = -\infty, \quad \alpha_k^i < \alpha_{k+1}^i. \quad (24.8)$$

For such sequences, define a grid on $\mathbb{R}^n$ denoted by $\mathcal{G}$ or $\mathcal{F}$ as the collection of boxes of the form

$$Q = \prod_{i=1}^n [\alpha_{j_i}^i, \alpha_{j_i+1}^i]. \quad (24.9)$$

If $\mathcal{G}$ is a grid, $\mathcal{F}$ is called a refinement of $\mathcal{G}$ if every box of $\mathcal{G}$ is the union of boxes of $\mathcal{F}$.

Definition 24.3.4 Let f be a bounded function which equals zero off a bounded set, D , and let $\mathcal{G}$ be a grid. For $Q \in \mathcal{G}$, define

$$M_Q(f) \equiv \sup \{f(\mathbf{x}) : \mathbf{x} \in Q\}, \quad m_Q(f) \equiv \inf \{f(\mathbf{x}) : \mathbf{x} \in Q\}. \quad (24.10)$$

Also define for Q a box, the volume of Q , denoted by $v(Q)$ by

$$v(Q) \equiv \prod_{i=1}^n (b_i - a_i), \quad Q \equiv \prod_{i=1}^n [a_i, b_i].$$

Now define upper sums, $\mathcal{U}_{\mathcal{G}}(f)$ and lower sums, $\mathcal{L}_{\mathcal{G}}(f)$ with respect to the indicated grid, by the formulas

$$\mathcal{U}_{\mathcal{G}}(f) \equiv \sum_{Q \in \mathcal{G}} M_Q(f) v(Q), \quad \mathcal{L}_{\mathcal{G}}(f) \equiv \sum_{Q \in \mathcal{G}} m_Q(f) v(Q).$$

Then a function of n variables is Riemann integrable if there is a unique number between all the upper and lower sums. This number is the value of the integral.

In this book most integrals will involve no more than three variables. However, this does not mean an integral of a function of more than three variables is unimportant. Therefore, I will begin to refer to the general case when theorems are stated.

Definition 24.3.5 For $E \subseteq \mathbb{R}^n$,

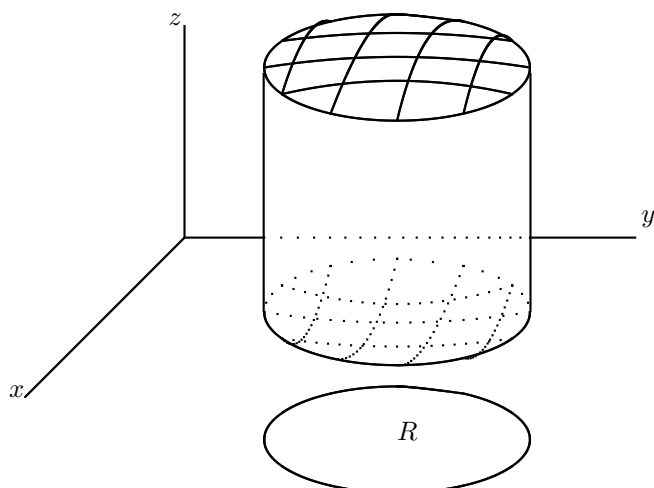
$$\mathcal{X}_E(\mathbf{x}) \equiv \begin{cases} 1 & \text{if } \mathbf{x} \in E \\ 0 & \text{if } \mathbf{x} \notin E \end{cases}.$$

Define $\int_E f dV \equiv \int \mathcal{X}_E f dV$ when $f \mathcal{X}_E \in \mathcal{R}(\mathbb{R}^n)$.

¹All of these fundamental questions about integrals can be considered more easily in the context of the Lebesgue integral. However, this integral is more abstract than the Riemann integral.

24.3.2 Iterated Integrals

As before, the integral is often computed by using an iterated integral. In general it is impossible to set up an iterated integral for finding $\int_E f dV$ for arbitrary regions, E but when the region is sufficiently simple, one can make progress. Suppose the region, E over which the integral is to be taken is of the form $E = \{(x, y, z) : a(x, y) \leq z \leq b(x, y)\}$ for $(x, y) \in R$, a two dimensional region. This is illustrated in the following picture in which the bottom surface is the graph of $z = a(x, y)$ and the top is the graph of $z = b(x, y)$.



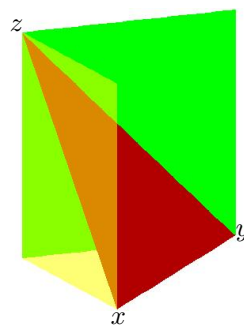
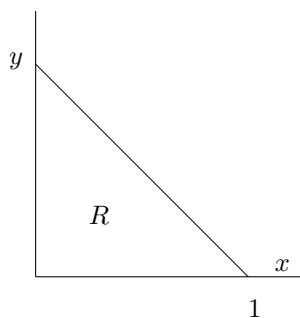
Then

$$\int_E f dV = \int_R \int_{a(x,y)}^{b(x,y)} f(x, y, z) dz dA$$

It might be helpful to think of $dV = dz dA$. Now $\int_{a(x,y)}^{b(x,y)} f(x, y, z) dz$ is a function of x and y and so you have reduced the triple integral to a double integral over R of this function of x and y . Similar reasoning would apply if the region in $\mathbb{R}^3$ were of the form $\{(x, y, z) : a(y, z) \leq x \leq b(y, z)\}$ or $\{(x, y, z) : a(x, z) \leq y \leq b(x, z)\}$.

Example 24.3.6 Find the volume of the region, E in the first octant between $z = 1 - (x + y)$ and $z = 0$.

In this case, R is the region shown.



Thus the region, E is between the plane $z = 1 - (x + y)$ on the top, $z = 0$ on the bottom, and over R shown above. Thus

$$\begin{aligned}\int_E 1 dV &= \int_R \int_0^{1-(x+y)} dz dA \\ &= \int_0^1 \int_0^{1-x} \int_0^{1-(x+y)} dz dy dx = \frac{1}{6}\end{aligned}$$

Of course iterated integrals have a life of their own although this will not be explored here. You can just write them down and go to work on them. Here are some examples.

Example 24.3.7 Find $\int_2^3 \int_3^x \int_{3y}^x (x - y) dz dy dx$.

The inside integral yields $\int_{3y}^x (x - y) dz = x^2 - 4xy + 3y^2$. Next this must be integrated with respect to y to give $\int_3^x (x^2 - 4xy + 3y^2) dy = -3x^2 + 18x - 27$. Finally the third integral gives

$$\int_2^3 \int_3^x \int_{3y}^x (x - y) dz dy dx = \int_2^3 (-3x^2 + 18x - 27) dx = -1.$$

Example 24.3.8 Find $\int_0^\pi \int_0^{3y} \int_0^{y+z} \cos(x + y) dx dz dy$.

The inside integral is $\int_0^{y+z} \cos(x + y) dx = 2 \cos z \sin y \cos y + 2 \sin z \cos^2 y - \sin z - \sin y$. Now this has to be integrated.

$$\begin{aligned}\int_0^{3y} \int_0^{y+z} \cos(x + y) dx dz &= \int_0^{3y} (2 \cos z \sin y \cos y + 2 \sin z \cos^2 y - \sin z - \sin y) dz \\ &= -1 - 16 \cos^5 y + 20 \cos^3 y - 5 \cos y - 3(\sin y)y + 2 \cos^2 y.\end{aligned}$$

Finally, this last expression must be integrated from 0 to π . Thus

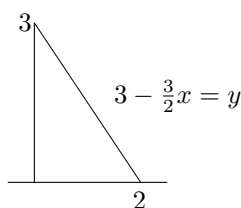
$$\begin{aligned}&\int_0^\pi \int_0^{3y} \int_0^{y+z} \cos(x + y) dx dz dy \\ &= \int_0^\pi (-1 - 16 \cos^5 y + 20 \cos^3 y - 5 \cos y - 3(\sin y)y + 2 \cos^2 y) dy \\ &= -3\pi\end{aligned}$$

Example 24.3.9 Here is an iterated integral: $\int_0^2 \int_0^{3-\frac{3}{2}x} \int_0^{x^2} dz dy dx$. Write as an iterated integral in the order $dz dx dy$.

The inside integral is just a function of x and y . (In fact, only a function of x .) The order of the last two integrals must be interchanged. Thus the iterated integral which needs to be done in a different order is

$$\int_0^2 \int_0^{3-\frac{3}{2}x} f(x, y) dy dx.$$

As usual, it is important to draw a picture and then go from there.



Thus this double integral equals

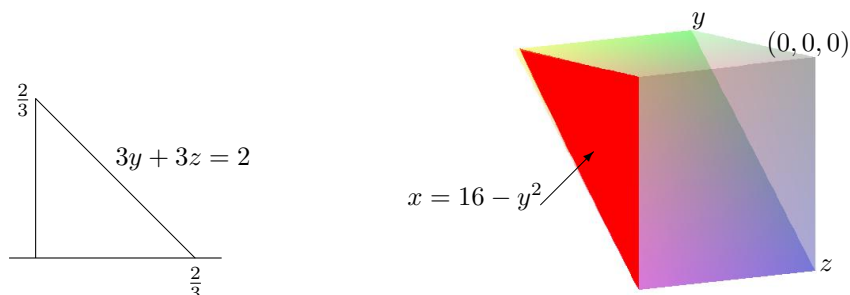
$$\int_0^3 \int_0^{\frac{2}{3}(3-y)} f(x, y) \, dx \, dy.$$

Now substituting in for $f(x, y)$,

$$\int_0^3 \int_0^{\frac{2}{3}(3-y)} \int_0^{x^2} dz \, dx \, dy.$$

Example 24.3.10 Find the volume of the bounded region determined by $3y + 3z = 2$, $x = 16 - y^2$, $y = 0$, $x = 0$.

In the yz plane, the first of the following pictures corresponds to $x = 0$.



Therefore, the outside integrals taken with respect to z and y are of the form $\int_0^{\frac{2}{3}} \int_0^{\frac{2}{3}-y} dz \, dy$ and now for any choice of (y, z) in the above triangular region, x goes from 0 to $16 - y^2$. Therefore, the iterated integral is

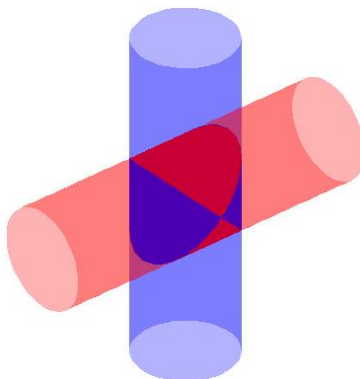
$$\int_0^{\frac{2}{3}} \int_0^{\frac{2}{3}-y} \int_0^{16-y^2} dx \, dz \, dy = \frac{860}{243}$$

Example 24.3.11 Find the volume of the region determined by the intersection of the two cylinders, $x^2 + y^2 \leq 9$ and $y^2 + z^2 \leq 9$.

The first listed cylinder intersects the xy plane in the disk, $x^2 + y^2 \leq 9$. What is the volume of the three dimensional region which is between this disk and the two surfaces, $z = \sqrt{9 - y^2}$ and $z = -\sqrt{9 - y^2}$? An iterated integral for the volume is

$$\int_{-3}^3 \int_{-\sqrt{9-y^2}}^{\sqrt{9-y^2}} \int_{-\sqrt{9-y^2}}^{\sqrt{9-y^2}} dz \, dx \, dy = 144.$$

Note I drew no picture of the three dimensional region. If you are interested, here it is.



One of the cylinders is parallel to the z axis, $x^2 + y^2 \leq 9$ and the other is parallel to the x axis, $y^2 + z^2 \leq 9$. I did not need to be able to draw such a nice picture in order to work this problem. This is the key to doing these. Draw pictures in two dimensions and reason from the two dimensional pictures rather than attempt to wax artistic and consider all three dimensions at once. These problems are hard enough without making them even harder by attempting to be an artist.

24.4 Exercises

- Find the following iterated integrals.

(a) $\int_{-1}^3 \int_0^{2z} \int_y^{z+1} (x+y) \, dx \, dy \, dz$

(b) $\int_0^1 \int_0^z \int_y^{z^2} (y+z) \, dx \, dy \, dz$

(c) $\int_0^3 \int_1^x \int_2^{3x-y} \sin(x) \, dz \, dy \, dx$

(d) $\int_0^1 \int_x^{2x} \int_y^{2y} dz \, dy \, dx$

(e) $\int_2^4 \int_2^{2x} \int_{2y}^x dz \, dy \, dx$

(f) $\int_0^3 \int_0^{2-5x} \int_0^{2-x-2y} 2x \, dz \, dy \, dx$

(g) $\int_0^2 \int_0^{1-3x} \int_0^{3-3x-2y} x \, dz \, dy \, dx$

(h) $\int_0^\pi \int_0^{3y} \int_0^{y+z} \cos(x+y) \, dx \, dz \, dy$

(i) $\int_0^\pi \int_0^{4y} \int_0^{y+z} \sin(x+y) \, dx \, dz \, dy$

- Fill in the missing limits. $\int_0^1 \int_0^z \int_0^z f(x, y, z) \, dx \, dy \, dz = \int_?^? \int_?^? \int_?^? f(x, y, z) \, dx \, dz \, dy$,

$$\int_0^1 \int_0^z \int_0^{2z} f(x, y, z) \, dx \, dy \, dz = \int_?^? \int_?^? \int_?^? f(x, y, z) \, dy \, dz \, dx,$$

$$\int_0^1 \int_0^z \int_0^z f(x, y, z) \, dx \, dy \, dz = \int_?^? \int_?^? \int_?^? f(x, y, z) \, dz \, dy \, dx,$$

$$\int_0^1 \int_{z/2}^{\sqrt{z}} \int_0^{y+z} f(x, y, z) \, dx \, dy \, dz = \int_?^? \int_?^? \int_?^? f(x, y, z) \, dx \, dz \, dy,$$

$$\int_4^6 \int_2^6 \int_0^4 f(x, y, z) \, dx \, dy \, dz = \int_?^? \int_?^? \int_?^? f(x, y, z) \, dz \, dy \, dx.$$

- Find the volume of R where R is the bounded region formed by the plane $\frac{1}{5}x + y + \frac{1}{4}z = 1$ and the planes $x = 0, y = 0, z = 0$.
- Find the volume of R where R is the bounded region formed by the plane $\frac{1}{5}x + \frac{1}{2}y + \frac{1}{4}z = 1$ and the planes $x = 0, y = 0, z = 0$.
- Find the volume of R where R is the bounded region formed by the plane $\frac{1}{5}x + \frac{1}{3}y + \frac{1}{4}z = 1$ and the planes $x = 0, y = 0, z = 0$.
- Find the volume of the bounded region determined by $3y + z = 3, x = 4 - y^2, y = 0, x = 0$.
- Find the volume of the region bounded by $x^2 + y^2 = 16, z = 3x, z = 0$, and $x \geq 0$.
- Find the volume of R where R is the bounded region formed by the plane $\frac{1}{4}x + \frac{1}{2}y + \frac{1}{4}z = 1$ and the planes $x = 0, y = 0, z = 0$.
- Here is an iterated integral: $\int_0^3 \int_0^{3-x} \int_0^{x^2} dz \, dy \, dx$. Write as an iterated integral in the following orders: $dz \, dx \, dy$, $dx \, dz \, dy$, $dx \, dy \, dz$, $dy \, dx \, dz$, $dy \, dz \, dx$.

10. Find the volume of the bounded region determined by $2y + z = 3, x = 9 - y^2, y = 0, x = 0, z = 0$.
11. Find the volume of the bounded region determined by $y + 2z = 3, x = 9 - y^2, y = 0, x = 0$.
12. Find the volume of the bounded region determined by $y + z = 2, x = 3 - y^2, y = 0, x = 0$.
13. Find the volume of the region bounded by $x^2 + y^2 = 25, z = x, z = 0$, and $x \geq 0$.
Your answer should be $\frac{250}{3}$.
14. Find the volume of the region bounded by $x^2 + y^2 = 9, z = 3x, z = 0$, and $x \geq 0$.

24.4.1 Mass And Density

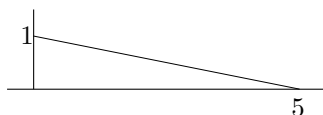
As an example of the use of triple integrals, consider a solid occupying a set of points, $U \subseteq \mathbb{R}^3$ having density ρ . Thus ρ is a function of position and the total mass of the solid equals

$$\int_U \rho dV.$$

This is just like the two dimensional case. The mass of an infinitesimal chunk of the solid located at $\mathbf{x}$ would be $\rho(\mathbf{x}) dV$ and so the total mass is just the sum of all these, $\int_U \rho(\mathbf{x}) dV$.

Example 24.4.1 Find the volume of R where R is the bounded region formed by the plane $\frac{1}{5}x + y + \frac{1}{5}z = 1$ and the planes $x = 0, y = 0, z = 0$.

When $z = 0$, the plane becomes $\frac{1}{5}x + y = 1$. Thus the intersection of this plane with the xy plane is this line shown in the following picture.



Therefore, the bounded region is between the triangle formed in the above picture by the x axis, the y axis and the above line and the surface given by $\frac{1}{5}x + y + \frac{1}{5}z = 1$ or $z = 5(1 - (\frac{1}{5}x + y)) = 5 - x - 5y$. Therefore, an iterated integral which yields the volume is

$$\int_0^5 \int_0^{1-\frac{1}{5}x} \int_0^{5-x-5y} dz dy dx = \frac{25}{6}.$$

Example 24.4.2 Find the mass of the bounded region, R formed by the plane $\frac{1}{3}x + \frac{1}{3}y + \frac{1}{5}z = 1$ and the planes $x = 0, y = 0, z = 0$ if the density is $\rho(x, y, z) = z$.

This is done just like the previous example except in this case there is a function to integrate. Thus the answer is

$$\int_0^3 \int_0^{3-x} \int_0^{5-\frac{5}{3}x-\frac{5}{3}y} z dz dy dx = \frac{75}{8}.$$

Example 24.4.3 Find the total mass of the bounded solid determined by $z = 9 - x^2 - y^2$ and $x, y, z \geq 0$ if the mass is given by $\rho(x, y, z) = z$

When $z = 0$ the surface, $z = 9 - x^2 - y^2$ intersects the xy plane in a circle of radius 3 centered at $(0, 0)$. Since $x, y \geq 0$, it is only a quarter of a circle of interest, the part where both these variables are nonnegative. For each (x, y) inside this quarter circle, z goes from 0 to $9 - x^2 - y^2$. Therefore, the iterated integral is of the form,

$$\int_0^3 \int_0^{\sqrt{9-x^2}} \int_0^{9-x^2-y^2} z \, dz \, dy \, dx = \frac{243}{8}\pi$$

Example 24.4.4 Find the volume of the bounded region determined by $x \geq 0, y \geq 0, z \geq 0$, and $\frac{1}{7}x + y + \frac{1}{4}z = 1$, and $x + \frac{1}{7}y + \frac{1}{4}z = 1$.

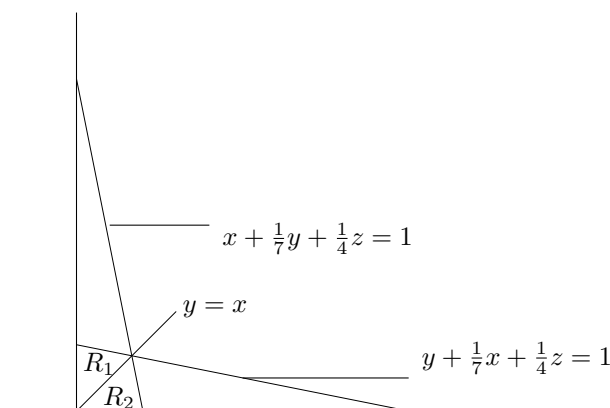
When $z = 0$, the plane $\frac{1}{7}x + y + \frac{1}{4}z = 1$ intersects the xy plane in the line whose equation is

$$\frac{1}{7}x + y = 1$$

while the plane, $x + \frac{1}{7}y + \frac{1}{4}z = 1$ intersects the xy plane in the line whose equation is

$$x + \frac{1}{7}y = 1.$$

Furthermore, the two planes intersect when $x = y$ as can be seen from the equations, $x + \frac{1}{7}y = 1 - \frac{z}{4}$ and $\frac{1}{7}x + y = 1 - \frac{z}{4}$ which imply $x = y$. Thus the two dimensional picture to look at is depicted in the following picture.



You see in this picture, the base of the region in the xy plane is the union of the two triangles, R_1 and R_2 . For $(x, y) \in R_1$, z goes from 0 to what it needs to be to be on the plane, $\frac{1}{7}x + y + \frac{1}{4}z = 1$. Thus z goes from 0 to $4(1 - \frac{1}{7}x - y)$. Similarly, on R_2 , z goes from 0 to $4(1 - \frac{1}{7}y - x)$. Therefore, the integral needed is

$$\int_{R_1} \int_0^{4(1-\frac{1}{7}x-y)} dz \, dV + \int_{R_2} \int_0^{4(1-\frac{1}{7}y-x)} dz \, dV$$

and now it only remains to consider $\int_{R_1} dV$ and $\int_{R_2} dV$. The point of intersection of these lines shown in the above picture is $(\frac{7}{8}, \frac{7}{8})$ and so an iterated integral is

$$\int_0^{7/8} \int_x^{1-\frac{x}{7}} \int_0^{4(1-\frac{1}{7}x-y)} dz \, dy \, dx + \int_0^{7/8} \int_y^{1-\frac{y}{7}} \int_0^{4(1-\frac{1}{7}y-x)} dz \, dx \, dy = \frac{7}{6}$$

24.5 Exercises

1. Find the volume of the region determined by the intersection of the two cylinders, $x^2 + y^2 \leq 16$ and $y^2 + z^2 \leq 16$.
2. Find the volume of the region determined by the intersection of the two cylinders, $x^2 + y^2 \leq 9$ and $y^2 + z^2 \leq 9$.
3. Find the volume of the region bounded by $x^2 + y^2 = 4$, $z = 0$, $z = 5 - y$
4. Find $\int_0^2 \int_0^{6-2z} \int_{\frac{1}{2}x}^{3-z} (3-z) \cos(y^2) \, dy \, dx \, dz$.
5. Find $\int_0^1 \int_0^{18-3z} \int_{\frac{1}{3}x}^{6-z} (6-z) \exp(y^2) \, dy \, dx \, dz$.
6. Find $\int_0^2 \int_0^{24-4z} \int_{\frac{1}{4}y}^{6-z} (6-z) \exp(x^2) \, dx \, dy \, dz$.
7. Find $\int_0^1 \int_0^{10-2z} \int_{\frac{1}{2}y}^{5-z} \frac{\sin x}{x} \, dx \, dy \, dz$.

Hint: Interchange order of integration.

8. Find the mass of the bounded region, R formed by the plane $\frac{1}{4}x + \frac{1}{2}y + \frac{1}{3}z = 1$ and the planes $x = 0, y = 0, z = 0$ if the density is $\rho(x, y, z) = y$
9. Find the mass of the bounded region, R formed by the plane $\frac{1}{2}x + \frac{1}{2}y + \frac{1}{4}z = 1$ and the planes $x = 0, y = 0, z = 0$ if the density is $\rho(x, y, z) = z^2$
10. Find the mass of the bounded region, R formed by the plane $\frac{1}{4}x + \frac{1}{2}y + \frac{1}{4}z = 1$ and the planes $x = 0, y = 0, z = 0$ if the density is $\rho(x, y, z) = y + z$
11. Find the mass of the bounded region, R formed by the plane $\frac{1}{4}x + \frac{1}{2}y + \frac{1}{5}z = 1$ and the planes $x = 0, y = 0, z = 0$ if the density is $\rho(x, y, z) = y$
12. Find $\int_0^1 \int_0^{12-4z} \int_{\frac{1}{4}y}^{3-z} \frac{\sin x}{x} \, dx \, dy \, dz$.
13. Find $\int_0^{20} \int_0^2 \int_{\frac{1}{5}y}^{6-z} \frac{\sin x}{x} \, dx \, dz \, dy + \int_{20}^{30} \int_0^{6-\frac{1}{5}y} \int_{\frac{1}{5}y}^{6-z} \frac{\sin x}{x} \, dx \, dz \, dy$.
14. Find the volume of the bounded region determined by $x \geq 0, y \geq 0, z \geq 0$, and $\frac{1}{2}x + y + \frac{1}{2}z = 1$, and $x + \frac{1}{2}y + \frac{1}{2}z = 1$.
15. Find the volume of the bounded region determined by $x \geq 0, y \geq 0, z \geq 0$, and $\frac{1}{7}x + y + \frac{1}{3}z = 1$, and $x + \frac{1}{7}y + \frac{1}{3}z = 1$.

The Integral In Other Coordinates

25.0.1 Outcomes

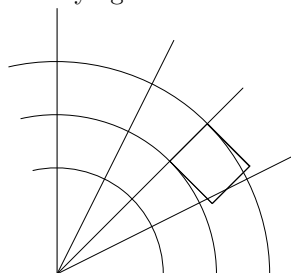
1. Represent a region in polar coordinates and use to evaluate integrals.
2. Represent a region in spherical or cylindrical coordinates and use to evaluate integrals.
3. Convert integrals in rectangular coordinates to integrals in polar coordinates and use to evaluate the integral.
4. Evaluate integrals in any coordinate system using the Jacobian.
5. Evaluate areas and volumes using another coordinate system.
6. Understand the transformation equations between spherical, polar and cylindrical coordinates and be able to change algebraic expressions from one system to another.
7. Use multiple integrals in an appropriate coordinate system to calculate the volume, mass, moments, center of gravity and moment of inertia.

25.1 Polar Coordinates

Recall the relation between the rectangular coordinates and polar coordinates is

$$\mathbf{x}(r, \theta) \equiv \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r \cos(\theta) \\ r \sin(\theta) \end{pmatrix}, r \geq 0, \theta \in [0, 2\pi)$$

Now consider the part of grid obtained by fixing θ at various values and varying r and then by fixing r at various values and varying θ .



The idea is that these lines obtained by fixing one or the other coordinate are very close together, much closer than drawn and so we would expect the area of one of the little curvy quadrilaterals to be close to the area of the parallelogram shown. Consider this parallelogram. The two sides originating at the intersection of two of the grid lines as shown are approximately equal to

$$\mathbf{x}_r(r, \theta) dr, \mathbf{x}_\theta(r, \theta) d\theta$$

where dr and $d\theta$ are the respective small changes in the variables r and θ . Thus the area of one of those little curvy shapes should be approximately equal to

$$|\mathbf{x}_r(r, \theta) dr \times \mathbf{x}_\theta(r, \theta) d\theta|$$

by Definition 16.4.2, where these vectors are extended as 0 in the third component in order to take the cross product. This reduces to

$$dA = \left| \det \begin{pmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{pmatrix} \right| dr d\theta = r dr d\theta$$

which is the increment of area in polar coordinates, taking the place of $dx dy$. The integral is really about taking the value of the function integrated multiplied by dA and adding these products. Here is an example.

Example 25.1.1 Find the area of a circle of radius a .

The variable r goes from 0 to a and the angle θ goes from 0 to 2π . Therefore, the area is

$$\int_D dA = \int_0^{2\pi} \int_0^a r dr d\theta = \pi a^2$$

Example 25.1.2 The density equals r . Find the total mass of a disk of radius a .

This is easy to do in polar coordinates. The disk involved has θ going from 0 to 2π and r from 0 to a . Therefore, the integral to work is just

$$\int_0^{2\pi} \int_0^a \overbrace{r r dr d\theta}^{dA} = \frac{2}{3} \pi a^3$$

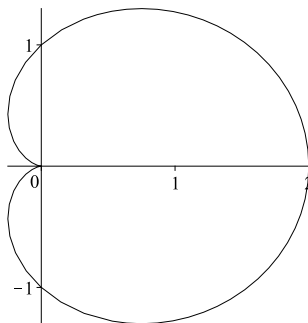
Notice how in these examples the circular disk is really a rectangle $[0, 2\pi] \times [0, a]$. This is why polar coordinates are so useful. The next example was worked earlier from a different point of view.

Example 25.1.3 Find the area of the inside of the cardioid $r = 1 + \cos \theta$, $\theta \in [0, 2\pi]$.

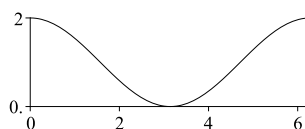
Here the integral is

$$\int_0^{2\pi} \int_0^{1+\cos(\theta)} r dr d\theta = \frac{3}{2} \pi$$

To see how impossible this problem is in rectangular coordinates, draw the graph of the cardioid.



How would you go about setting this up in rectangular coordinates? It would be very hard if not impossible but is easy in polar coordinates. This is because in polar coordinates the region integrated over is the region below the curve in the following picture.



Example 25.1.4 Let R denote the inside of the cardioid $r = 1 + \cos \theta$ for $\theta \in [0, 2\pi]$. Find

$$\int_R x dA$$

Here the convenient increment of area is $r dr d\theta$ and so the integral is

$$\int_0^{2\pi} \int_0^{1+\cos(\theta)} x r dr d\theta$$

Now you need to change x to the right coordinates. Thus the integral equals

$$\int_0^{2\pi} \int_0^{1+\cos(\theta)} (r \cos(\theta)) r dr d\theta = \frac{5}{4}\pi$$

A case where this sort of problem occurs is when you find the mass of a plate given the density.

Definition 25.1.5 Suppose a material occupies a region of the plane R . The density λ is a nonnegative function of position with the property that if $B \subseteq R$, then the mass of B is given by

$$\int_B \lambda dA$$

In particular, this is true of $B = R$.

Example 25.1.6 Let R denote the inside of the polar curve $r = 2 + \sin \theta$. Let $\lambda = 3 + x$. Find the total mass of R .

As above, this is

$$\int_0^{2\pi} \int_0^{2+\sin(\theta)} (3 + r \cos(\theta)) r dr d\theta = \frac{27}{2}\pi$$

25.2 Exercises

1. Sketch a graph in polar coordinates of $r = 2 + \sin(\theta)$ and find the area of the enclosed region.
2. Sketch a graph in polar coordinates of $r = \sin(4\theta)$ and find the area of the region enclosed. **Hint:** In this case, you need to worry and fuss about $r < 0$.

3. Suppose the density is $\lambda(x, y) = 2 - x$ and the region is the interior of the cardioid $r = 1 + \cos \theta$. Find the total mass.
4. Suppose the density is $\lambda = 4 - x - y$ and find the mass of the plate which is between the concentric circles $r = 1$ and $r = 2$.
5. Suppose the density is $\lambda = 4 - x - y$ and find the mass of the plate which is inside the polar graph of $r = 1 + \sin(\theta)$.
6. Suppose the density is $2 + x$. Find the mass of the plate which is the inside of the polar curve $r = \sin(2\theta)$. **Hint:** This is one of those fussy things with negative radius.
7. The moment about the x axis of a plate with density λ occupying the region R is defined as

$$y_c = \int_R y \lambda dA$$

The moment about the y axis of the same plate is

$$x_c = \int_R x \lambda dA$$

If $\lambda = 2 - x$, find the moments about the x and y axes of the plate inside $r = 2 + \sin(\theta)$.

8. Let $D = \{(x, y) : x^2 + y^2 \leq 25\}$. Find $\int_D e^{25x^2+25y^2} dx dy$. **Hint:** This is an integral of the form $\int_D f(x, y) dA$. Write in polar coordinates and it will be fairly easy.
9. Let $D = \{(x, y) : x^2 + y^2 \leq 16\}$. Find $\int_D \cos(9x^2 + 9y^2) dx dy$. **Hint:** This is an integral of the form $\int_D f(x, y) dA$. Write in polar coordinates and it will be fairly easy.
10. Derive a formula for area between two polar graphs using the increment of area of polar coordinates.
11. Use polar coordinates to evaluate the following integral. Here S is given in terms of the polar coordinates. $\int_S \sin(2x^2 + 2y^2) dV$ where $r \leq 2$ and $0 \leq \theta \leq \frac{3}{2}\pi$.
12. Find $\int_S e^{2x^2+2y^2} dV$ where S is given in terms of the polar coordinates, $r \leq 2$ and $0 \leq \theta \leq \pi$.
13. Find $\int_S \frac{y}{x} dV$ where S is described in polar coordinates as $1 \leq r \leq 2$ and $0 \leq \theta \leq \pi/4$.
14. Find $\int_S \left(\left(\frac{y}{x} \right)^2 + 1 \right) dV$ where S is given in polar coordinates as $1 \leq r \leq 2$ and $0 \leq \theta \leq \frac{1}{6}\pi$.

25.3 Cylindrical And Spherical Coordinates

Cylindrical coordinates are defined as follows.

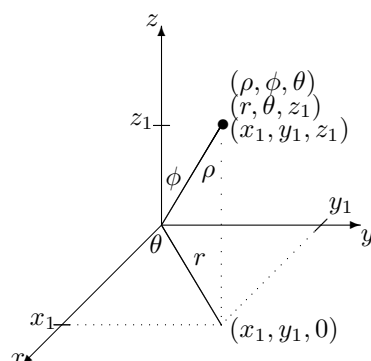
$$\begin{aligned} \mathbf{x}(r, \theta, z) &\equiv \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r \cos(\theta) \\ r \sin(\theta) \\ z \end{pmatrix}, \\ r &\geq 0, \theta \in [0, 2\pi), z \in \mathbb{R} \end{aligned}$$

Spherical coordinates are a little harder. Recall these are given by

$$\mathbf{x}(\rho, \theta, \phi) \equiv \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \rho \sin(\phi) \cos(\theta) \\ \rho \sin(\phi) \sin(\theta) \\ \rho \cos(\phi) \end{pmatrix},$$

$$\rho \geq 0, \theta \in [0, 2\pi), \phi \in [0, \pi]$$

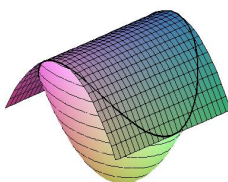
The following picture relates the various coordinates.



In this picture, ρ is the distance between the origin, the point whose Cartesian coordinates are $(0, 0, 0)$ and the point indicated by a dot and labelled as (x_1, y_1, z_1) , (r, θ, z_1) , and (ρ, ϕ, θ) . The angle between the positive z axis and the line between the origin and the point indicated by a dot is denoted by ϕ , and θ is the angle between the positive x axis and the line joining the origin to the point $(x_1, y_1, 0)$ as shown, while r is the length of this line. Thus $r = \rho \sin(\phi)$ and is the usual polar coordinate while θ is the other polar coordinate. Letting z_1 denote the usual z coordinate of a point in three dimensions, like the one shown as a dot, (r, θ, z_1) are the cylindrical coordinates of the dotted point. The spherical coordinates are determined by (ρ, ϕ, θ) . When ρ is specified, this indicates that the point of interest is on some sphere of radius ρ which is centered at the origin. Then when ϕ is given, the location of the point is narrowed down to a circle of “latitude” and finally, θ determines which point is on this circle by specifying a circle of “longitude”. Let $\phi \in [0, \pi]$, $\theta \in [0, 2\pi)$, and $\rho \in [0, \infty)$. The picture shows how to relate these new coordinate systems to Cartesian coordinates. Note that **θ is the same** in the two coordinate systems and that $\rho \sin \phi = r$.

The increment of three dimensional volume in cylindrical coordinates is $dV = r dr d\theta dz$. It is just a chunk of two dimensional area, $r dr d\theta$ times the height dz which gives three dimensional volume. Here is an example.

Example 25.3.1 Find the volume of the three dimensional region between the graphs of $z = 4 - 2y^2$ and $z = 4x^2 + 2y^2$.



Where do the two surfaces intersect? This happens when $4x^2 + 2y^2 = 4 - 2y^2$ which is the curve in the xy plane, $x^2 + y^2 = 1$. Thus (x, y) is on the inside of this circle while

z goes from $4x^2 + 2y^2$ to $4 - 2y^2$. Denoting the unit disk by D , the desired integral is

$$\int_D \int_{4x^2+2y^2}^{4-2y^2} dz dA$$

I will use the dA which corresponds to polar coordinates so this will then be in cylindrical coordinates. Thus the above equals

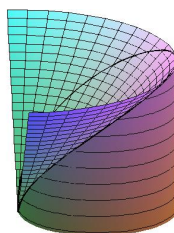
$$\int_0^{2\pi} \int_0^1 \int_{4(r^2 \cos^2(\theta)) + 2(r^2 \sin^2(\theta))}^{4-2(r^2 \sin^2(\theta))} dz r dr d\theta = 2\pi$$

Note this is really not much different than simply using polar coordinates to integrate the difference of the two values of z . This is

$$\begin{aligned} \int_D 4 - 2y^2 - (4x^2 + 2y^2) dA &= \int_D (4 - 4r^2) dA \\ &= \int_0^{2\pi} \int_0^1 (4 - 4r^2) r dr d\theta = 2\pi \end{aligned}$$

Here is another example.

Example 25.3.2 Find the volume of the three dimensional region between the graphs of $z = 0$, $z = \sqrt{x^2 + y^2}$, and the cylinder $(x - 1)^2 + y^2 = 1$.

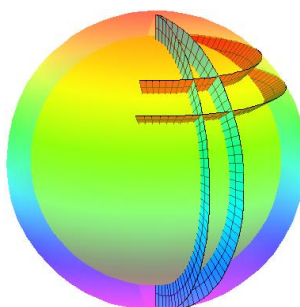


Consider the cylinder. It reduces to $r^2 = 2r \cos \theta$ or more simply $r = 2 \cos \theta$. This is the graph of a circle having radius 1 and centered at $(1, 0)$. Therefore, $\theta \in [-\pi/2, \pi/2]$. It follows that the cylindrical coordinate description of this volume is

$$\int_{-\pi/2}^{\pi/2} \int_0^{2 \cos \theta} \int_0^r dz r dr d\theta = \frac{32}{9}$$

You need to always worry about the limits of the integrals when you do these problems.

What is the increment of volume in spherical coordinates? There are two ways to see what this is, through art and through a systematic procedure. First consider art. Here is a picture.



In the picture there are two concentric spheres formed by making ρ two different constants and surfaces which correspond to θ assuming two different constants and ϕ assuming two different constants. These intersecting surfaces form the little box in the picture. What is the volume of this little box? Length $\approx \rho d\phi$, width $\approx \rho \sin(\phi) d\theta$, height $\approx d\rho$ and so the volume increment for spherical coordinates is

$$dV = \rho^2 \sin(\phi) d\rho d\theta d\phi$$

Now what is really going on? Consider the lower left corner of the little box. Fixing θ and ϕ at their values at this point and differentiating with respect to ρ leads to a little vector of the form

$$\begin{pmatrix} \sin(\phi) \cos(\theta) \\ \sin(\phi) \sin(\theta) \\ \cos(\phi) \end{pmatrix} d\rho$$

which points out from the surface of the sphere. Next keeping ρ and θ constant and differentiating only with respect to ϕ leads to an infinitesimal vector in the direction of a line of longitude,

$$\begin{pmatrix} \rho \cos(\phi) \cos(\theta) \\ \rho \cos(\phi) \sin(\theta) \\ -\rho \sin(\phi) \end{pmatrix} d\phi$$

and finally keeping ρ and ϕ constant and differentiating with respect to θ leads to the third infinitesimal vector which points in the direction of a line of latitude.

$$\begin{pmatrix} -\rho \sin(\phi) \sin(\theta) \\ \rho \sin(\phi) \cos(\theta) \\ 0 \end{pmatrix} d\theta$$

To find the increment of volume, we just need to take the absolute value of the determinant which has these vectors as columns, (Remember this is the absolute value of the box product.) exactly as was the case for polar coordinates. This will also yield

$$dV = \rho^2 \sin(\phi) d\rho d\theta d\phi.$$

However, in contrast to the drawing of pictures, this procedure is completely general and will handle all curvilinear coordinate systems and in any dimension. This is discussed more later.

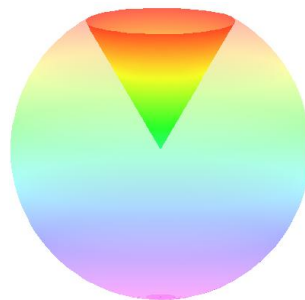
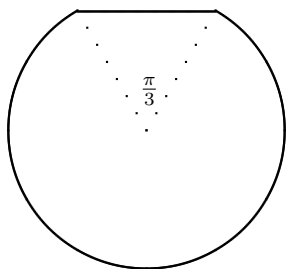
Example 25.3.3 Find the volume of a ball, B_R of radius R .

In this case, $U = (0, R] \times [0, \pi] \times [0, 2\pi)$ and use spherical coordinates. Then 25.4 yields a set in $\mathbb{R}^3$ which clearly differs from the ball of radius R only by a set having volume equal to zero. It leaves out the point at the origin is all. Therefore, the volume of the ball is

$$\begin{aligned} \int_{B_R} 1 dV &= \int_U \rho^2 \sin \phi dV \\ &= \int_0^R \int_0^\pi \int_0^{2\pi} \rho^2 \sin \phi d\theta d\phi d\rho = \frac{4}{3} R^3 \pi. \end{aligned}$$

The reason this was effortless, is that the ball, B_R is realized as a box in terms of the spherical coordinates. Remember what was pointed out earlier about setting up iterated integrals over boxes.

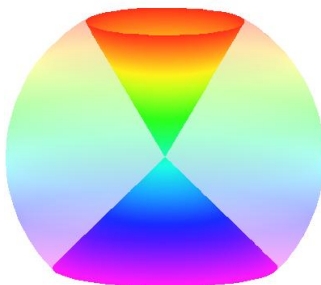
Example 25.3.4 A cone is cut out of a ball of radius R as shown in the following picture, the diagram on the left being a side view. The angle of the cone is $\pi/3$. Find the volume of what is left.



Use spherical coordinates. This volume is then

$$\int_{\pi/6}^{\pi} \int_0^{2\pi} \int_0^R \rho^2 \sin(\phi) d\rho d\theta d\phi = \frac{2}{3}\pi R^3 + \frac{1}{3}\sqrt{3}\pi R^3$$

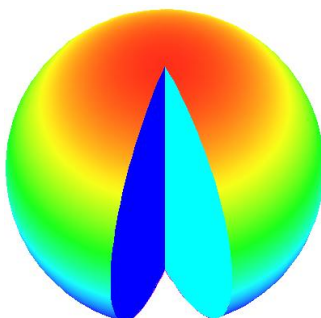
Now change the example a little by cutting out a cone at the bottom which has an angle of $\pi/2$ as shown. What is the volume of what is left?



This time you would have the volume equals

$$\int_{\pi/6}^{3\pi/4} \int_0^{2\pi} \int_0^R \rho^2 \sin(\phi) d\rho d\theta d\phi = \frac{1}{3}\sqrt{2}\pi R^3 + \frac{1}{3}\sqrt{3}\pi R^3$$

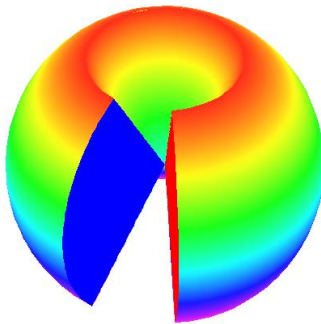
Example 25.3.5 Next suppose the ball of radius R is a sort of an orange and you remove a slice as shown in the picture. What is the volume of what is left? Assume the slice is formed by the two half planes $\theta = 0$ and $\theta = \pi/4$.



Using spherical coordinates, this gives for the volume

$$\int_0^{\pi} \int_{\pi/4}^{2\pi} \int_0^R \rho^2 \sin(\phi) d\rho d\theta d\phi = \frac{7}{6}\pi R^3$$

Example 25.3.6 Now remove the same two cones as in the above examples along with the same slice and find the volume of what is left.

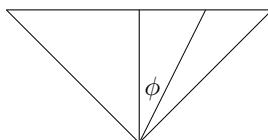


This time you need

$$\int_{\pi/6}^{3\pi/4} \int_{\pi/4}^{2\pi} \int_0^R \rho^2 \sin(\phi) d\rho d\theta d\phi = \frac{7}{24}\sqrt{2}\pi R^3 + \frac{7}{24}\sqrt{3}\pi R^3$$

Example 25.3.7 Set up the integrals to find the volume of the cone $0 \leq z \leq 4, z = \sqrt{x^2 + y^2}$.

This is entirely the wrong coordinate system to use for this problem but it is a good exercise. Here is a side view.



You need to figure out what ρ is as a function of ϕ which goes from 0 to $\pi/4$. You should get

$$\int_0^{2\pi} \int_0^{\pi/4} \int_0^{4 \sec(\phi)} \rho^2 \sin(\phi) d\rho d\phi d\theta = \frac{64}{3}\pi$$

Example 25.3.8 Find the volume element for cylindrical coordinates.

In cylindrical coordinates,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r \cos \theta \\ r \sin \theta \\ z \end{pmatrix}$$

Therefore, the Jacobian determinant is

$$\det \begin{pmatrix} \cos \theta & -r \sin \theta & 0 \\ \sin \theta & r \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} = r.$$

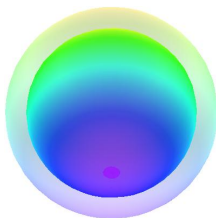
It follows the volume element in cylindrical coordinates is $r d\theta dr dz$.

Example 25.3.9 In the cone of Example 25.3.7 set up the integrals for finding the volume in cylindrical coordinates.

This is a better coordinate system for this example than spherical coordinates. This time you should get

$$\int_0^{2\pi} \int_0^4 \int_r^4 r dz dr d\theta = \frac{64}{3}\pi$$

Example 25.3.10 *This example uses spherical coordinates to verify an important conclusion about gravitational force. Let the hollow sphere, H be defined by $a^2 < x^2 + y^2 + z^2 < b^2$*



and suppose this hollow sphere has constant density taken to equal 1. Now place a unit mass at the point $(0, 0, z_0)$ where $|z_0| \in [a, b]$. Show that the force of gravity acting on this unit mass is $\left(\alpha G \int_H \frac{(z - z_0)}{[x^2 + y^2 + (z - z_0)^2]^{3/2}} dV \right) \mathbf{k}$ and then show that if $|z_0| > b$ then the force of gravity acting on this point mass is the same as if the entire mass of the hollow sphere were placed at the origin, while if $|z_0| < a$, the total force acting on the point mass from gravity equals zero. Here G is the gravitation constant and α is the density. In particular, this shows that the force a planet exerts on an object is as though the entire mass of the planet were situated at its center¹.

Without loss of generality, assume $z_0 > 0$. Let dV be a little chunk of material located at the point (x, y, z) of H the hollow sphere. Then according to Newton's law of gravity, the force this small chunk of material exerts on the given point mass equals

$$\frac{x\mathbf{i} + y\mathbf{j} + (z - z_0)\mathbf{k}}{|x\mathbf{i} + y\mathbf{j} + (z - z_0)\mathbf{k}|} \frac{1}{(x^2 + y^2 + (z - z_0)^2)^{3/2}} G\alpha dV =$$

$$(x\mathbf{i} + y\mathbf{j} + (z - z_0)\mathbf{k}) \frac{1}{(x^2 + y^2 + (z - z_0)^2)^{3/2}} G\alpha dV$$

Therefore, the total force is

$$\int_H (x\mathbf{i} + y\mathbf{j} + (z - z_0)\mathbf{k}) \frac{1}{(x^2 + y^2 + (z - z_0)^2)^{3/2}} G\alpha dV.$$

By the symmetry of the sphere, the $\mathbf{i}$ and $\mathbf{j}$ components will cancel out when the integral is taken. This is because there is the same amount of stuff for negative x and y as there is for positive x and y . Hence what remains is

$$\alpha G \mathbf{k} \int_H \frac{(z - z_0)}{[x^2 + y^2 + (z - z_0)^2]^{3/2}} dV$$

as claimed. Now for the interesting part, the integral is evaluated. In spherical coordinates this integral is

$$\int_0^{2\pi} \int_a^b \int_0^\pi \frac{(\rho \cos \phi - z_0) \rho^2 \sin \phi}{(\rho^2 + z_0^2 - 2\rho z_0 \cos \phi)^{3/2}} d\phi d\rho d\theta. \quad (25.1)$$

Rewrite the inside integral and use integration by parts to obtain this inside integral equals

$$\frac{1}{2z_0} \int_0^\pi (\rho^2 \cos \phi - \rho z_0) \frac{(2z_0 \rho \sin \phi)}{(\rho^2 + z_0^2 - 2\rho z_0 \cos \phi)^{3/2}} d\phi =$$

¹This was shown by Newton in 1685 and allowed him to assert his law of gravitation applied to the planets as though they were point masses. It was a major accomplishment.

$$\begin{aligned} & \frac{1}{2z_0} \left(-2 \frac{-\rho^2 - \rho z_0}{\sqrt{(\rho^2 + z_0^2 + 2\rho z_0)}} + 2 \frac{\rho^2 - \rho z_0}{\sqrt{(\rho^2 + z_0^2 - 2\rho z_0)}} \right. \\ & \quad \left. - \int_0^\pi 2\rho^2 \frac{\sin \phi}{\sqrt{(\rho^2 + z_0^2 - 2\rho z_0 \cos \phi)}} d\phi \right). \end{aligned} \quad (25.2)$$

There are some cases to consider here.

First suppose $z_0 < a$ so the point is on the inside of the hollow sphere and it is always the case that $\rho > z_0$. Then in this case, the two first terms reduce to

$$\frac{2\rho(\rho + z_0)}{\sqrt{(\rho + z_0)^2}} + \frac{2\rho(\rho - z_0)}{\sqrt{(\rho - z_0)^2}} = \frac{2\rho(\rho + z_0)}{(\rho + z_0)} + \frac{2\rho(\rho - z_0)}{\rho - z_0} = 4\rho$$

and so the expression in 25.2 equals

$$\begin{aligned} & \frac{1}{2z_0} \left(4\rho - \int_0^\pi 2\rho^2 \frac{\sin \phi}{\sqrt{(\rho^2 + z_0^2 - 2\rho z_0 \cos \phi)}} d\phi \right) \\ &= \frac{1}{2z_0} \left(4\rho - \frac{1}{z_0} \int_0^\pi \rho \frac{2\rho z_0 \sin \phi}{\sqrt{(\rho^2 + z_0^2 - 2\rho z_0 \cos \phi)}} d\phi \right) \\ &= \frac{1}{2z_0} \left(4\rho - \frac{2\rho}{z_0} (\rho^2 + z_0^2 - 2\rho z_0 \cos \phi)^{1/2} \Big|_0^\pi \right) \\ &= \frac{1}{2z_0} \left(4\rho - \frac{2\rho}{z_0} [(\rho + z_0) - (\rho - z_0)] \right) = 0. \end{aligned}$$

Therefore, in this case the inner integral of 25.1 equals zero and so the original integral will also be zero.

The other case is when $z_0 > b$ and so it is always the case that $z_0 > \rho$. In this case the first two terms of 25.2 are

$$\frac{2\rho(\rho + z_0)}{\sqrt{(\rho + z_0)^2}} + \frac{2\rho(\rho - z_0)}{\sqrt{(\rho - z_0)^2}} = \frac{2\rho(\rho + z_0)}{(\rho + z_0)} + \frac{2\rho(\rho - z_0)}{z_0 - \rho} = 0.$$

Therefore in this case, 25.2 equals

$$\begin{aligned} & \frac{1}{2z_0} \left(- \int_0^\pi 2\rho^2 \frac{\sin \phi}{\sqrt{(\rho^2 + z_0^2 - 2\rho z_0 \cos \phi)}} d\phi \right) \\ &= \frac{-\rho}{2z_0^2} \left(\int_0^\pi \frac{2\rho z_0 \sin \phi}{\sqrt{(\rho^2 + z_0^2 - 2\rho z_0 \cos \phi)}} d\phi \right) \end{aligned}$$

which equals

$$\begin{aligned} & \frac{-\rho}{z_0^2} \left((\rho^2 + z_0^2 - 2\rho z_0 \cos \phi)^{1/2} \Big|_0^\pi \right) \\ &= \frac{-\rho}{z_0^2} [(\rho + z_0) - (z_0 - \rho)] = -\frac{2\rho^2}{z_0^2}. \end{aligned}$$

Thus the inner integral of 25.1 reduces to the above simple expression. Therefore, 25.1 equals

$$\int_0^{2\pi} \int_a^b \left(-\frac{2}{z_0^2} \rho^2 \right) d\rho d\theta = -\frac{4}{3} \pi \frac{b^3 - a^3}{z_0^2}$$

and so

$$\begin{aligned} & \alpha G \mathbf{k} \int_H \frac{(z - z_0)}{[x^2 + y^2 + (z - z_0)^2]^{3/2}} dV \\ &= \alpha G \mathbf{k} \left(-\frac{4}{3} \pi \frac{b^3 - a^3}{z_0^2} \right) = -\mathbf{k} G \frac{\text{total mass}}{z_0^2}. \end{aligned}$$

25.4 Exercises

1. Find the volume of the region bounded by $z = 0$, $x^2 + (y - 2)^2 = 4$, and $z = \sqrt{x^2 + y^2}$.
2. Find the volume of the region, $z \geq 0$, $x^2 + y^2 \leq 4$, and $z \leq 4 - \sqrt{x^2 + y^2}$.
3. Find the volume of the region which is between the surfaces $z = 5y^2 + 9x^2$ and $z = 9 - 4y^2$.
4. Find the volume of the region which is between $z = x^2 + y^2$ and $z = 5 - 4x$. **Hint:** You might want to change variables at some point.
5. The ice cream in a sugar cone is described in spherical coordinates by $\rho \in [0, 10]$, $\phi \in [0, \frac{1}{3}\pi]$, $\theta \in [0, 2\pi]$. If the units are in centimeters, find the total volume in cubic centimeters of this ice cream.
6. Find the volume between $z = 3 - x^2 - y^2$ and $z = 2\sqrt{(x^2 + y^2)}$.
7. A ball of radius 3 is placed in a drill press and a hole of radius 2 is drilled out with the center of the hole a diameter of the ball. What is the volume of the material which remains?
8. Find the volume of the cone defined by $z \in [0, 4]$ having angle $\pi/2$. Use spherical coordinates.
9. A ball of radius 9 has density equal to $\sqrt{x^2 + y^2 + z^2}$ in rectangular coordinates. The top of this ball is sliced off by a plane of the form $z = 2$. Write integrals for the mass of what is left. In spherical coordinates and in cylindrical coordinates.
10. A ball of radius 4 has a cone taken out of the top which has an angle of $\pi/2$ and then a cone taken out of the bottom which has an angle of $\pi/3$. Then a slice, $\theta \in [0, \pi/4]$ is removed. What is the volume of what is left?
11. In Example 25.3.10 on Page 568 check out all the details by working the integrals to be sure the steps are right.
12. What if the hollow sphere in Example 25.3.10 were in two dimensions and everything, including Newton's law still held? Would similar conclusions hold? Explain.

25.5 The General Procedure

As mentioned above, the fundamental concept of an integral is a sum of things of the form $f(\mathbf{x}) dV$ where dV is an “infinitesimal” chunk of volume located at the point, $\mathbf{x}$. Up to now, this infinitesimal chunk of volume has had the form of a box with sides $dx_1, \dots, dx_n$ so $dV = dx_1 dx_2 \cdots dx_n$ but its form is not important. It could just as well be an infinitesimal parallelepiped for example. In what follows, this is what it will be.

First recall the following fundamental definition on Page 394.

Definition 25.5.1 Let $\mathbf{u}_1, \dots, \mathbf{u}_p$ be vectors in $\mathbb{R}^k$. The parallelepiped determined by these vectors will be denoted by $P(\mathbf{u}_1, \dots, \mathbf{u}_p)$ and it is defined as

$$P(\mathbf{u}_1, \dots, \mathbf{u}_p) \equiv \left\{ \sum_{j=1}^p s_j \mathbf{u}_j : s_j \in [0, 1] \right\}.$$

Now define the volume of this parallelepiped.

$$\text{volume of } P(\mathbf{u}_1, \dots, \mathbf{u}_p) \equiv (\det(\mathbf{u}_i \cdot \mathbf{u}_j))^{1/2}.$$

The dot product is used to determine this volume of a parallelepiped spanned by the given vectors and you should note that it is only the dot product that matters. Now consider spherical coordinates, ρ, ϕ , and θ . Suppose the relationship between coordinates $\mathbf{u}$ and Cartesian coordinates $\mathbf{x} = (x, y, z)$ is

$$x = f_1(u_1, u_2, u_3), y = f_2(u_1, u_2, u_3), z = f_3(u_1, u_2, u_3) \quad (25.3)$$

where $\mathbf{u} \in U$ an open set in $\mathbb{R}^3$ and corresponding to such a $\mathbf{u} \in U$ there exists a unique point, $(x, y, z) \in V$ as above. Suppose at the point $\mathbf{u}_0 \in U$, there is an infinitesimal box having sides du_1, du_2, du_3 . Then this little box would correspond to something in V . What? Consider the mapping from U to V defined by

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} f_1(u_1, u_2, u_3) \\ f_2(u_1, u_2, u_3) \\ f_3(u_1, u_2, u_3) \end{pmatrix} = \mathbf{f}(\mathbf{u}) \quad (25.4)$$

which takes a point $\mathbf{u}$ in U and sends it to the point in V which is identified as $(x, y, z)^T \equiv \mathbf{x}$. What happens to a point of the infinitesimal box? Such a point is of the form

$$(u_{01} + s_1 du_1, u_{02} + s_2 du_2, u_{03} + s_3 du_3),$$

where $s_i \geq 0$ and $\sum_i s_i \leq 1$. Also, from the definition of the derivative,

$$\begin{aligned} \mathbf{f}(u_{01} + s_1 du_1, u_{02} + s_2 du_2, u_{03} + s_3 du_3) - \mathbf{f}(u_{01}, u_{02}, u_{03}) = \\ D\mathbf{f}(u_{01}, u_{02}, u_{03}) \begin{pmatrix} s_1 du_1 \\ s_2 du_2 \\ s_3 du_3 \end{pmatrix} + \mathbf{o} \begin{pmatrix} s_1 du_1 \\ s_2 du_2 \\ s_3 du_3 \end{pmatrix} \end{aligned}$$

where the last term may be taken equal to $\mathbf{0}$ because the vector, $(s_1 du_1, s_2 du_2, s_3 du_3)^T$ is infinitesimal, meaning nothing precise, but conveying the idea that it is surpassingly small. Therefore, a point of this infinitesimal box is sent to the vector,

$$\begin{aligned} & \overbrace{\left(\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1}, \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_2}, \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_3} \right)}^{=D\mathbf{f}(u_{01}, u_{02}, u_{03})} \begin{pmatrix} s_1 d\rho \\ s_2 d\phi \\ s_3 d\theta \end{pmatrix} = \\ & s_1 \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1} du_1 + s_2 \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_2} du_2 + s_3 \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_3} du_3, \end{aligned}$$

a point of the infinitesimal parallelepiped determined by the vectors

$$\left\{ \frac{\partial \mathbf{x}(u_{01}, u_{02}, u_{03})}{\partial u_1} du_1, \frac{\partial \mathbf{x}(u_{01}, u_{02}, u_{03})}{\partial u_2} du_2, \frac{\partial \mathbf{x}(u_{01}, u_{02}, u_{03})}{\partial u_3} du_3 \right\}.$$

The situation is no different for general coordinate systems in any dimension. In general, $\mathbf{x} = \mathbf{f}(\mathbf{u})$ where $\mathbf{u} \in U$, a subset of $\mathbb{R}^n$ and $\mathbf{x}$ is a point in V , a subset of n dimensional space. Thus, letting the Cartesian coordinates of $\mathbf{x}$ be given by $\mathbf{x} = (x_1, \dots, x_n)^T$, each x_i being a function of $\mathbf{u}$, an infinitesimal box located at $\mathbf{u}_0$ corresponds to an infinitesimal parallelepiped located at $\mathbf{f}(\mathbf{u}_0)$ which is determined by the n vectors $\left\{ \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_i} du_i \right\}_{i=1}^n$. From Definition 25.5.1, the volume of this infinitesimal parallelepiped located at $\mathbf{f}(\mathbf{u}_0)$ is given by

$$\det \left(\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_i} du_i \cdot \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_j} du_j \right)^{1/2} \quad (25.5)$$

in which there is no sum on the repeated index. Now in general, if there are n vectors in $\mathbb{R}^n$, $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$,

$$\det(\mathbf{v}_i \cdot \mathbf{v}_j)^{1/2} = |\det(\mathbf{v}_1, \dots, \mathbf{v}_n)| \quad (25.6)$$

where this last matrix is the $n \times n$ matrix which has the i^{th} column equal to $\mathbf{v}_i$. The reason for this is that the matrix whose ij^{th} entry is $\mathbf{v}_i \cdot \mathbf{v}_j$ is just the product of the two matrices,

$$\begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{pmatrix} (\mathbf{v}_1, \dots, \mathbf{v}_n)$$

where the first on the left is the matrix having the i^{th} row equal to $\mathbf{v}_i^T$ while the matrix on the right is just the matrix having the i^{th} column equal to $\mathbf{v}_i$. Therefore, since the determinant of a matrix equals the determinant of its transpose,

$$\begin{aligned} \det(\mathbf{v}_i \cdot \mathbf{v}_j) &= \det \left(\begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{pmatrix} (\mathbf{v}_1, \dots, \mathbf{v}_n) \right) \\ &= \det(\mathbf{v}_1, \dots, \mathbf{v}_n)^2 \end{aligned}$$

and so taking square roots yields 25.6. Therefore, from the properties of determinants, 25.5 equals

$$\begin{aligned} &\left| \det \left(\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1} du_1, \dots, \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_n} du_n \right) \right| = \\ &\left| \det \left(\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1}, \dots, \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_n} \right) \right| du_1 \cdots du_n \end{aligned}$$

and this is the infinitesimal chunk of volume corresponding to the point $\mathbf{f}(\mathbf{u}_0)$ in V .

Definition 25.5.2 Let $\mathbf{x} = \mathbf{f}(\mathbf{u})$ be as described above. Then the symbol,

$$\frac{\partial(x_1, \dots, x_n)}{\partial(u_1, \dots, u_n)},$$

called the *Jacobian determinant*, is defined by

$$\det \left(\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1}, \dots, \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_n} \right) \equiv \frac{\partial(x_1, \dots, x_n)}{\partial(u_1, \dots, u_n)}.$$

Also, the symbol, $\left| \frac{\partial(x_1, \dots, x_n)}{\partial(u_1, \dots, u_n)} \right| du_1 \cdots du_n$ is called the *volume element* or *increment of volume* or *increment of area*.

This has given motivation for the following fundamental procedure often called the **change of variables formula** which holds under fairly general conditions.

Procedure 25.5.3 Suppose U is an open subset of $\mathbb{R}^n$ for $n > 0$ and suppose $\mathbf{f} : U \rightarrow \mathbf{f}(U)$ is a C^1 function which is one to one, $\mathbf{x} = \mathbf{f}(\mathbf{u})$.² Then if $h : \mathbf{f}(U) \rightarrow \mathbb{R}$,

$$\int_U h(\mathbf{f}(\mathbf{u})) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(u_1, \dots, u_n)} \right| dV = \int_{\mathbf{f}(U)} h(\mathbf{x}) dV.$$

25.6 Exercises

1. Verify the three dimensional volume increment in spherical coordinates is

$$\rho^2 \sin(\phi) d\rho d\phi d\theta.$$

2. Find the area of the bounded region, R , determined by $5x + y = 1$, $5x + y = 9$, $y = 2x$, and $y = 5x$.
3. Find the area of the bounded region, R , determined by $y + 2x = 6$, $y + 2x = 10$, $y = 3x$, and $y = 4x$.
4. A solid, R is determined by $3x + y = 2$, $3x + y = 4$, $y = x$, and $y = 2x$ and the density is $\rho = x$. Find the total mass of R .
5. A solid, R is determined by $4x + 2y = 1$, $4x + 2y = 9$, $y = x$, and $y = 6x$ and the density is $\rho = y$. Find the total mass of R .
6. A solid, R is determined by $3x + y = 3$, $3x + y = 10$, $y = 3x$, and $y = 5x$ and the density is $\rho = y^{-1}$. Find the total mass of R .
7. Find a 2×2 matrix, A which maps the equilateral triangle having vertices at

$$(0, 0), (1, 0), \text{ and } \left(1/2, \sqrt{3}/2\right)$$

to the triangle having vertices at $(0, 0)$, (a, b) , and (c, d) where (c, d) is not a multiple of (a, b) . Find the area of this last triangle by using the cross product. Next find the area of this triangle using the change of variables formula and the fact that the area of the equilateral triangle is $\frac{\sqrt{3}}{4}$.

8. Find the volume of the region, E , bounded by the ellipsoid, $\frac{1}{4}x^2 + y^2 + z^2 = 1$.
9. Here are three vectors. $(4, 1, 2)^T$, $(5, 0, 2)^T$, and $(3, 1, 3)^T$. These vectors determine a parallelepiped, R , which is occupied by a solid having density $\rho = x$. Find the mass of this solid.
10. Here are three vectors. $(5, 1, 6)^T$, $(6, 0, 6)^T$, and $(4, 1, 7)^T$. These vectors determine a parallelepiped, R , which is occupied by a solid having density $\rho = y$. Find the mass of this solid.

²This will cause non overlapping infinitesimal boxes in U to be mapped to non overlapping infinitesimal parallelepipeds in V .

Also, in the context of the Riemann integral we should say more about the set U in any case the function, h . These conditions are mainly technical however, and since a mathematically respectable treatment will not be attempted for this theorem in this part of the book, I think it best to give a memorable version of it which is essentially correct in all examples of interest.

11. Here are three vectors. $(5, 2, 9)^T$, $(6, 1, 9)^T$, and $(4, 2, 10)^T$. These vectors determine a parallelepiped, R , which is occupied by a solid having density $\rho = y + x$. Find the mass of this solid.
12. Compute the volume of a sphere of radius R using cylindrical coordinates.
13. Fill in all details for the following argument that

$$\int_0^\infty e^{-x^2} dx = \frac{1}{2}\sqrt{\pi}.$$

Let $I = \int_0^\infty e^{-x^2} dx$. Then

$$I^2 = \int_0^\infty \int_0^\infty e^{-(x^2+y^2)} dx dy = \int_0^{\pi/2} \int_0^\infty r e^{-r^2} dr d\theta = \frac{1}{4}\pi$$

from which the result follows.

14. Show that $\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$. Here σ is a positive number called the standard deviation and μ is a number called the mean.
15. Show using Problem 13 $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$. Recall $\Gamma(\alpha) \equiv \int_0^\infty e^{-t} t^{\alpha-1} dt$.
16. Let $p, q > 0$ and define $B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx$. Show that

$$\Gamma(p)\Gamma(q) = B(p, q)\Gamma(p+q).$$

Hint: It is fairly routine if you start with the left side and proceed to change variables.

25.7 The Moment Of Inertia And Center Of Mass

The methods used to evaluate multiple integrals make possible the determination of centers of mass and moments of inertia for solids. This leads to the following definition.

Definition 25.7.1 *Let a solid occupy a region R such that its density is $\rho(\mathbf{x})$ for $\mathbf{x}$ a point in R and let L be a line. For $\mathbf{x} \in R$, let $l(\mathbf{x})$ be the distance from the point, $\mathbf{x}$ to the line L . The moment of inertia of the solid is defined as*

$$I = \int_R l(\mathbf{x})^2 \rho(\mathbf{x}) dV.$$

Letting $(\bar{x}, \bar{y}, \bar{z})$ denote the Cartesian coordinates of the center of mass,

$$\begin{aligned} \bar{x} &= \frac{\int_R x \rho(\mathbf{x}) dV}{\int_R \rho(\mathbf{x}) dV}, \quad \bar{y} = \frac{\int_R y \rho(\mathbf{x}) dV}{\int_R \rho(\mathbf{x}) dV}, \\ \bar{z} &= \frac{\int_R z \rho(\mathbf{x}) dV}{\int_R \rho(\mathbf{x}) dV} \end{aligned}$$

where x, y, z are the Cartesian coordinates of the point at $\mathbf{x}$.

The reason the moment of inertia is of interest has to do with the total kinetic energy of a solid occupying the region R which is rotating about the line L . Suppose its angular velocity is ω . Then the kinetic energy of an infinitesimal chunk of volume located at

point $\mathbf{x}$ is $\frac{1}{2}\rho(\mathbf{x})(l(\mathbf{x})\omega)^2 dV$. Then using an integral to add these up, it follows the total kinetic energy is

$$\frac{1}{2} \int_R \rho(\mathbf{x}) l(\mathbf{x})^2 dV \omega^2 = \frac{1}{2} I \omega^2$$

Thus in the consideration of a rotating body, the moment of inertia takes the place of mass when angular velocity takes the place of speed.

As to the center of mass, its significance is that it gives the point at which the mass will balance. See Page 386 to see this explained with point masses. The only difference is that here the sums need to be replaced with integrals.

Example 25.7.2 *Let a solid occupy the three dimensional region R and suppose the density is ρ . What is the moment of inertia of this solid about the z axis? What is the center of mass?*

Here the little masses would be of the form $\rho(\mathbf{x}) dV$ where $\mathbf{x}$ is a point of R . Therefore, the contribution of this mass to the moment of inertia would be

$$(x^2 + y^2) \rho(\mathbf{x}) dV$$

where the Cartesian coordinates of the point $\mathbf{x}$ are (x, y, z) . Then summing these up as an integral, yields the following for the moment of inertia.

$$\int_R (x^2 + y^2) \rho(\mathbf{x}) dV. \quad (25.7)$$

To find the center of mass, sum up $\mathbf{r}\rho dV$ for the points in R and divide by the total mass. In Cartesian coordinates, where $\mathbf{r} = (x, y, z)$, this means to sum up vectors of the form $(x\rho dV, y\rho dV, z\rho dV)$ and divide by the total mass. Thus the Cartesian coordinates of the center of mass are

$$\left(\frac{\int_R x\rho dV}{\int_R \rho dV}, \frac{\int_R y\rho dV}{\int_R \rho dV}, \frac{\int_R z\rho dV}{\int_R \rho dV} \right) = \frac{\int_R \mathbf{r}\rho dV}{\int_R \rho dV}.$$

Here is a specific example.

Example 25.7.3 *Find the moment of inertia about the z axis and center of mass of the solid which occupies the region, R defined by $9 - (x^2 + y^2) \geq z \geq 0$ if the density is $\rho(x, y, z) = \sqrt{x^2 + y^2}$.*

This moment of inertia is $\int_R (x^2 + y^2) \sqrt{x^2 + y^2} dV$ and the easiest way to find this integral is to use cylindrical coordinates. Thus the answer is

$$\int_0^{2\pi} \int_0^3 \int_0^{9-r^2} r^3 r dz dr d\theta = \frac{8748}{35} \pi.$$

To find the center of mass, note the x and y coordinates of the center of mass,

$$\frac{\int_R x\rho dV}{\int_R \rho dV}, \frac{\int_R y\rho dV}{\int_R \rho dV}$$

both equal zero because the above shape is symmetric about the z axis and ρ is also symmetric in its values. Thus $x\rho dV$ will cancel with $-x\rho dV$ and a similar conclusion will hold for the y coordinate. It only remains to find the z coordinate of the center of mass, $\bar{z}$. In polar coordinates, $\rho = r$ and so,

$$\bar{z} = \frac{\int_R z\rho dV}{\int_R \rho dV} = \frac{\int_0^{2\pi} \int_0^3 \int_0^{9-r^2} zr^2 dz dr d\theta}{\int_0^{2\pi} \int_0^3 \int_0^{9-r^2} r^2 dz dr d\theta} = \frac{18}{7}.$$

Thus the center of mass will be $(0, 0, \frac{18}{7})$.

25.8 Exercises

1. Let R denote the finite region bounded by $z = 4 - x^2 - y^2$ and the xy plane. Find z_c , the z coordinate of the center of mass if the density, σ is a constant.
2. Let R denote the finite region bounded by $z = 4 - x^2 - y^2$ and the xy plane. Find z_c , the z coordinate of the center of mass if the density, σ is equals $\sigma(x, y, z) = z$.
3. Find the mass and center of mass of the region between the surfaces $z = -y^2 + 8$ and $z = 2x^2 + y^2$ if the density equals $\sigma = 1$.
4. Find the mass and center of mass of the region between the surfaces $z = -y^2 + 8$ and $z = 2x^2 + y^2$ if the density equals $\sigma(x, y, z) = x^2$.
5. The two cylinders, $x^2 + y^2 = 4$ and $y^2 + z^2 = 4$ intersect in a region, R . Find the mass and center of mass if the density, σ , is given by $\sigma(x, y, z) = z^2$.
6. The two cylinders, $x^2 + y^2 = 4$ and $y^2 + z^2 = 4$ intersect in a region, R . Find the mass and center of mass if the density, σ , is given by $\sigma(x, y, z) = 4 + z$.
7. Find the mass and center of mass of the set, (x, y, z) such that $\frac{x^2}{4} + \frac{y^2}{9} + z^2 \leq 1$ if the density is $\sigma(x, y, z) = 4 + y + z$.
8. Let R denote the finite region bounded by $z = 9 - x^2 - y^2$ and the xy plane. Find the moment of inertia of this shape about the z axis given the density equals 1.
9. Let R denote the finite region bounded by $z = 9 - x^2 - y^2$ and the xy plane. Find the moment of inertia of this shape about the x axis given the density equals 1.
10. Let B be a solid ball of constant density and radius R . Find the moment of inertia about a line through a diameter of the ball. You should get $\frac{2}{5}R^2M$ where M is the mass..
11. Let B be a solid ball of density, $\sigma = \rho$ where ρ is the distance to the center of the ball which has radius R . Find the moment of inertia about a line through a diameter of the ball. Write your answer in terms of the total mass and the radius as was done in the constant density case.
12. Let C be a solid cylinder of constant density and radius R . Find the moment of inertia about the axis of the cylinder
You should get $\frac{1}{2}R^2M$ where M is the mass.
13. Let C be a solid cylinder of constant density and radius R and mass M and let B be a solid ball of radius R and mass M . The cylinder and the ball are placed on the top of an inclined plane and allowed to roll to the bottom. Which one will arrive first and why?
14. A ball of radius 4 has a cone taken out of the top which has an angle of $\pi/2$ and then a cone taken out of the bottom which has an angle of $\pi/3$. If the density is $\lambda = \rho$, find the z component of the center of mass.
15. A ball of radius 4 has a cone taken out of the top which has an angle of $\pi/2$ and then a cone taken out of the bottom which has an angle of $\pi/3$. If the density is $\lambda = \rho$, find the moment of inertia about the z axis.

16. Suppose a solid of mass M occupying the region, B has moment of inertia, I_l about a line, l which passes through the center of mass of M and let l_1 be another line parallel to l and at a distance of a from l . Then the parallel axis theorem states $I_{l_1} = I_l + a^2 M$. Prove the parallel axis theorem. **Hint:** Choose axes such that the z axis is l and l_1 passes through the point $(a, 0)$ in the xy plane.
17. * Using the parallel axis theorem find the moment of inertia of a solid ball of radius R and mass M about an axis located at a distance of a from the center of the ball. Your answer should be $Ma^2 + \frac{2}{5}MR^2$.
18. Consider all axes in computing the moment of inertia of a solid. Will the smallest possible moment of inertia always result from using an axis which goes through the center of mass?
19. Find the moment of inertia of a solid thin rod of length l , mass M , and constant density about an axis through the center of the rod perpendicular to the axis of the rod. You should get $\frac{1}{12}l^2 M$.
20. Using the parallel axis theorem, find the moment of inertia of a solid thin rod of length l , mass M , and constant density about an axis through an end of the rod perpendicular to the axis of the rod. You should get $\frac{1}{3}l^2 M$.
21. Let the angle between the z axis and the sides of a right circular cone be α . Also assume the height of this cone is h . Find the z coordinate of the center of mass of this cone in terms of α and h assuming the density is constant.
22. Let the angle between the z axis and the sides of a right circular cone be α . Also assume the height of this cone is h . Assuming the density is $\sigma = 1$, find the moment of inertia about the z axis in terms of α and h .
23. Let R denote the part of the solid ball, $x^2 + y^2 + z^2 \leq R^2$ which lies in the first octant. That is $x, y, z \geq 0$. Find the coordinates of the center of mass if the density is constant. Your answer for one of the coordinates for the center of mass should be $(3/8)R$.
24. Show that in general for $\mathbf{L}$ angular momentum,

$$\frac{d\mathbf{L}}{dt} = \mathbf{\Gamma}$$

where $\mathbf{\Gamma}$ is the total torque,

$$\mathbf{\Gamma} \equiv \sum \mathbf{r}_i \times \mathbf{F}_i$$

where $\mathbf{F}_i$ is the force on the i^{th} point mass.

The Integral On Two Dimensional Surfaces In $\mathbb{R}^3$

26.0.1 Outcomes

1. Find the area of a surface.
2. Define and compute integrals over surfaces given parametrically.

26.1 The Two Dimensional Area In $\mathbb{R}^3$

Consider the boundary of some three dimensional region such that a function, f is defined on this boundary. Imagine taking the value of this function at a point, multiplying this value by the area of an infinitesimal chunk of area located at this point and then adding these up. This is just the notion of the integral presented earlier only now there is a difference because this infinitesimal chunk of area should be considered as two dimensional even though it is in three dimensions. However, it is not really all that different from what was done earlier. It all depends on the following fundamental definition which is just a review of the fact presented earlier that the area of a parallelogram determined by two vectors in $\mathbb{R}^3$ is the norm of the cross product of the two vectors.

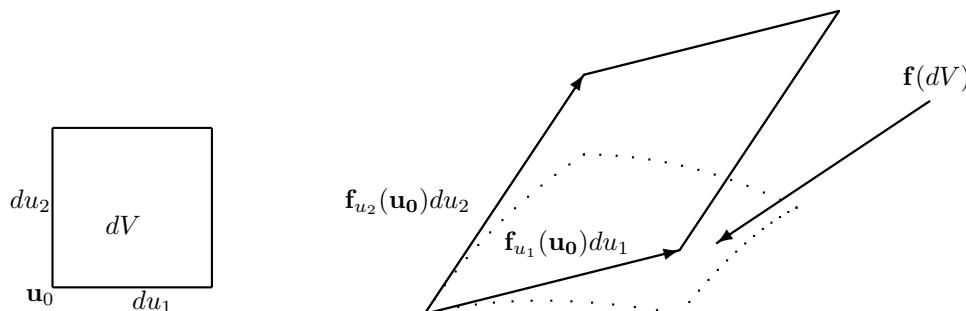
Definition 26.1.1 *Let $\mathbf{u}_1, \mathbf{u}_2$ be vectors in $\mathbb{R}^3$. The 2 dimensional parallelogram determined by these vectors will be denoted by $P(\mathbf{u}_1, \mathbf{u}_2)$ and it is defined as*

$$P(\mathbf{u}_1, \mathbf{u}_2) \equiv \left\{ \sum_{j=1}^2 s_j \mathbf{u}_j : s_j \in [0, 1] \right\}.$$

Then the area of this parallelogram is

$$\text{area } P(\mathbf{u}_1, \mathbf{u}_2) \equiv |\mathbf{u}_1 \times \mathbf{u}_2|.$$

Suppose then that $\mathbf{x} = \mathbf{f}(\mathbf{u})$ where $\mathbf{u} \in U$, a subset of $\mathbb{R}^2$ and $\mathbf{x}$ is a point in V , a subset of 3 dimensional space. Thus, letting the Cartesian coordinates of $\mathbf{x}$ be given by $\mathbf{x} = (x_1, x_2, x_3)^T$, each x_i being a function of $\mathbf{u}$, an infinitesimal rectangle located at $\mathbf{u}_0$ corresponds to an infinitesimal parallelogram located at $\mathbf{f}(\mathbf{u}_0)$ which is determined by the 2 vectors $\left\{ \frac{\partial \mathbf{f}(\mathbf{u}_0)}{\partial u_i} du_i \right\}_{i=1}^2$, each of which is tangent to the surface defined by $\mathbf{x} = \mathbf{f}(\mathbf{u})$. (No sum on the repeated index.)



From Definition 26.1.1, the two dimensional volume of this infinitesimal parallelepiped located at $\mathbf{f}(\mathbf{u}_0)$ is given by

$$\left| \frac{\partial \mathbf{f}(\mathbf{u}_0)}{\partial u_1} du_1 \times \frac{\partial \mathbf{f}(\mathbf{u}_0)}{\partial u_2} du_2 \right| = \left| \frac{\partial \mathbf{f}(\mathbf{u}_0)}{\partial u_1} \times \frac{\partial \mathbf{f}(\mathbf{u}_0)}{\partial u_2} \right| du_1 du_2 \quad (26.1)$$

$$= |\mathbf{f}_{u_1} \times \mathbf{f}_{u_2}| du_1 du_2 \quad (26.2)$$

It might help to think of a lizard. The infinitesimal parallelepiped is like a very small scale on a lizard. This is the essence of the idea. To define the area of the lizard sum up areas of individual scales¹. If the scales are small enough, their sum would serve as a good approximation to the area of the lizard.



This motivates the following fundamental procedure which I hope is extremely familiar from the earlier material.

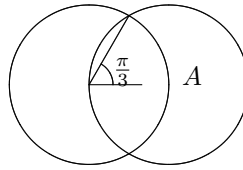
Procedure 26.1.2 Suppose U is a subset of $\mathbb{R}^2$ and suppose $\mathbf{f} : U \rightarrow \mathbf{f}(U) \subseteq \mathbb{R}^3$ is a one to one and C^1 function. Then if $h : \mathbf{f}(U) \rightarrow \mathbb{R}$, define the 2 dimensional surface integral, $\int_{\mathbf{f}(U)} h(\mathbf{x}) dA$ according to the following formula.

$$\int_{\mathbf{f}(U)} h(\mathbf{x}) dA \equiv \int_U h(\mathbf{f}(\mathbf{u})) |\mathbf{f}_{u_1}(\mathbf{u}) \times \mathbf{f}_{u_2}(\mathbf{u})| du_1 du_2.$$

Definition 26.1.3 It is customary to write $|\mathbf{f}_{u_1}(\mathbf{u}) \times \mathbf{f}_{u_2}(\mathbf{u})| = \frac{\partial(x_1, x_2, x_3)}{\partial(u_1, u_2)}$ because this new notation generalizes to far more general situations for which the cross product is not defined. For example, one can consider three dimensional surfaces in $\mathbb{R}^8$.

¹This beautiful lizard is a *Sceloporus magister*. It was photographed by C. Riley Nelson who is in the Zoology department at Brigham Young University © 2004 in Kane Co. Utah. The lizard is a little less than one foot in length.

Example 26.1.4 Find the area of the region labeled A in the following picture. The two circles are of radius 1, one has center $(0, 0)$ and the other has center $(1, 0)$.



The circles bounding these disks are $x^2 + y^2 = 1$ and $(x - 1)^2 + y^2 = x^2 + y^2 - 2x + 1 = 1$. Therefore, in polar coordinates these are of the form $r = 1$ and $r = 2 \cos \theta$.

The set A corresponds to the set U , in the (θ, r) plane determined by $\theta \in [-\frac{\pi}{3}, \frac{\pi}{3}]$ and for each value of θ in this interval, r goes from 1 up to $2 \cos \theta$. Therefore, the area of this region is of the form,

$$\int_U 1 dV = \int_{-\pi/3}^{\pi/3} \int_1^{2 \cos \theta} \frac{\partial(x_1, x_2, x_3)}{\partial(\theta, r)} dr d\theta.$$

It is necessary to find $\frac{\partial(x_1, x_2)}{\partial(\theta, r)}$. The mapping $\mathbf{f} : U \rightarrow \mathbb{R}^2$ takes the form $\mathbf{f}(\theta, r) = (r \cos \theta, r \sin \theta)^T$. Here $x_3 = 0$ and so

$$\frac{\partial(x_1, x_2, x_3)}{\partial(\theta, r)} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial x_1}{\partial \theta} & \frac{\partial x_2}{\partial \theta} & \frac{\partial x_3}{\partial \theta} \\ \frac{\partial x_1}{\partial r} & \frac{\partial x_2}{\partial r} & \frac{\partial x_3}{\partial r} \end{vmatrix} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -r \sin \theta & r \cos \theta & 0 \\ \cos \theta & \sin \theta & 0 \end{vmatrix} = r$$

Therefore, the area element is $r dr d\theta$. It follows the desired area is

$$\int_{-\pi/3}^{\pi/3} \int_1^{2 \cos \theta} r dr d\theta = \frac{1}{2} \sqrt{3} + \frac{1}{3} \pi.$$

Example 26.1.5 Consider the surface given by $z = x^2$ for $(x, y) \in [0, 1] \times [0, 1] = U$. Find the surface area of this surface.

The first step in using the above is to write this surface in the form $\mathbf{x} = \mathbf{f}(\mathbf{u})$. This is easy to do if you let $\mathbf{u} = (x, y)$. Then $\mathbf{f}(x, y) = (x, y, x^2)$. If you like, let $x = u_1$ and $y = u_2$. What is $\frac{\partial(x_1, x_2, x_3)}{\partial(x, y)} = |\mathbf{f}_x \times \mathbf{f}_y|$?

$$\mathbf{f}_x = \begin{pmatrix} 1 \\ 0 \\ 2x \end{pmatrix}, \mathbf{f}_y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

and so

$$|\mathbf{f}_x \times \mathbf{f}_y| = \left| \begin{pmatrix} 1 \\ 0 \\ 2x \end{pmatrix} \times \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right| = \sqrt{1 + 4x^2}$$

and so the area element is $\sqrt{1 + 4x^2} dx dy$ and the surface area is obtained by integrating the function, $h(\mathbf{x}) \equiv 1$. Therefore, this area is

$$\int_{\mathbf{f}(U)} dA = \int_0^1 \int_0^1 \sqrt{1 + 4x^2} dx dy = \frac{1}{2} \sqrt{5} - \frac{1}{4} \ln(-2 + \sqrt{5})$$

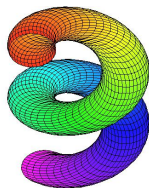
which can be obtained by using the trig. substitution, $2x = \tan \theta$ on the inside integral.

Note this all depends on being able to write the surface in the form, $\mathbf{x} = \mathbf{f}(\mathbf{u})$ for $\mathbf{u} \in U \subseteq \mathbb{R}^p$. Surfaces obtained in this form are called parametrically defined surfaces. These are best but sometimes you have some other description of a surface and in these cases things can get pretty intractable. For example, you might have a level surface of the form $3x^2 + 4y^4 + z^6 = 10$. In this case, you could solve for z using methods of algebra. Thus $z = \sqrt[6]{10 - 3x^2 - 4y^4}$ and a parametric description of part of this level surface is $(x, y, \sqrt[6]{10 - 3x^2 - 4y^4})$ for $(x, y) \in U$ where $U = \{(x, y) : 3x^2 + 4y^4 \leq 10\}$. But what if the level surface was something like

$$\sin(x^2 + \ln(7 + y^2 \sin x)) + \sin(zx)e^z = 11 \sin(xyz)?$$

I really don't see how to use methods of algebra to solve for some variable in terms of the others. It isn't even clear to me whether there are any points $(x, y, z) \in \mathbb{R}^3$ satisfying this particular relation. However, if a point satisfying this relation can be identified, the implicit function theorem from advanced calculus can usually be used to assert one of the variables is a function of the others, proving the existence of a parametrization at least locally. The problem is, this theorem doesn't give the answer in terms of known functions so this isn't much help. Finding a parametric description of a surface is a hard problem and there are no easy answers. This is a good example which illustrates the gulf between theory and practice.

Example 26.1.6 Let $U = [0, 12] \times [0, 2\pi]$ and let $\mathbf{f} : U \rightarrow \mathbb{R}^3$ be given by $\mathbf{f}(t, s) \equiv (2 \cos t + \cos s, 2 \sin t + \sin s, t)^T$. Find a double integral for the surface area. A graph of this surface is drawn below.



It looks like something you would use to make sausages². Anyway,

$$\mathbf{f}_t = \begin{pmatrix} -2 \sin t \\ 2 \cos t \\ 1 \end{pmatrix}, \mathbf{f}_s = \begin{pmatrix} -\sin s \\ \cos s \\ 0 \end{pmatrix}$$

and

$$\mathbf{f}_t \times \mathbf{f}_s = \begin{pmatrix} -\cos s \\ -\sin s \\ -2 \sin t \cos s + 2 \cos t \sin s \end{pmatrix}$$

and so

$$\frac{\partial(x_1, x_2, x_3)}{\partial(t, s)} = |\mathbf{f}_t \times \mathbf{f}_s| = \sqrt{5 - 4 \sin^2 t \sin^2 s - 8 \sin t \sin s \cos t \cos s - 4 \cos^2 t \cos^2 s}.$$

Therefore, the desired integral giving the area is

$$\int_0^{2\pi} \int_0^{12} \sqrt{5 - 4 \sin^2 t \sin^2 s - 8 \sin t \sin s \cos t \cos s - 4 \cos^2 t \cos^2 s} dt ds.$$

If you really needed to find the number this equals, how would you go about finding it? This is an interesting question and there is no single right answer. You should think about this. Here is an example for which you will be able to find the integrals.

²At Volwerth's in Hancock Michigan, they make excellent sausages and hot dogs. The best are made from "natural casings" which are the linings of intestines.

Example 26.1.7 Let $U = [0, 2\pi] \times [0, 2\pi]$ and for $(t, s) \in U$, let

$$\mathbf{f}(t, s) = (2 \cos t + \cos t \cos s, -2 \sin t - \sin t \cos s, \sin s)^T.$$

Find the area of $\mathbf{f}(U)$. This is the surface of a donut shown below. The fancy name for this shape is a torus.



To find its area,

$$\mathbf{f}_t = \begin{pmatrix} -2 \sin t - \sin t \cos s \\ -2 \cos t - \cos t \cos s \\ 0 \end{pmatrix}, \mathbf{f}_s = \begin{pmatrix} -\cos t \sin s \\ \sin t \sin s \\ \cos s \end{pmatrix}$$

and so $|\mathbf{f}_t \times \mathbf{f}_s| = (\cos s + 2)$ so the area element is $(\cos s + 2) ds dt$ and the area is

$$\int_0^{2\pi} \int_0^{2\pi} (\cos s + 2) ds dt = 8\pi^2$$

Example 26.1.8 Let $U = [0, 2\pi] \times [0, 2\pi]$ and for $(t, s) \in U$, let

$$\mathbf{f}(t, s) = (2 \cos t + \cos t \cos s, -2 \sin t - \sin t \cos s, \sin s)^T.$$

Find

$$\int_{\mathbf{f}(U)} h dV$$

where $h(x, y, z) = x^2$.

Everything is the same as the preceding example except this time it is an integral of a function. The area element is $(\cos s + 2) ds dt$ and so the integral called for is

$$\int_{\mathbf{f}(U)} h dA = \int_0^{2\pi} \int_0^{2\pi} \left(\overbrace{2 \cos t + \cos t \cos s}^{x \text{ on the surface}} \right)^2 (\cos s + 2) ds dt = 22\pi^2$$

26.1.1 Surfaces Of The Form $z = f(x, y)$

The special case where a surface is in the form $z = f(x, y)$, $(x, y) \in U$, yields a simple formula which is used most often in this situation. You write the surface parametrically in the form $\mathbf{f}(x, y) = (x, y, f(x, y))^T$ such that $(x, y) \in U$. Then

$$\mathbf{f}_x = \begin{pmatrix} 1 \\ 0 \\ f_x \end{pmatrix}, \mathbf{f}_y = \begin{pmatrix} 0 \\ 1 \\ f_y \end{pmatrix}$$

and

$$|\mathbf{f}_x \times \mathbf{f}_y| = \sqrt{1 + f_y^2 + f_x^2}$$

so the area element is

$$\sqrt{1 + f_y^2 + f_x^2} dx dy.$$

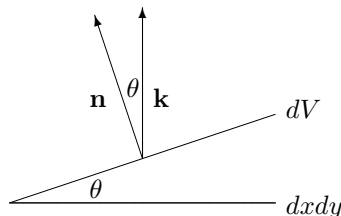
When the surface of interest comes in this simple form, people generally use this area element directly rather than worrying about a parametrization and taking cross products.

In the case where the surface is of the form $x = f(y, z)$ for $(y, z) \in U$, the area element is obtained similarly and is

$$\sqrt{1 + f_y^2 + f_z^2} dy dz.$$

I think you can guess what the area element is if $y = f(x, z)$.

There is also a simple geometric description of these area elements. Consider the surface $z = f(x, y)$. This is a level surface of the function of three variables $z - f(x, y)$. In fact the surface is simply $z - f(x, y) = 0$. Now consider the gradient of this function of three variables. The gradient is perpendicular to the surface and the third component is positive in this case. This gradient is $(-f_x, -f_y, 1)$ and so the unit upward normal is just $\frac{1}{\sqrt{1+f_x^2+f_y^2}}(-f_x, -f_y, 1)$. Now consider the following picture.



In this picture, you are looking at a chunk of area on the surface seen on edge and so it seems reasonable to expect to have $dx dy = dV \cos \theta$. But it is easy to find $\cos \theta$ from the picture and the properties of the dot product.

$$\cos \theta = \frac{\mathbf{n} \cdot \mathbf{k}}{|\mathbf{n}| |\mathbf{k}|} = \frac{1}{\sqrt{1 + f_x^2 + f_y^2}}.$$

Therefore, $dA = \sqrt{1 + f_x^2 + f_y^2} dx dy$ as claimed.

Example 26.1.9 Let $z = \sqrt{x^2 + y^2}$ where $(x, y) \in U$ for $U = \{(x, y) : x^2 + y^2 \leq 4\}$. Find

$$\int_S h dS$$

where $h(x, y, z) = x + z$ and S is the surface described as $(x, y, \sqrt{x^2 + y^2})$ for $(x, y) \in U$.

Here you can see directly the angle in the above picture is $\frac{\pi}{4}$ and so $dV = \sqrt{2} dx dy$. If you don't see this or if it is unclear, simply compute $\sqrt{1 + f_x^2 + f_y^2}$ and you will find it is $\sqrt{2}$. Therefore, using polar coordinates,

$$\begin{aligned} \int_S h dS &= \int_U (x + \sqrt{x^2 + y^2}) \sqrt{2} dA \\ &= \sqrt{2} \int_0^{2\pi} \int_0^2 (r \cos \theta + r) r dr d\theta \\ &= \frac{16}{3} \sqrt{2} \pi. \end{aligned}$$

One other issue is worth mentioning. Suppose $\mathbf{f}_i : U_i \rightarrow \mathbb{R}^3$ where U_i are sets in $\mathbb{R}^2$ and suppose $\mathbf{f}_1(U_1)$ intersects $\mathbf{f}_2(U_2)$ along C where $C = \mathbf{h}(V)$ for $V \subseteq \mathbb{R}^1$. Then

define integrals and areas over $\mathbf{f}_1(U_1) \cup \mathbf{f}_2(U_2)$ as follows.

$$\int_{\mathbf{f}_1(U_1) \cup \mathbf{f}_2(U_2)} g \, dA \equiv \int_{\mathbf{f}_1(U_1)} g \, dA + \int_{\mathbf{f}_2(U_2)} g \, dA.$$

Admittedly, the set C gets added in twice but this doesn't matter because its 2 dimensional volume equals zero and therefore, the integrals over this set will also be zero.

I have been purposely vague about precise mathematical conditions necessary for the above procedures. This is because the precise mathematical conditions which are usually cited are very technical and at the same time far too restrictive. The most general conditions under which these sorts of procedures are valid include things like Lipschitz functions defined on very general sets. These are functions satisfying a Lipschitz condition of the form $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq K |\mathbf{x} - \mathbf{y}|$. For example, $y = |x|$ is Lipschitz continuous. However, this function does not have a derivative at every point. So it is with Lipschitz functions. However, it turns out these functions have derivatives at enough points to push everything through but this requires considerations involving the Lebesgue integral. Lipschitz functions are also not the most general kind of function for which the above is valid.

26.2 Exercises

- Find a parametrization for the intersection of the planes $4x + 2y + 4z = 3$ and $6x - 2y = -1$.
- Find a parametrization for the intersection of the plane $3x + y + z = 1$ and the circular cylinder $x^2 + y^2 = 1$.
- Find a parametrization for the intersection of the plane $3x + 2y + 4z = 4$ and the elliptic cylinder $x^2 + 4z^2 = 16$.
- Find a parametrization for the straight line joining $(1, 3, 1)$ and $(-2, 5, 3)$.
- Find a parametrization for the intersection of the surfaces $4y + 3z = 3x^2 + 2$ and $3y + 2z = -x + 3$.
- Find the area of S if S is the part of the circular cylinder $x^2 + y^2 = 4$ which lies between $z = 0$ and $z = 2 + y$.
- Find the area of S if S is the part of the cone $x^2 + y^2 = 16z^2$ between $z = 0$ and $z = h$.
- Parametrizing the cylinder $x^2 + y^2 = a^2$ by $x = a \cos v, y = a \sin v, z = u$, show that the area element is $dA = a \, du \, dv$.
- Find the area enclosed by the limaçon $r = 2 + \cos \theta$.
- Find the surface area of the paraboloid $z = h(1 - x^2 - y^2)$ between $z = 0$ and $z = h$. Take a limit of this area as h decreases to 0.
- Evaluate $\int_S (1 + x) \, dA$ where S is the part of the plane $4x + y + 3z = 12$ which is in the first octant.
- Evaluate $\int_S (1 + x) \, dA$ where S is the part of the cylinder $x^2 + y^2 = 9$ between $z = 0$ and $z = h$.
- Evaluate $\int_S (1 + x) \, dA$ where S is the hemisphere $x^2 + y^2 + z^2 = 4$ between $x = 0$ and $x = 2$.

14. For $(\theta, \alpha) \in [0, 2\pi] \times [0, 2\pi]$, let

$$\mathbf{f}(\theta, \alpha) \equiv (\cos \theta (4 + \cos \alpha), -\sin \theta (4 + \cos \alpha), \sin \alpha)^T.$$

Find the area of $\mathbf{f}([0, 2\pi] \times [0, 2\pi])$. **Hint:** Check whether $\mathbf{f}_\theta \cdot \mathbf{f}_\alpha = 0$. This might make the computations reasonable.

15. For $(\theta, \alpha) \in [0, 2\pi] \times [0, 2\pi]$, let $\mathbf{f}(\theta, \alpha) \equiv$

$$(\cos \theta (3 + 2 \cos \alpha), -\sin \theta (3 + 2 \cos \alpha), 2 \sin \alpha)^T.$$

Also let $h(\mathbf{x}) = \cos \alpha$ where α is such that

$$\mathbf{x} = (\cos \theta (3 + 2 \cos \alpha), -\sin \theta (3 + 2 \cos \alpha), 2 \sin \alpha)^T.$$

Find $\int_{\mathbf{f}([0, 2\pi] \times [0, 2\pi])} h \, dA$. **Hint:** Check whether $\mathbf{f}_\theta \cdot \mathbf{f}_\alpha = 0$. This might make the computations reasonable.

16. For $(\theta, \alpha) \in [0, 2\pi] \times [0, 2\pi]$, let $\mathbf{f}(\theta, \alpha) \equiv$

$$(\cos \theta (4 + 3 \cos \alpha), -\sin \theta (4 + 3 \cos \alpha), 3 \sin \alpha)^T.$$

Also let $h(\mathbf{x}) = \cos^2 \theta$ where θ is such that

$$\mathbf{x} = (\cos \theta (4 + 3 \cos \alpha), -\sin \theta (4 + 3 \cos \alpha), 3 \sin \alpha)^T.$$

Find $\int_{\mathbf{f}([0, 2\pi] \times [0, 2\pi])} h \, dA$. **Hint:** Check whether $\mathbf{f}_\theta \cdot \mathbf{f}_\alpha = 0$. This might make the computations reasonable.

17. In spherical coordinates, $\phi = c, \rho \in [0, R]$ determines a cone. Find the area of this cone.

Calculus Of Vector Fields

27.0.1 Outcomes

1. Define and evaluate the divergence of a vector field in terms of Cartesian coordinates.
2. Define and evaluate the Curl of a vector field in Cartesian coordinates.
3. Discover vector identities involving the gradient, divergence, and curl.
4. Recall and verify the divergence theorem.
5. Apply the divergence theorem.

27.1 Divergence And Curl Of A Vector Field

Here the important concepts of divergence and curl are defined.

Definition 27.1.1 Let $\mathbf{f} : U \rightarrow \mathbb{R}^p$ for $U \subseteq \mathbb{R}^p$ denote a vector field. A scalar valued function is called a **scalar field**. The function, $\mathbf{f}$ is called a C^k **vector field** if the function, $\mathbf{f}$ is a C^k function. For a C^1 vector field, as just described $\nabla \cdot \mathbf{f}(\mathbf{x}) \equiv \text{div } \mathbf{f}(\mathbf{x})$ known as the **divergence**, is defined as

$$\nabla \cdot \mathbf{f}(\mathbf{x}) \equiv \text{div } \mathbf{f}(\mathbf{x}) \equiv \sum_{i=1}^p \frac{\partial f_i}{\partial x_i}(\mathbf{x}).$$

Using the repeated summation convention, this is often written as

$$f_{i,i}(\mathbf{x}) \equiv \partial_i f_i(\mathbf{x})$$

where the comma indicates a partial derivative is being taken with respect to the i^{th} variable and ∂_i denotes differentiation with respect to the i^{th} variable. In words, the divergence is the sum of the i^{th} derivative of the i^{th} component function of $\mathbf{f}$ for all values of i . If $p = 3$, the **curl** of the vector field yields another vector field and it is defined as follows.

$$(\text{curl } (\mathbf{f})(\mathbf{x}))_i \equiv (\nabla \times \mathbf{f}(\mathbf{x}))_i \equiv \varepsilon_{ijk} \partial_j f_k(\mathbf{x})$$

where here ∂_j means the partial derivative with respect to x_j and the subscript of i in $(\text{curl } (\mathbf{f})(\mathbf{x}))_i$ means the i^{th} Cartesian component of the vector, $\text{curl } (\mathbf{f})(\mathbf{x})$. Thus the curl is evaluated by expanding the following determinant along the top row.

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ f_1(x, y, z) & f_2(x, y, z) & f_3(x, y, z) \end{vmatrix}.$$

Note the similarity with the cross product. Sometimes the curl is called *rot*. (Short for rotation not decay.) Also

$$\nabla^2 f \equiv \nabla \cdot (\nabla f).$$

This last symbol is important enough that it is given a name, the **Laplacian**. It is also denoted by Δ . Thus $\nabla^2 f = \Delta f$. In addition for $\mathbf{f}$ a vector field, the symbol $\mathbf{f} \cdot \nabla$ is defined as a “differential operator” in the following way.

$$\mathbf{f} \cdot \nabla (\mathbf{g}) \equiv f_1(\mathbf{x}) \frac{\partial \mathbf{g}(\mathbf{x})}{\partial x_1} + f_2(\mathbf{x}) \frac{\partial \mathbf{g}(\mathbf{x})}{\partial x_2} + \cdots + f_p(\mathbf{x}) \frac{\partial \mathbf{g}(\mathbf{x})}{\partial x_p}.$$

Thus $\mathbf{f} \cdot \nabla$ takes vector fields and makes them into new vector fields.

This definition is in terms of a given coordinate system but later coordinate free definitions of the curl and div are presented. For now, everything is defined in terms of a given Cartesian coordinate system. The divergence and curl have profound physical significance and this will be discussed later. For now it is important to understand their definition in terms of coordinates. Be sure you understand that for $\mathbf{f}$ a vector field, $\text{div } \mathbf{f}$ is a scalar field meaning it is a scalar valued function of three variables. For a scalar field, f , ∇f is a vector field described earlier on Page 517. For $\mathbf{f}$ a vector field having values in $\mathbb{R}^3$, $\text{curl } \mathbf{f}$ is another vector field.

Example 27.1.2 Let $\mathbf{f}(\mathbf{x}) = xy\mathbf{i} + (z - y)\mathbf{j} + (\sin(x) + z)\mathbf{k}$. Find $\text{div } \mathbf{f}$ and $\text{curl } \mathbf{f}$.

First the divergence of $\mathbf{f}$ is

$$\frac{\partial(xy)}{\partial x} + \frac{\partial(z - y)}{\partial y} + \frac{\partial(\sin(x) + z)}{\partial z} = y + (-1) + 1 = y.$$

Now $\text{curl } \mathbf{f}$ is obtained by evaluating

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ xy & z - y & \sin(x) + z \end{vmatrix} =$$

$$\mathbf{i} \left(\frac{\partial}{\partial y} (\sin(x) + z) - \frac{\partial}{\partial z} (z - y) \right) - \mathbf{j} \left(\frac{\partial}{\partial x} (\sin(x) + z) - \frac{\partial}{\partial z} (xy) \right) +$$

$$\mathbf{k} \left(\frac{\partial}{\partial x} (z - y) - \frac{\partial}{\partial y} (xy) \right) = -\mathbf{i} - \cos(x)\mathbf{j} - x\mathbf{k}.$$

27.1.1 Vector Identities

There are many interesting identities which relate the gradient, divergence and curl.

Theorem 27.1.3 Assuming $\mathbf{f}, \mathbf{g}$ are a C^2 vector fields whenever necessary, the following identities are valid.

1. $\nabla \cdot (\nabla \times \mathbf{f}) = 0$
2. $\nabla \times \nabla \phi = \mathbf{0}$
3. $\nabla \times (\nabla \times \mathbf{f}) = \nabla (\nabla \cdot \mathbf{f}) - \nabla^2 \mathbf{f}$ where $\nabla^2 \mathbf{f}$ is a vector field whose i^{th} component is $\nabla^2 f_i$.
4. $\nabla \cdot (\mathbf{f} \times \mathbf{g}) = \mathbf{g} \cdot (\nabla \times \mathbf{f}) - \mathbf{f} \cdot (\nabla \times \mathbf{g})$
5. $\nabla \times (\mathbf{f} \times \mathbf{g}) = (\nabla \cdot \mathbf{g}) \mathbf{f} - (\nabla \cdot \mathbf{f}) \mathbf{g} + (\mathbf{g} \cdot \nabla) \mathbf{f} - (\mathbf{f} \cdot \nabla) \mathbf{g}$

Proof: These are all easy to establish if you use the repeated index summation convention and the reduction identities discussed on Page 391.

$$\begin{aligned}
 \nabla \cdot (\nabla \times \mathbf{f}) &= \partial_i (\nabla \times \mathbf{f})_i \\
 &= \partial_i (\varepsilon_{ijk} \partial_j f_k) \\
 &= \varepsilon_{ijk} \partial_i (\partial_j f_k) \\
 &= \varepsilon_{jik} \partial_j (\partial_i f_k) \\
 &= -\varepsilon_{ijk} \partial_j (\partial_i f_k) \\
 &= -\varepsilon_{ijk} \partial_i (\partial_j f_k) \\
 &= -\nabla \cdot (\nabla \times \mathbf{f}).
 \end{aligned}$$

This establishes the first formula. The second formula is done similarly. Now consider the third.

$$\begin{aligned}
 (\nabla \times (\nabla \times \mathbf{f}))_i &= \varepsilon_{ijk} \partial_j (\nabla \times \mathbf{f})_k \\
 &= \varepsilon_{ijk} \partial_j (\varepsilon_{krs} \partial_r f_s) \\
 &= \varepsilon_{ijk} \varepsilon_{krs} \partial_j (\partial_r f_s) \\
 &= (\delta_{ir} \delta_{js} - \delta_{is} \delta_{jr}) \partial_j (\partial_r f_s) \\
 &= \partial_j (\partial_i f_j) - \partial_j (\partial_j f_i) \\
 &= \partial_i (\partial_j f_j) - \partial_j (\partial_j f_i) \\
 &= (\nabla (\nabla \cdot \mathbf{f}) - \nabla^2 \mathbf{f})_i
 \end{aligned}$$

This establishes the third identity.

Consider the fourth identity.

$$\begin{aligned}
 \nabla \cdot (\mathbf{f} \times \mathbf{g}) &= \partial_i (\mathbf{f} \times \mathbf{g})_i \\
 &= \partial_i \varepsilon_{ijk} f_j g_k \\
 &= \varepsilon_{ijk} (\partial_i f_j) g_k + \varepsilon_{ijk} f_j (\partial_i g_k) \\
 &= (\varepsilon_{kij} \partial_i f_j) g_k - (\varepsilon_{jik} \partial_i g_k) f_k \\
 &= \nabla \times \mathbf{f} \cdot \mathbf{g} - \nabla \times \mathbf{g} \cdot \mathbf{f}.
 \end{aligned}$$

This proves the fourth identity.

Consider the fifth.

$$\begin{aligned}
 (\nabla \times (\mathbf{f} \times \mathbf{g}))_i &= \varepsilon_{ijk} \partial_j (\mathbf{f} \times \mathbf{g})_k \\
 &= \varepsilon_{ijk} \partial_j \varepsilon_{krs} f_r g_s \\
 &= \varepsilon_{kij} \varepsilon_{krs} \partial_j (f_r g_s) \\
 &= (\delta_{ir} \delta_{js} - \delta_{is} \delta_{jr}) \partial_j (f_r g_s) \\
 &= \partial_j (f_i g_j) - \partial_j (f_j g_i) \\
 &= (\partial_j g_j) f_i + g_j \partial_j f_i - (\partial_j f_j) g_i - f_j (\partial_j g_i) \\
 &= ((\nabla \cdot \mathbf{g}) \mathbf{f} + (\mathbf{g} \cdot \nabla) (\mathbf{f}) - (\nabla \cdot \mathbf{f}) \mathbf{g} - (\mathbf{f} \cdot \nabla) (\mathbf{g}))_i
 \end{aligned}$$

and this establishes the fifth identity.

I think the important thing about the above is not that these identities can be proved and are valid as much as the method by which they were proved. The reduction identities on Page 391 were used to discover the identities. There is a difference between proving something someone tells you about and both discovering what should be proved and proving it. This notation and the reduction identity make the discovery of vector identities fairly routine and this is why these things are of great significance.

27.1.2 Vector Potentials

One of the above identities says $\nabla \cdot (\nabla \times \mathbf{f}) = 0$. Suppose now $\nabla \cdot \mathbf{g} = 0$. Does it follow that there exists $\mathbf{f}$ such that $\mathbf{g} = \nabla \times \mathbf{f}$? It turns out that this is usually the case and when such an $\mathbf{f}$ exists, it is called a **vector potential**. Here is one way to do it, assuming everything is defined so the following formulas make sense.

$$\mathbf{f}(x, y, z) = \left(\int_0^z g_2(x, y, t) dt, - \int_0^z g_1(x, y, t) dt + \int_0^x g_3(t, y, 0) dt, 0 \right)^T. \quad (27.1)$$

In verifying this you need to use the following manipulation which will generally hold under reasonable conditions but which has not been carefully shown yet.

$$\frac{\partial}{\partial x} \int_a^b h(x, t) dt = \int_a^b \frac{\partial h}{\partial x}(x, t) dt. \quad (27.2)$$

The above formula seems plausible because the integral is a sort of a sum and the derivative of a sum is the sum of the derivatives. However, this sort of sloppy reasoning will get you into all sorts of trouble. The formula involves the interchange of two limit operations, the integral and the limit of a difference quotient. Such an interchange can only be accomplished through a theorem. The following gives the necessary result. This lemma is stated without proof.

Lemma 27.1.4 Suppose h and $\frac{\partial h}{\partial x}$ are continuous on the rectangle $R = [c, d] \times [a, b]$. Then 27.2 holds.

Proof: Let Δx be such that $x, x + \Delta x$ are both in $[c, d]$. By Theorem 17.8.23 on Page 429 there exists $\delta > 0$ such that if $|(x, t) - (x_1, t_1)| < \delta$, then

$$\left| \frac{\partial h}{\partial x}(x, t) - \frac{\partial h}{\partial x}(x_1, t_1) \right| < \frac{\varepsilon}{b-a}.$$

Let $|\Delta x| < \delta$. Then

$$\begin{aligned} & \left| \int_a^b \frac{h(x + \Delta x, t) - h(x, t)}{\Delta x} dt - \int_a^b \frac{\partial h}{\partial x}(x, t) dt \right| \\ & \leq \int_a^b \left| \frac{h(x + \Delta x, t) - h(x, t)}{\Delta x} - \frac{\partial h}{\partial x}(x, t) \right| dt \\ & = \int_a^b \left| \frac{\partial h(x + \theta_t \Delta x)}{\partial x} - \frac{\partial h}{\partial x}(x, t) \right| dt < \int_a^b \frac{\varepsilon}{b-a} dt = \varepsilon. \end{aligned}$$

Here θ_t is a number between 0 and 1 and going from the second to the third line is an application of the mean value theorem. This proves the lemma.

The second formula of Theorem 27.1.3 states $\nabla \times \nabla \phi = \mathbf{0}$. This suggests the following question: Suppose $\nabla \times \mathbf{f} = \mathbf{0}$, does it follow there exists ϕ , a scalar field such that $\nabla \phi = \mathbf{f}$? The answer to this is often yes and a theorem will be given and proved after the presentation of Stoke's theorem. This scalar field, ϕ , is called a **scalar potential** for $\mathbf{f}$.

27.1.3 The Weak Maximum Principle

There is also a fundamental result having great significance which involves ∇^2 called the maximum principle. This principle says that if $\nabla^2 u \geq 0$ on a bounded open set, U , then u achieves its maximum value on the boundary of U .

Theorem 27.1.5 Let U be a bounded open set in $\mathbb{R}^n$ and suppose

$$u \in C^2(U) \cap C(\overline{U})$$

such that $\nabla^2 u \geq 0$ in U . Then letting $\partial U = \overline{U} \setminus U$, it follows that

$$\max \{u(\mathbf{x}) : \mathbf{x} \in \overline{U}\} = \max \{u(\mathbf{x}) : \mathbf{x} \in \partial U\}.$$

Proof: If this is not so, there exists $\mathbf{x}_0 \in U$ such that $u(\mathbf{x}_0) > \max \{u(\mathbf{x}) : \mathbf{x} \in \partial U\} \equiv M$. Since U is bounded, there exists $\varepsilon > 0$ such that

$$u(\mathbf{x}_0) > \max \{u(\mathbf{x}) + \varepsilon |\mathbf{x}|^2 : \mathbf{x} \in \partial U\}.$$

Therefore, $u(\mathbf{x}) + \varepsilon |\mathbf{x}|^2$ also has its maximum in U because for ε small enough,

$$u(\mathbf{x}_0) + \varepsilon |\mathbf{x}_0|^2 > u(\mathbf{x}_0) > \max \{u(\mathbf{x}) + \varepsilon |\mathbf{x}|^2 : \mathbf{x} \in \partial U\}$$

for all $\mathbf{x} \in \partial U$.

Now let $\mathbf{x}_1$ be the point in U at which $u(\mathbf{x}) + \varepsilon |\mathbf{x}|^2$ achieves its maximum. As an exercise you should show that $\nabla^2(f + g) = \nabla^2 f + \nabla^2 g$ and therefore, $\nabla^2(u(\mathbf{x}) + \varepsilon |\mathbf{x}|^2) = \nabla^2 u(\mathbf{x}) + 2n\varepsilon$. (Why?) Therefore,

$$0 \geq \nabla^2 u(\mathbf{x}_1) + 2n\varepsilon \geq 2n\varepsilon,$$

a contradiction. This proves the theorem.

27.2 Exercises

1. Find $\operatorname{div} \mathbf{f}$ and $\operatorname{curl} \mathbf{f}$ where $\mathbf{f}$ is

(a) $(xyz, x^2 + \ln(xy), \sin x^2 + z)^T$

(b) $(\sin x, \sin y, \sin z)^T$

(c) $(f(x), g(y), h(z))^T$

(d) $(x - 2, y - 3, z - 6)^T$

(e) $(y^2, 2xy, \cos z)^T$

(f) $(f(y, z), g(x, z), h(y, z))^T$

2. Prove formula 2 of Theorem 27.1.3.

3. Show that if u and v are C^2 functions, then $\operatorname{curl}(u\nabla v) = \nabla u \times \nabla v$.

4. Simplify the expression $\mathbf{f} \times (\nabla \times \mathbf{g}) + \mathbf{g} \times (\nabla \times \mathbf{f}) + (\mathbf{f} \cdot \nabla) \mathbf{g} + (\mathbf{g} \cdot \nabla) \mathbf{f}$.

5. Simplify $\nabla \times (\mathbf{v} \times \mathbf{r})$ where $\mathbf{r} = (x, y, z)^T = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ and $\mathbf{v}$ is a constant vector.

6. Discover a formula which simplifies $\nabla \cdot (v\nabla u)$.

7. Verify that $\nabla \cdot (u\nabla v) - \nabla \cdot (v\nabla u) = u\nabla^2 v - v\nabla^2 u$.

8. Verify that $\nabla^2(uv) = v\nabla^2 u + 2(\nabla u \cdot \nabla v) + u\nabla^2 v$.

9. Functions, u , which satisfy $\nabla^2 u = 0$ are called harmonic functions. Show that the following functions are harmonic where ever they are defined.

- (a) $2xy$
- (b) $x^2 - y^2$
- (c) $\sin x \cosh y$
- (d) $\ln(x^2 + y^2)$
- (e) $1/\sqrt{x^2 + y^2 + z^2}$

10. Verify the formula given in 27.1 is a vector potential for $\mathbf{g}$ assuming that $\operatorname{div} \mathbf{g} = 0$.
11. Show that if $\nabla^2 u_k = 0$ for each $k = 1, 2, \dots, m$, and c_k is a constant, then

$$\nabla^2 \left(\sum_{k=1}^m c_k u_k \right) = 0$$

also.

12. In Theorem 27.1.5, why is $\nabla^2 (\varepsilon |\mathbf{x}|^2) = 2n\varepsilon$?
13. Using Theorem 27.1.5 prove the following: Let $f \in C(\partial U)$ (f is continuous on ∂U .) where U is a bounded open set. Then there exists at most one solution, $u \in C^2(U) \cap C(\overline{U})$ and $\nabla^2 u = 0$ in U with $u = f$ on ∂U . **Hint:** Suppose there are two solutions, u_i , $i = 1, 2$ and let $w = u_1 - u_2$. Then use the maximum principle.
14. Suppose $\mathbf{B}$ is a vector field and $\nabla \times \mathbf{A} = \mathbf{B}$. Thus $\mathbf{A}$ is a vector potential for $\mathbf{B}$. Show that $\mathbf{A} + \nabla \phi$ is also a vector potential for $\mathbf{B}$. Here ϕ is just a C^2 scalar field. Thus the vector potential is not unique.

27.3 The Divergence Theorem

The divergence theorem relates an integral over a set to one on the boundary of the set. It is also called Gauss's theorem.

Definition 27.3.1 A subset, V of $\mathbb{R}^3$ is called *cylindrical in the x direction* if it is of the form

$$V = \{(x, y, z) : \phi(y, z) \leq x \leq \psi(y, z) \text{ for } (y, z) \in D\}$$

where D is a subset of the yz plane. V is *cylindrical in the z direction* if

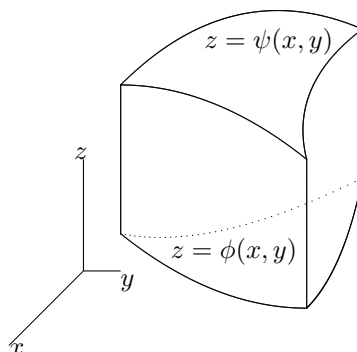
$$V = \{(x, y, z) : \phi(x, y) \leq z \leq \psi(x, y) \text{ for } (x, y) \in D\}$$

where D is a subset of the xy plane, and V is *cylindrical in the y direction* if

$$V = \{(x, y, z) : \phi(x, z) \leq y \leq \psi(x, z) \text{ for } (x, z) \in D\}$$

where D is a subset of the xz plane. If V is cylindrical in the z direction, denote by ∂V the boundary of V defined to be the points of the form $(x, y, \phi(x, y))$, $(x, y, \psi(x, y))$ for $(x, y) \in D$, along with points of the form (x, y, z) where $(x, y) \in \partial D$ and $\phi(x, y) \leq z \leq \psi(x, y)$. Points on ∂D are defined to be those for which every open ball contains points which are in D as well as points which are not in D . A similar definition holds for ∂V in the case that V is cylindrical in one of the other directions.

The following picture illustrates the above definition in the case of V cylindrical in the z direction.



Of course, many three dimensional sets are cylindrical in each of the coordinate directions. For example, a ball or a rectangle or a tetrahedron are all cylindrical in each direction. The following lemma allows the exchange of the volume integral of a partial derivative for an area integral in which the derivative is replaced with multiplication by an appropriate component of the unit exterior normal.

Lemma 27.3.2 *Suppose V is cylindrical in the z direction and that ϕ and ψ are the functions in the above definition. Assume ϕ and ψ are C^1 functions and suppose F is a C^1 function defined on V . Also, let $\mathbf{n} = (n_x, n_y, n_z)$ be the unit exterior normal to ∂V . Then*

$$\int_V \frac{\partial F}{\partial z}(x, y, z) dV = \int_{\partial V} F n_z dA.$$

Proof: From the fundamental theorem of calculus,

$$\begin{aligned} \int_V \frac{\partial F}{\partial z}(x, y, z) dV &= \int_D \int_{\phi(x, y)}^{\psi(x, y)} \frac{\partial F}{\partial z}(x, y, z) dz dx dy \\ &= \int_D [F(x, y, \psi(x, y)) - F(x, y, \phi(x, y))] dx dy \end{aligned} \quad (27.3)$$

Now the unit exterior normal on the top of V , the surface $(x, y, \psi(x, y))$ is

$$\frac{1}{\sqrt{\psi_x^2 + \psi_y^2 + 1}} (-\psi_x, -\psi_y, 1).$$

This follows from the observation that the top surface is the level surface, $z - \psi(x, y) = 0$ and so the gradient of this function of three variables is perpendicular to the level surface. It points in the correct direction because the z component is positive. Therefore, on the top surface,

$$n_z = \frac{1}{\sqrt{\psi_x^2 + \psi_y^2 + 1}}$$

Similarly, the unit normal to the surface on the bottom is

$$\frac{1}{\sqrt{\phi_x^2 + \phi_y^2 + 1}} (\phi_x, \phi_y, -1)$$

and so on the bottom surface,

$$n_z = \frac{-1}{\sqrt{\phi_x^2 + \phi_y^2 + 1}}$$

Note that here the z component is negative because since it is the outer normal it must point down. On the lateral surface, the one where $(x, y) \in \partial D$ and $z \in [\phi(x, y), \psi(x, y)]$, $n_z = 0$.

The area element on the top surface is $dA = \sqrt{\psi_x^2 + \psi_y^2 + 1} dx dy$ while the area element on the bottom surface is $\sqrt{\phi_x^2 + \phi_y^2 + 1} dx dy$. Therefore, the last expression in 27.3 is of the form,

$$\begin{aligned} & \int_D F(x, y, \psi(x, y)) \overbrace{\frac{1}{\sqrt{\psi_x^2 + \psi_y^2 + 1}}}^{n_z} \overbrace{\sqrt{\psi_x^2 + \psi_y^2 + 1} dx dy}^{dA} + \\ & \int_D F(x, y, \phi(x, y)) \left(\overbrace{\frac{-1}{\sqrt{\phi_x^2 + \phi_y^2 + 1}}}^{n_z} \right) \overbrace{\sqrt{\phi_x^2 + \phi_y^2 + 1} dx dy}^{dA} \\ & + \int_{\text{Lateral surface}} F n_z dA, \end{aligned}$$

the last term equaling zero because on the lateral surface, $n_z = 0$. Therefore, this reduces to $\int_{\partial V} F n_z dA$ as claimed.

The following corollary is entirely similar to the above.

Corollary 27.3.3 *If V is cylindrical in the y direction, then*

$$\int_V \frac{\partial F}{\partial y} dV = \int_{\partial V} F n_y dA$$

and if V is cylindrical in the x direction, then

$$\int_V \frac{\partial F}{\partial x} dV = \int_{\partial V} F n_x dA$$

With this corollary, here is a proof of the divergence theorem.

Theorem 27.3.4 *Let V be cylindrical in each of the coordinate directions and let $\mathbf{F}$ be a C^1 vector field defined on V . Then*

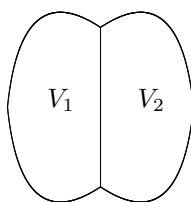
$$\int_V \nabla \cdot \mathbf{F} dV = \int_{\partial V} \mathbf{F} \cdot \mathbf{n} dA.$$

Proof: From the above lemma and corollary,

$$\begin{aligned} \int_V \nabla \cdot \mathbf{F} dV &= \int_V \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z} dV \\ &= \int_{\partial V} (F_1 n_x + F_2 n_y + F_3 n_z) dA \\ &= \int_{\partial V} \mathbf{F} \cdot \mathbf{n} dA. \end{aligned}$$

This proves the theorem.

The divergence theorem holds for much more general regions than this. Suppose for example you have a complicated region which is the union of finitely many disjoint regions of the sort just described which are cylindrical in each of the coordinate directions. Then the volume integral over the union of these would equal the sum of the integrals over the disjoint regions. If the boundaries of two of these regions intersect, then the area integrals will cancel out on the intersection because the unit exterior normals will point in opposite directions. Therefore, the sum of the integrals over the boundaries of these disjoint regions will reduce to an integral over the boundary of the union of these. Hence the divergence theorem will continue to hold. For example, consider the following picture. If the divergence theorem holds for each V_i in the following picture, then it holds for the union of these two.



General formulations of the divergence theorem involve Hausdorff measures and the Lebesgue integral, a better integral than the old fashioned Riemann integral which has been obsolete now for almost 100 years. When all is said and done, one finds that the conclusion of the divergence theorem is usually true and the theorem can be used with confidence.

Example 27.3.5 Let $V = [0, 1] \times [0, 1] \times [0, 1]$. That is, V is the cube in the first octant having the lower left corner at $(0, 0, 0)$ and the sides of length 1. Let $\mathbf{F}(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. Find the flux integral in which $\mathbf{n}$ is the unit exterior normal.

$$\int_{\partial V} \mathbf{F} \cdot \mathbf{n} dS$$

You can certainly inflict much suffering on yourself by breaking the surface up into 6 pieces corresponding to the 6 sides of the cube, finding a parametrization for each face and adding up the appropriate flux integrals. For example, $\mathbf{n} = \mathbf{k}$ on the top face and $\mathbf{n} = -\mathbf{k}$ on the bottom face. On the top face, a parametrization is $(x, y, 1) : (x, y) \in [0, 1] \times [0, 1]$. The area element is just $dx dy$. It isn't really all that hard to do it this way but it is much easier to use the divergence theorem. The above integral equals

$$\int_V \operatorname{div}(\mathbf{F}) dV = \int_V 3 dV = 3.$$

Example 27.3.6 This time, let V be the unit ball, $\{(x, y, z) : x^2 + y^2 + z^2 \leq 1\}$ and let $\mathbf{F}(x, y, z) = x^2\mathbf{i} + y\mathbf{j} + (z - 1)\mathbf{k}$. Find

$$\int_{\partial V} \mathbf{F} \cdot \mathbf{n} dS.$$

As in the above you could do this by brute force. A parametrization of the ∂V is obtained as

$$x = \sin \phi \cos \theta, \quad y = \sin \phi \sin \theta, \quad z = \cos \phi$$

where $(\phi, \theta) \in (0, \pi) \times (0, 2\pi]$. Now this does not include all the ball but it includes all but the point at the top and at the bottom. As far as the flux integral is concerned

these points contribute nothing to the integral so you can neglect them. Then you can grind away and get the flux integral which is desired. However, it is so much easier to use the divergence theorem! Using spherical coordinates,

$$\begin{aligned}\int_{\partial V} \mathbf{F} \cdot \mathbf{n} dS &= \int_V \operatorname{div}(\mathbf{F}) dV = \int_V (2x + 1 + 1) dV \\ &= \int_0^\pi \int_0^{2\pi} \int_0^1 (2 + 2\rho \sin(\phi) \cos \theta) \rho^2 \sin(\phi) d\rho d\theta d\phi = \frac{8}{3}\pi\end{aligned}$$

Example 27.3.7 Suppose V is an open set in $\mathbb{R}^3$ for which the divergence theorem holds. Let $\mathbf{F}(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. Then show that

$$\int_{\partial V} \mathbf{F} \cdot \mathbf{n} dS = 3 \times \text{volume}(V).$$

This follows from the divergence theorem.

$$\int_{\partial V} \mathbf{F} \cdot \mathbf{n} dS = \int_V \operatorname{div}(\mathbf{F}) dV = 3 \int_V dV = 3 \times \text{volume}(V).$$

The message of the divergence theorem is the relation between the volume integral and an area integral. This is the exciting thing about this marvelous theorem. It is not its utility as a method for evaluations of boring problems. This will be shown in the examples of its use which follow.

27.3.1 Coordinate Free Concept Of Divergence

The divergence theorem also makes possible a coordinate free definition of the divergence.

Theorem 27.3.8 Let $B(\mathbf{x}, \delta)$ be the ball centered at $\mathbf{x}$ having radius δ and let $\mathbf{F}$ be a C^1 vector field. Then letting $v(B(\mathbf{x}, \delta))$ denote the volume of $B(\mathbf{x}, \delta)$ given by

$$\int_{B(\mathbf{x}, \delta)} dV,$$

it follows

$$\operatorname{div} \mathbf{F}(\mathbf{x}) = \lim_{\delta \rightarrow 0^+} \frac{1}{v(B(\mathbf{x}, \delta))} \int_{\partial B(\mathbf{x}, \delta)} \mathbf{F} \cdot \mathbf{n} dA. \quad (27.4)$$

Proof: The divergence theorem holds for balls because they are cylindrical in every direction. Therefore,

$$\frac{1}{v(B(\mathbf{x}, \delta))} \int_{\partial B(\mathbf{x}, \delta)} \mathbf{F} \cdot \mathbf{n} dA = \frac{1}{v(B(\mathbf{x}, \delta))} \int_{B(\mathbf{x}, \delta)} \operatorname{div} \mathbf{F}(\mathbf{y}) dV.$$

Therefore, since $\operatorname{div} \mathbf{F}(\mathbf{x})$ is a constant,

$$\begin{aligned}& \left| \operatorname{div} \mathbf{F}(\mathbf{x}) - \frac{1}{v(B(\mathbf{x}, \delta))} \int_{\partial B(\mathbf{x}, \delta)} \mathbf{F} \cdot \mathbf{n} dA \right| \\ &= \left| \operatorname{div} \mathbf{F}(\mathbf{x}) - \frac{1}{v(B(\mathbf{x}, \delta))} \int_{B(\mathbf{x}, \delta)} \operatorname{div} \mathbf{F}(\mathbf{y}) dV \right| \\ &= \left| \frac{1}{v(B(\mathbf{x}, \delta))} \int_{B(\mathbf{x}, \delta)} (\operatorname{div} \mathbf{F}(\mathbf{x}) - \operatorname{div} \mathbf{F}(\mathbf{y})) dV \right|\end{aligned}$$

$$\begin{aligned} &\leq \frac{1}{v(B(\mathbf{x}, \delta))} \int_{B(\mathbf{x}, \delta)} |\operatorname{div} \mathbf{F}(\mathbf{x}) - \operatorname{div} \mathbf{F}(\mathbf{y})| dV \\ &\leq \frac{1}{v(B(\mathbf{x}, \delta))} \int_{B(\mathbf{x}, \delta)} \frac{\varepsilon}{2} dV < \varepsilon \end{aligned}$$

whenever ε is small enough due to the continuity of $\operatorname{div} \mathbf{F}$. Since ε is arbitrary, this shows 27.4.

How is this definition independent of coordinates? It only involves geometrical notions of volume and dot product. This is why. Imagine rotating the coordinate axes, keeping all distances the same and expressing everything in terms of the new coordinates. The divergence would still have the same value because of this theorem.

27.4 Some Applications Of The Divergence Theorem

27.4.1 Hydrostatic Pressure

Imagine a fluid which does not move which is acted on by an acceleration, $\mathbf{g}$. Of course the acceleration is usually the acceleration of gravity. Also let the density of the fluid be ρ , a function of position. What can be said about the pressure, p , in the fluid? Let $B(\mathbf{x}, \varepsilon)$ be a small ball centered at the point, $\mathbf{x}$. Then the force the fluid exerts on this ball would equal

$$-\int_{\partial B(\mathbf{x}, \varepsilon)} p \mathbf{n} dA.$$

Here $\mathbf{n}$ is the unit exterior normal at a small piece of $\partial B(\mathbf{x}, \varepsilon)$ having area dA . By the divergence theorem, (see Problem 1 on Page 611) this integral equals

$$-\int_{B(\mathbf{x}, \varepsilon)} \nabla p dV.$$

Also the force acting on this small ball of fluid is

$$\int_{B(\mathbf{x}, \varepsilon)} \rho \mathbf{g} dV.$$

Since it is given that the fluid does not move, the sum of these forces must equal zero. Thus

$$\int_{B(\mathbf{x}, \varepsilon)} \rho \mathbf{g} dV = \int_{B(\mathbf{x}, \varepsilon)} \nabla p dV.$$

Since this must hold for any ball in the fluid of any radius, it must be that

$$\nabla p = \rho \mathbf{g}. \quad (27.5)$$

It turns out that the pressure in a lake at depth z is equal to $62.5z$. This is easy to see from 27.5. In this case, $\mathbf{g} = g\mathbf{k}$ where $g = 32$ feet/sec². The weight of a cubic foot of water is 62.5 pounds. Therefore, the mass in slugs of this water is $62.5/32$. Since it is a cubic foot, this is also the density of the water in slugs per cubic foot. Also, it is normally assumed that water is incompressible¹. Therefore, this is the mass of water at any depth. Therefore,

$$\frac{\partial p}{\partial x} \mathbf{i} + \frac{\partial p}{\partial y} \mathbf{j} + \frac{\partial p}{\partial z} \mathbf{k} = \frac{62.5}{32} \times 32 \mathbf{k}.$$

and so p does not depend on x and y and is only a function of z . It follows $p(0) = 0$, and $p'(z) = 62.5$. Therefore, $p(x, y, z) = 62.5z$. This establishes the claim. This is interesting but 27.5 is more interesting because it does not require ρ to be constant.

¹There is no such thing as an incompressible fluid but this doesn't stop people from making this assumption.

27.4.2 Archimedes Law Of Buoyancy

Archimedes principle states that when a solid body is immersed in a fluid the net force acting on the body by the fluid is directly up and equals the total weight of the fluid displaced.

Denote the set of points in three dimensions occupied by the body as V . Then for dA an increment of area on the surface of this body, the force acting on this increment of area would equal $-p dA \mathbf{n}$ where $\mathbf{n}$ is the exterior unit normal. Therefore, since the fluid does not move,

$$\int_{\partial V} -p \mathbf{n} dA = \int_V -\nabla p dV = \int_V \rho g dV \mathbf{k}$$

Which equals the total weight of the displaced fluid and you note the force is directed upward as claimed. Here ρ is the density and 27.5 is being used. There is an interesting point in the above explanation. Why does the second equation hold? Imagine that V were filled with fluid. Then the equation follows from 27.5 because in this equation $\mathbf{g} = -g\mathbf{k}$.

27.4.3 Equations Of Heat And Diffusion

Let $\mathbf{x}$ be a point in three dimensional space and let (x_1, x_2, x_3) be Cartesian coordinates of this point. Let there be a three dimensional body having density, $\rho = \rho(\mathbf{x}, t)$.

The heat flux, $\mathbf{J}$, in the body is defined as a vector which has the following property.

$$\text{Rate at which heat crosses } S = \int_S \mathbf{J} \cdot \mathbf{n} dA$$

where $\mathbf{n}$ is the unit normal in the desired direction. Thus if V is a three dimensional body,

$$\text{Rate at which heat leaves } V = \int_{\partial V} \mathbf{J} \cdot \mathbf{n} dA$$

where $\mathbf{n}$ is the unit exterior normal.

Fourier's law of heat conduction states that the heat flux, $\mathbf{J}$ satisfies $\mathbf{J} = -k \nabla(u)$ where u is the temperature and $k = k(u, \mathbf{x}, t)$ is called the coefficient of thermal conductivity. This changes depending on the material. It also can be shown by experiment to change with temperature. This equation for the heat flux states that the heat flows from hot places toward colder places in the direction of greatest rate of decrease in temperature. Let $c(\mathbf{x}, t)$ denote the specific heat of the material in the body. This means the amount of heat within V is given by the formula $\int_V \rho(\mathbf{x}, t) c(\mathbf{x}, t) u(\mathbf{x}, t) dV$. Suppose also there are sources for the heat within the material given by $f(\mathbf{x}, u, t)$. If f is positive, the heat is increasing while if f is negative the heat is decreasing. For example such sources could result from a chemical reaction taking place. Then the divergence theorem can be used to verify the following equation for u . Such an equation is called a reaction diffusion equation.

$$\frac{\partial}{\partial t} (\rho(\mathbf{x}, t) c(\mathbf{x}, t) u(\mathbf{x}, t)) = \nabla \cdot (k(u, \mathbf{x}, t) \nabla u(\mathbf{x}, t)) + f(\mathbf{x}, u, t). \quad (27.6)$$

Take an arbitrary V for which the divergence theorem holds. Then the time rate of change of the heat in V is

$$\frac{d}{dt} \int_V \rho(\mathbf{x}, t) c(\mathbf{x}, t) u(\mathbf{x}, t) dV = \int_V \frac{\partial (\rho(\mathbf{x}, t) c(\mathbf{x}, t) u(\mathbf{x}, t))}{\partial t} dV$$

where, as in the preceding example, this is a physical derivation so the consideration of hard mathematics is not necessary. Therefore, from the Fourier law of heat conduction, $\frac{d}{dt} \int_V \rho(\mathbf{x}, t) c(\mathbf{x}, t) u(\mathbf{x}, t) dV =$

$$\begin{aligned} \int_V \frac{\partial (\rho(\mathbf{x}, t) c(\mathbf{x}, t) u(\mathbf{x}, t))}{\partial t} dV &= \overbrace{\int_{\partial V} -\mathbf{J} \cdot \mathbf{n} dA}^{\text{rate at which heat enters}} + \int_V f(\mathbf{x}, u, t) dV \\ &= \int_{\partial V} k \nabla(u) \cdot \mathbf{n} dA + \int_V f(\mathbf{x}, u, t) dV = \int_V (\nabla \cdot (k \nabla(u)) + f) dV. \end{aligned}$$

Since this holds for every sample volume, V it must be the case that the above reaction diffusion equation, 27.6 holds. Note that more interesting equations can be obtained by letting more of the quantities in the equation depend on temperature. However, the above is a fairly hard equation and people usually assume the coefficient of thermal conductivity depends only on $\mathbf{x}$ and that the reaction term, f depends only on $\mathbf{x}$ and t and that ρ and c are constant. Then it reduces to the much easier equation,

$$\frac{\partial}{\partial t} u(\mathbf{x}, t) = \frac{1}{\rho c} \nabla \cdot (k(\mathbf{x}) \nabla u(\mathbf{x}, t)) + f(\mathbf{x}, t). \quad (27.7)$$

This is often referred to as the heat equation. Sometimes there are modifications of this in which k is not just a scalar but a matrix to account for different heat flow properties in different directions. However, they are not much harder than the above. The major mathematical difficulties result from allowing k to depend on temperature.

It is known that the heat equation is not correct even if the thermal conductivity did not depend on u because it implies infinite speed of propagation of heat. However, this does not prevent people from using it.

27.4.4 Balance Of Mass

Let $\mathbf{y}$ be a point in three dimensional space and let (y_1, y_2, y_3) be Cartesian coordinates of this point. Let V be a region in three dimensional space and suppose a fluid having density, $\rho(\mathbf{y}, t)$ and velocity, $\mathbf{v}(\mathbf{y}, t)$ is flowing through this region. Then the mass of fluid leaving V per unit time is given by the area integral, $\int_{\partial V} \rho(\mathbf{y}, t) \mathbf{v}(\mathbf{y}, t) \cdot \mathbf{n} dA$ while the total mass of the fluid enclosed in V at a given time is $\int_V \rho(\mathbf{y}, t) dV$. Also suppose mass originates at the rate $f(\mathbf{y}, t)$ per cubic unit per unit time within this fluid. Then the conclusion which can be drawn through the use of the divergence theorem is the following fundamental equation known as the mass balance equation.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = f(\mathbf{y}, t) \quad (27.8)$$

To see this is so, take an arbitrary V for which the divergence theorem holds. Then the time rate of change of the mass in V is

$$\frac{\partial}{\partial t} \int_V \rho(\mathbf{y}, t) dV = \int_V \frac{\partial \rho(\mathbf{y}, t)}{\partial t} dV$$

where the derivative was taken under the integral sign with respect to t . (This is a physical derivation and therefore, it is not necessary to fuss with the hard mathematics related to the change of limit operations. You should expect this to be true under fairly general conditions because the integral is a sort of sum and the derivative of a sum is

the sum of the derivatives.) Therefore, the rate of change of mass, $\frac{\partial}{\partial t} \int_V \rho(\mathbf{y}, t) dV$, equals

$$\begin{aligned} \int_V \frac{\partial \rho(\mathbf{y}, t)}{\partial t} dV &= \overbrace{- \int_{\partial V} \rho(\mathbf{y}, t) \mathbf{v}(\mathbf{y}, t) \cdot \mathbf{n} dA}^{\text{rate at which mass enters}} + \int_V f(\mathbf{y}, t) dV \\ &= - \int_V (\nabla \cdot (\rho(\mathbf{y}, t) \mathbf{v}(\mathbf{y}, t)) + f(\mathbf{y}, t)) dV. \end{aligned}$$

Since this holds for every sample volume, V it must be the case that the equation of continuity holds. Again, there are interesting mathematical questions here which can be explored but since it is a physical derivation, it is not necessary to dwell too much on them. If all the functions involved are continuous, it is certainly true but it is true under far more general conditions than that.

Also note this equation applies to many situations and f might depend on more than just $\mathbf{y}$ and t . In particular, f might depend also on temperature and the density, ρ . This would be the case for example if you were considering the mass of some chemical and f represented a chemical reaction. Mass balance is a general sort of equation valid in many contexts.

27.4.5 Balance Of Momentum

This example is a little more substantial than the above. It concerns the balance of momentum for a continuum. To see a full description of all the physics involved, you should consult a book on continuum mechanics. The situation is of a material in three dimensions and it deforms and moves about in three dimensions. This means this material is not a rigid body. Let B_0 denote an open set identifying a chunk of this material at time $t = 0$ and let B_t be an open set which identifies the same chunk of material at time $t > 0$.

Let $\mathbf{y}(t, \mathbf{x}) = (y_1(t, \mathbf{x}), y_2(t, \mathbf{x}), y_3(t, \mathbf{x}))$ denote the position with respect to Cartesian coordinates at time t of the point whose position at time $t = 0$ is $\mathbf{x} = (x_1, x_2, x_3)$. The coordinates, $\mathbf{x}$ are sometimes called the reference coordinates and sometimes the material coordinates and sometimes the Lagrangian coordinates. The coordinates, $\mathbf{y}$ are called the Eulerian coordinates or sometimes the spacial coordinates and the function, $(t, \mathbf{x}) \rightarrow \mathbf{y}(t, \mathbf{x})$ is called the motion. Thus

$$\mathbf{y}(0, \mathbf{x}) = \mathbf{x}. \quad (27.9)$$

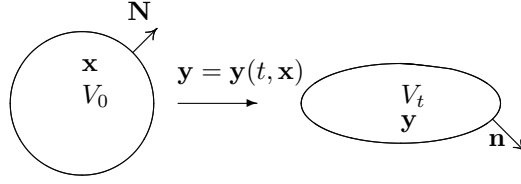
The derivative,

$$D_2 \mathbf{y}(t, \mathbf{x}) \equiv D_{\mathbf{x}} \mathbf{y}(t, \mathbf{x})$$

is called the deformation gradient. Recall the notation means you fix t and consider the function, $\mathbf{x} \rightarrow \mathbf{y}(t, \mathbf{x})$, taking its derivative. Since it is a linear transformation, it is represented by the usual matrix, whose ij^{th} entry is given by

$$F_{ij}(\mathbf{x}) = \frac{\partial y_i(t, \mathbf{x})}{\partial x_j}.$$

Let $\rho(t, \mathbf{y})$ denote the density of the material at time t at the point, $\mathbf{y}$ and let $\rho_0(\mathbf{x})$ denote the density of the material at the point, $\mathbf{x}$. Thus $\rho_0(\mathbf{x}) = \rho(0, \mathbf{x}) = \rho(0, \mathbf{y}(0, \mathbf{x}))$. The first task is to consider the relationship between $\rho(t, \mathbf{y})$ and $\rho_0(\mathbf{x})$. The following picture is useful to illustrate the ideas.



Lemma 27.4.1 $\rho_0(\mathbf{x}) = \rho(t, \mathbf{y}(t, \mathbf{x})) \det(F)$ and in any reasonable physical motion, $\det(F) > 0$.

Proof: Let V_0 represent a small chunk of material at $t = 0$ and let V_t represent the same chunk of material at time t . I will be a little sloppy and refer to V_0 as the small chunk of material at time $t = 0$ and V_t as the chunk of material at time t rather than an open set representing the chunk of material. Then by the change of variables formula for multiple integrals,

$$\int_{V_t} dV = \int_{V_0} |\det(F)| dV.$$

If $\det(F) = 0$ for some t the above formula shows that the chunk of material went from positive volume to zero volume and this is not physically possible. Therefore, it is impossible that $\det(F)$ can equal zero. However, at $t = 0$, $F = I$, the identity because of 27.9. Therefore, $\det(F) = 1$ at $t = 0$ and if it is assumed $t \rightarrow \det(F)$ is continuous it follows by the intermediate value theorem that $\det(F) > 0$ for all t . Of course it is not known for sure this function is continuous but the above shows why it is at least reasonable to expect $\det(F) > 0$.

Now using the change of variables formula,

$$\begin{aligned} \text{mass of } V_t &= \int_{V_t} \rho(t, \mathbf{y}) dV = \int_{V_0} \rho(t, \mathbf{y}(t, \mathbf{x})) \det(F) dV \\ &= \text{mass of } V_0 = \int_{V_0} \rho_0(\mathbf{x}) dV. \end{aligned}$$

Since V_0 is arbitrary, it follows

$$\rho_0(\mathbf{x}) = \rho(t, \mathbf{y}(t, \mathbf{x})) \det(F)$$

as claimed. Note this shows that $\det(F)$ is a magnification factor for the density.

Now consider a small chunk of material, V_t at time t which corresponds to V_0 at time $t = 0$. The total linear momentum of this material at time t is

$$\int_{V_t} \rho(t, \mathbf{y}) \mathbf{v}(t, \mathbf{y}) dV$$

where $\mathbf{v}$ is the velocity. By Newton's second law, the time rate of change of this linear momentum should equal the total force acting on the chunk of material. In the following derivation, $dV(\mathbf{y})$ will indicate the integration is taking place with respect to the variable, $\mathbf{y}$. By Lemma 27.4.1 and the change of variables formula for multiple

integrals

$$\begin{aligned}
& \frac{d}{dt} \left(\int_{V_t} \rho(t, \mathbf{y}) \mathbf{v}(t, \mathbf{y}) dV(\mathbf{y}) \right) \\
&= \frac{d}{dt} \left(\int_{V_0} \rho(t, \mathbf{y}(t, \mathbf{x})) \mathbf{v}(t, \mathbf{y}(t, \mathbf{x})) \det(F) dV(\mathbf{x}) \right) \\
&= \frac{d}{dt} \left(\int_{V_0} \rho_0(\mathbf{x}) \mathbf{v}(t, \mathbf{y}(t, \mathbf{x})) dV(\mathbf{x}) \right) \\
&= \int_{V_0} \rho_0(\mathbf{x}) \left[\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial y_i} \frac{\partial y_i}{\partial t} \right] dV(\mathbf{x}) \\
&= \int_{V_t} \overbrace{\rho(t, \mathbf{y}) \det(F)}^{\rho_0(\mathbf{x})} \left[\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial y_i} \frac{\partial y_i}{\partial t} \right] \frac{1}{\det(F)} dV(\mathbf{y}) \\
&= \int_{V_t} \rho(t, \mathbf{y}) \left[\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial y_i} \frac{\partial y_i}{\partial t} \right] dV(\mathbf{y}).
\end{aligned}$$

Having taken the derivative of the total momentum, it is time to consider the total force acting on the chunk of material.

The force comes from two sources, a body force, $\mathbf{b}$ and a force which acts on the boundary of the chunk of material called a traction force. Typically, the body force is something like gravity in which case, $\mathbf{b} = -g\rho\mathbf{k}$, assuming the Cartesian coordinate system has been chosen in the usual manner. The traction force is of the form

$$\int_{\partial V_t} \mathbf{s}(t, \mathbf{y}, \mathbf{n}) dA$$

where $\mathbf{n}$ is the unit exterior normal. Thus the traction force depends on position, time, and the orientation of the boundary of V_t . Cauchy showed the existence of a linear transformation, $T(t, \mathbf{y})$ such that $T(t, \mathbf{y}) \mathbf{n} = \mathbf{s}(t, \mathbf{y}, \mathbf{n})$. It follows there is a matrix, $T_{ij}(t, \mathbf{y})$ such that the i^{th} component of $\mathbf{s}$ is given by $\mathbf{s}_i(t, \mathbf{y}, \mathbf{n}) = T_{ij}(t, \mathbf{y}) n_j$. Cauchy also showed this matrix is symmetric, $T_{ij} = T_{ji}$. It is called the Cauchy stress. Using Newton's second law to equate the time derivative of the total linear momentum with the applied forces and using the usual repeated index summation convention,

$$\int_{V_t} \rho(t, \mathbf{y}) \left[\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial y_i} \frac{\partial y_i}{\partial t} \right] dV(\mathbf{y}) = \int_{V_t} \mathbf{b}(t, \mathbf{y}) dV(\mathbf{y}) + \int_{\partial V_t} T_{ij}(t, \mathbf{y}) n_j dA.$$

Here is where the divergence theorem is used. In the last integral, the multiplication by n_j is exchanged for the j^{th} partial derivative and an integral over V_t . Thus

$$\int_{V_t} \rho(t, \mathbf{y}) \left[\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial y_i} \frac{\partial y_i}{\partial t} \right] dV(\mathbf{y}) = \int_{V_t} \mathbf{b}(t, \mathbf{y}) dV(\mathbf{y}) + \int_{V_t} \frac{\partial (T_{ij}(t, \mathbf{y}))}{\partial y_j} dV(\mathbf{y}).$$

Since V_t was arbitrary, it follows

$$\begin{aligned}
\rho(t, \mathbf{y}) \left[\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial y_i} \frac{\partial y_i}{\partial t} \right] &= \mathbf{b}(t, \mathbf{y}) + \frac{\partial (T_{ij}(t, \mathbf{y}))}{\partial y_j} \\
&\equiv \mathbf{b}(t, \mathbf{y}) + \operatorname{div}(T)
\end{aligned}$$

where here $\operatorname{div} T$ is a vector whose i^{th} component is given by

$$(\operatorname{div} T)_i = \frac{\partial T_{ij}}{\partial y_j}.$$

The term, $\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial y_i} \frac{\partial y_i}{\partial t}$, is the total derivative with respect to t of the velocity $\mathbf{v}$. Thus you might see this written as

$$\rho \dot{\mathbf{v}} = \mathbf{b} + \operatorname{div}(\mathbf{T}).$$

The above formulation of the balance of momentum involves the spatial coordinates, $\mathbf{y}$ but people also like to formulate momentum balance in terms of the material coordinates, $\mathbf{x}$. Of course this changes everything.

The momentum in terms of the material coordinates is

$$\int_{V_0} \rho_0(\mathbf{x}) \mathbf{v}(t, \mathbf{x}) dV$$

and so, since $\mathbf{x}$ does not depend on t ,

$$\frac{d}{dt} \left(\int_{V_0} \rho_0(\mathbf{x}) \mathbf{v}(t, \mathbf{x}) dV \right) = \int_{V_0} \rho_0(\mathbf{x}) \mathbf{v}_t(t, \mathbf{x}) dV.$$

As indicated earlier, this is a physical derivation and so the mathematical questions related to interchange of limit operations are ignored. This must equal the total applied force. Thus

$$\int_{V_0} \rho_0(\mathbf{x}) \mathbf{v}_t(t, \mathbf{x}) dV = \int_{V_0} \mathbf{b}_0(t, \mathbf{x}) dV + \int_{\partial V_t} T_{ij} n_j dA, \quad (27.10)$$

the first term on the right being the contribution of the body force given per unit volume in the material coordinates and the last term being the traction force discussed earlier. The task is to write this last integral as one over ∂V_0 . For $\mathbf{y} \in \partial V_t$ there is a unit outer normal, $\mathbf{n}$. Here $\mathbf{y} = \mathbf{y}(t, \mathbf{x})$ for $\mathbf{x} \in \partial V_0$. Then define $\mathbf{N}$ to be the unit outer normal to V_0 at the point, $\mathbf{x}$. Near the point $\mathbf{y} \in \partial V_t$ the surface, ∂V_t is given parametrically in the form $\mathbf{y} = \mathbf{y}(s, t)$ for $(s, t) \in D \subseteq \mathbb{R}^2$ and it can be assumed the unit normal to ∂V_t near this point is

$$\mathbf{n} = \frac{\mathbf{y}_s(s, t) \times \mathbf{y}_t(s, t)}{|\mathbf{y}_s(s, t) \times \mathbf{y}_t(s, t)|}$$

with the area element given by $|\mathbf{y}_s(s, t) \times \mathbf{y}_t(s, t)| ds dt$. This is true for $\mathbf{y} \in P_t \subseteq \partial V_t$, a small piece of ∂V_t . Therefore, the last integral in 27.10 is the sum of integrals over small pieces of the form

$$\int_{P_t} T_{ij} n_j dA \quad (27.11)$$

where P_t is parametrized by $\mathbf{y}(s, t)$, $(s, t) \in D$. Thus the integral in 27.11 is of the form

$$\int_D T_{ij}(\mathbf{y}(s, t)) (\mathbf{y}_s(s, t) \times \mathbf{y}_t(s, t))_j ds dt.$$

By the chain rule this equals

$$\int_D T_{ij}(\mathbf{y}(s, t)) \left(\frac{\partial \mathbf{y}}{\partial x_\alpha} \frac{\partial x_\alpha}{\partial s} \times \frac{\partial \mathbf{y}}{\partial x_\beta} \frac{\partial x_\beta}{\partial t} \right)_j ds dt.$$

Remember $\mathbf{y} = \mathbf{y}(t, \mathbf{x})$ and it is always assumed the mapping $\mathbf{x} \rightarrow \mathbf{y}(t, \mathbf{x})$ is one to one and so, since on the surface ∂V_t near $\mathbf{y}$, the points are functions of (s, t) , it follows $\mathbf{x}$ is also a function of (s, t) . Now by the properties of the cross product, this last integral equals

$$\int_D T_{ij}(\mathbf{x}(s, t)) \frac{\partial x_\alpha}{\partial s} \frac{\partial x_\beta}{\partial t} \left(\frac{\partial \mathbf{y}}{\partial x_\alpha} \times \frac{\partial \mathbf{y}}{\partial x_\beta} \right)_j ds dt \quad (27.12)$$

where here $\mathbf{x}(s, t)$ is the point of ∂V_0 which corresponds with $\mathbf{y}(s, t) \in \partial V_t$. Thus

$$T_{ij}(\mathbf{x}(s, t)) = T_{ij}(\mathbf{y}(s, t)).$$

(Perhaps this is a slight abuse of notation because T_{ij} is defined on ∂V_t , not on ∂V_0 , but it avoids introducing extra symbols.) Next 27.12 equals

$$\begin{aligned} & \int_D T_{ij}(\mathbf{x}(s, t)) \frac{\partial x_\alpha}{\partial s} \frac{\partial x_\beta}{\partial t} \varepsilon_{jab} \frac{\partial y_a}{\partial x_\alpha} \frac{\partial y_b}{\partial x_\beta} ds dt \\ &= \int_D T_{ij}(\mathbf{x}(s, t)) \frac{\partial x_\alpha}{\partial s} \frac{\partial x_\beta}{\partial t} \varepsilon_{cab} \delta_{jc} \frac{\partial y_a}{\partial x_\alpha} \frac{\partial y_b}{\partial x_\beta} ds dt \\ &= \int_D T_{ij}(\mathbf{x}(s, t)) \frac{\partial x_\alpha}{\partial s} \frac{\partial x_\beta}{\partial t} \varepsilon_{cab} \overbrace{\frac{\partial y_c}{\partial x_p} \frac{\partial x_p}{\partial y_j}}^{=\delta_{jc}} \frac{\partial y_a}{\partial x_\alpha} \frac{\partial y_b}{\partial x_\beta} ds dt \\ &= \int_D T_{ij}(\mathbf{x}(s, t)) \frac{\partial x_\alpha}{\partial s} \frac{\partial x_\beta}{\partial t} \frac{\partial x_p}{\partial y_j} \varepsilon_{cab} \overbrace{\frac{\partial y_c}{\partial x_p} \frac{\partial y_a}{\partial x_\alpha} \frac{\partial y_b}{\partial x_\beta}}^{=\varepsilon_{p\alpha\beta} \det(F)} ds dt \\ &= \int_D (\det F) T_{ij}(\mathbf{x}(s, t)) \varepsilon_{p\alpha\beta} \frac{\partial x_\alpha}{\partial s} \frac{\partial x_\beta}{\partial t} \frac{\partial x_p}{\partial y_j} ds dt. \end{aligned}$$

Now $\frac{\partial x_p}{\partial y_j} = F_{pj}^{-1}$ and also

$$\varepsilon_{p\alpha\beta} \frac{\partial x_\alpha}{\partial s} \frac{\partial x_\beta}{\partial t} = (\mathbf{x}_s \times \mathbf{x}_t)_p$$

so the result just obtained is of the form

$$\begin{aligned} & \int_D (\det F) F_{pj}^{-1} T_{ij}(\mathbf{x}(s, t)) (\mathbf{x}_s \times \mathbf{x}_t)_p ds dt = \\ & \int_D (\det F) T_{ij}(\mathbf{x}(s, t)) (F^{-T})_{jp} (\mathbf{x}_s \times \mathbf{x}_t)_p ds dt. \end{aligned}$$

This has transformed the integral over P_t to one over P_0 , the part of ∂V_0 which corresponds with P_t . Thus the last integral is of the form

$$\int_{P_0} \det(F) (TF^{-T})_{ip} N_p dA$$

Summing these up over the pieces of ∂V_t and ∂V_0 yields the last integral in 27.10 equals

$$\int_{\partial V_0} \det(F) (TF^{-T})_{ip} N_p dA$$

and so the balance of momentum in terms of the material coordinates becomes

$$\int_{V_0} \rho_0(\mathbf{x}) \mathbf{v}_t(t, \mathbf{x}) dV = \int_{V_0} \mathbf{b}_0(t, \mathbf{x}) dV + \int_{\partial V_0} \det(F) (TF^{-T})_{ip} N_p dA$$

The matrix, $\det(F) (TF^{-T})_{ip}$ is called the Piola Kirchhoff stress, S . An application of the divergence theorem yields

$$\int_{V_0} \rho_0(\mathbf{x}) \mathbf{v}_t(t, \mathbf{x}) dV = \int_{V_0} \mathbf{b}_0(t, \mathbf{x}) dV + \int_{V_0} \frac{\partial (\det(F) (TF^{-T})_{ip})}{\partial x_p} dV.$$

Since V_0 is arbitrary, a balance law for momentum in terms of the material coordinates is obtained

$$\begin{aligned}\rho_0(\mathbf{x}) \mathbf{v}_t(t, \mathbf{x}) &= \mathbf{b}_0(t, \mathbf{x}) + \frac{\partial \left(\det(F) (TF^{-T})_{ip} \right)}{\partial x_p} \\ &= \mathbf{b}_0(t, \mathbf{x}) + \operatorname{div} \left(\det(F) (TF^{-T}) \right) \\ &= \mathbf{b}_0(t, \mathbf{x}) + \operatorname{div} S.\end{aligned}\quad (27.13)$$

As just shown, the relation between the Cauchy stress and the Piola Kirchhoff stress is

$$S = \det(F) (TF^{-T}), \quad (27.14)$$

perhaps not the first thing you would think of.

The main purpose of this presentation is to show how the divergence theorem is used in a significant way to obtain balance laws and to indicate a very interesting direction for further study. To continue, one needs to specify T or S as an appropriate function of things related to the motion, $\mathbf{y}$. Often the thing related to the motion is something called the strain and such relationships are known as constitutive laws.

27.4.6 Frame Indifference

The proper formulation of constitutive laws involves more physical considerations such as frame indifference in which it is required the response of the system cannot depend on the manner in which the Cartesian coordinate system for the spacial coordinates was chosen.

For $Q(t)$ an orthogonal transformation and

$$\mathbf{y}' = \mathbf{q}(t) + Q(t) \mathbf{y}, \quad \mathbf{n}' = Q \mathbf{n},$$

the new spacial coordinates are denoted by $\mathbf{y}'$. Recall an orthogonal transformation is just one which satisfies

$$Q(t)^T Q(t) = Q(t) Q(t)^T = I.$$

The stress has to do with the traction force area density produced by internal changes in the body and has nothing to do with the way the body is observed. Therefore, it is required that

$$T' \mathbf{n}' = QT \mathbf{n}$$

Thus

$$T' Q \mathbf{n} = QT \mathbf{n}$$

Since this is true for any $\mathbf{n}$ normal to the boundary of any piece of the material considered, it must be the case that

$$T' Q = QT$$

and so

$$T' = QTQ^T.$$

This is called frame indifference.

By 27.14, the Piola Kirchhoff stress, S is related to T by

$$S = \det(F) TF^{-T}, \quad F \equiv D_{\mathbf{x}} \mathbf{y}.$$

This stress involves the use of the material coordinates and a normal $\mathbf{N}$ to a piece of the body in reference configuration. Thus $S\mathbf{N}$ gives the force on a part of ∂V_t per unit area on ∂V_0 . Then for a different choice of spacial coordinates, $\mathbf{y}' = \mathbf{q}(t) + Q(t) \mathbf{y}$,

$$S' = \det(F') T' (F')^{-T}$$

but

$$F' = D_{\mathbf{x}} \mathbf{y}' = Q(t) D_{\mathbf{x}} \mathbf{y} = QF$$

and so frame indifference in terms of S is

$$\begin{aligned} S' &= \det(F) QTQ^T (QF)^{-T} \\ &= \det(F) QTQ^T QF^{-T} \\ &= QS \end{aligned}$$

This principle of frame indifference is sometimes ignored and there are certainly interesting mathematical models which have resulted from doing this, but such things cannot be considered physically acceptable.

There are also many other physical properties which can be included and which require a certain form for the constitutive equations. These considerations are outside the scope of this book and require a considerable amount of linear algebra.

There are also balance laws for energy which you may study later but these are more problematic than the balance laws for mass and momentum. However, the divergence theorem is used in these also.

27.4.7 Bernoulli's Principle

Consider a possibly moving fluid with constant density, ρ and let P denote the pressure in this fluid. If B is a part of this fluid the force exerted on B by the rest of the fluid is $\int_{\partial B} -P \mathbf{n} dA$ where $\mathbf{n}$ is the outer normal from B . Assume this is the only force which matters so for example there is no viscosity in the fluid. Thus the Cauchy stress in rectangular coordinates should be

$$T = \begin{pmatrix} -P & 0 & 0 \\ 0 & -P & 0 \\ 0 & 0 & -P \end{pmatrix}.$$

Then

$$\operatorname{div} T = -\nabla P.$$

Also suppose the only body force is from gravity, a force of the form

$$-\rho g \mathbf{k}$$

and so from the balance of momentum

$$\rho \dot{\mathbf{v}} = -\rho g \mathbf{k} - \nabla P(\mathbf{x}). \quad (27.15)$$

Now in all this the coordinates are the spacial coordinates and it is assumed they are rectangular. Thus

$$\mathbf{x} = (x, y, z)^T$$

and $\mathbf{v}$ is the velocity while $\dot{\mathbf{v}}$ is the total derivative of $\mathbf{v} = (v_1, v_2, v_3)^T$ given by $\mathbf{v}_t + v_i \mathbf{v}_{,i}$. Take the dot product of both sides of 27.15 with $\mathbf{v}$. This yields

$$(\rho/2) \frac{d}{dt} |\mathbf{v}|^2 = -\rho g \frac{dz}{dt} - \frac{d}{dt} P(\mathbf{x}).$$

Therefore,

$$\frac{d}{dt} \left(\frac{\rho |\mathbf{v}|^2}{2} + \rho g z + P(\mathbf{x}) \right) = 0$$

and so there is a constant, C' such that

$$\frac{\rho |\mathbf{v}|^2}{2} + \rho g z + P(\mathbf{x}) = C'$$

For convenience define γ to be the weight density of this fluid. Thus $\gamma = \rho g$. Divide by γ . Then

$$\frac{|\mathbf{v}|^2}{2g} + z + \frac{P(\mathbf{x})}{\gamma} = C.$$

this is Bernoulli's² principle. Note how if you keep the height the same, then if you raise $|\mathbf{v}|$, it follows the pressure drops.

This is often used to explain the lift of an airplane wing. The top surface is curved which forces the air to go faster over the top of the wing causing a drop in pressure which creates lift. It is also used to explain the concept of a venturi tube in which the air loses pressure due to being pinched which causes it to flow faster. In many of these applications, the assumptions used in which ρ is constant and there is no other contribution to the traction force on ∂B than pressure so in particular, there is no viscosity, are not correct. However, it is hoped that the effects of these deviations from the ideal situation above are small enough that the conclusions are still roughly true. You can see how using balance of momentum can be used to consider more difficult situations. For example, you might have a body force which is more involved than gravity.

27.4.8 The Wave Equation

As an example of how the balance law of momentum is used to obtain an important equation of mathematical physics, suppose $S = kF$ where k is a constant and F is the deformation gradient and let $\mathbf{u} \equiv \mathbf{y} - \mathbf{x}$. Thus $\mathbf{u}$ is the displacement. Then from 27.13 you can verify the following holds.

$$\rho_0(\mathbf{x}) \mathbf{u}_{tt}(t, \mathbf{x}) = \mathbf{b}_0(t, \mathbf{x}) + k\Delta \mathbf{u}(t, \mathbf{x}) \quad (27.16)$$

In the case where ρ_0 is a constant and $\mathbf{b}_0 = 0$, this yields

$$\mathbf{u}_{tt} - c\Delta \mathbf{u} = \mathbf{0}.$$

The wave equation is $u_{tt} - c\Delta u = 0$ and so the above gives three wave equations, one for each component.

27.4.9 A Negative Observation

Many of the above applications of the divergence theorem are based on the assumption that matter is continuously distributed in a way that the above arguments are correct. In other words, a continuum. However, there is no such thing as a continuum. It has been known for some time now that matter is composed of atoms. It is not continuously distributed through some region of space as it is in the above. Apologists for this contradiction with reality sometimes say to consider enough of the material in question that it is reasonable to think of it as a continuum. This mystical reasoning is then violated as soon as they go from the integral form of the balance laws to the differential equations expressing the traditional formulation of these laws. See Problem 10 below, for example. However, these laws continue to be used and seem to lead to useful physical models which have value in predicting the behavior of physical systems. This is what justifies their use, not any fundamental truth.

²There were many Bernoullis. This is Daniel Bernoulli. He seems to have been nicer than some of the others. Daniel was actually a doctor who was interested in mathematics. He lived from 1700-1782.

27.4.10 Volumes Of Balls In $\mathbb{R}^n$

Recall, $B(\mathbf{x}, r)$ denotes the set of all $\mathbf{y} \in \mathbb{R}^n$ such that $|\mathbf{y} - \mathbf{x}| < r$. By the change of variables formula for multiple integrals or simple geometric reasoning, all balls of radius r have the same volume. Furthermore, simple reasoning or change of variables formula will show that the volume of the ball of radius r equals $\alpha_n r^n$ where α_n will denote the volume of the unit ball in $\mathbb{R}^n$. With the divergence theorem, it is now easy to give a simple relationship between the surface area of the ball of radius r and the volume. By the divergence theorem,

$$\int_{B(\mathbf{0}, r)} \operatorname{div} \mathbf{x} \, dx = \int_{\partial B(\mathbf{0}, r)} \mathbf{x} \cdot \frac{\mathbf{x}}{|\mathbf{x}|} dA$$

because the unit outward normal on $\partial B(\mathbf{0}, r)$ is $\frac{\mathbf{x}}{|\mathbf{x}|}$. Therefore,

$$n\alpha_n r^n = rA(\partial B(\mathbf{0}, r))$$

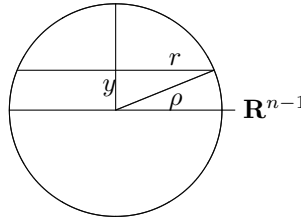
and so

$$A(\partial B(\mathbf{0}, r)) = n\alpha_n r^{n-1}.$$

You recall the surface area of $S^2 \equiv \{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}| = r\}$ is given by $4\pi r^2$ while the volume of the ball, $B(\mathbf{0}, r)$ is $\frac{4}{3}\pi r^3$. This follows the above pattern. You just take the derivative with respect to the radius of the volume of the ball of radius r to get the area of the surface of this ball. Let ω_n denote the area of the sphere $S^{n-1} = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| = 1\}$. I just showed that

$$\omega_n = n\alpha_n.$$

I want to find α_n now and also to get a relationship between ω_n and ω_{n-1} . Consider the following picture of the ball of radius ρ seen on the side.



Taking slices at height y as shown and using that these slices have $n-1$ dimensional area equal to $\alpha_{n-1} r^{n-1}$, it follows

$$\alpha_n \rho^n = 2 \int_0^\rho \alpha_{n-1} (\rho^2 - y^2)^{(n-1)/2} dy$$

In the integral, change variables, letting $y = \rho \cos \theta$. Then

$$\alpha_n \rho^n = 2\rho^n \alpha_{n-1} \int_0^{\pi/2} \sin^n(\theta) d\theta.$$

It follows that

$$\alpha_n = 2\alpha_{n-1} \int_0^{\pi/2} \sin^n(\theta) d\theta. \quad (27.17)$$

Consequently,

$$\omega_n = \frac{2n\omega_{n-1}}{n-1} \int_0^{\pi/2} \sin^n(\theta) d\theta. \quad (27.18)$$

This is a little messier than I would like.

$$\begin{aligned}
 \int_0^{\pi/2} \sin^n(\theta) d\theta &= -\cos\theta \sin^{n-1}\theta \Big|_0^{\pi/2} + (n-1) \int_0^{\pi/2} \cos^2\theta \sin^{n-2}\theta \\
 &= (n-1) \int_0^{\pi/2} (1 - \sin^2\theta) \sin^{n-2}(\theta) d\theta \\
 &= (n-1) \int_0^{\pi/2} \sin^{n-2}(\theta) d\theta - (n-1) \int_0^{\pi/2} \sin^n(\theta) d\theta
 \end{aligned}$$

Hence

$$n \int_0^{\pi/2} \sin^n(\theta) d\theta = (n-1) \int_0^{\pi/2} \sin^{n-2}(\theta) d\theta \quad (27.19)$$

and so 27.18 is of the form

$$\omega_n = 2\omega_{n-1} \int_0^{\pi/2} \sin^{n-2}(\theta) d\theta. \quad (27.20)$$

So what is α_n explicitly? Clearly $\alpha_1 = 2$ and $\alpha_2 = \pi$.

Theorem 27.4.2 $\alpha_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}$ where Γ denotes the gamma function, defined for $\alpha > 0$ by

$$\Gamma(\alpha) \equiv \int_0^\infty e^{-t} t^{\alpha-1} dt.$$

Proof: Recall that $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$. Now note the given formula holds if $n = 1$ because

$$\Gamma\left(\frac{1}{2} + 1\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}.$$

(I leave it as an exercise for you to verify that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.) Thus

$$\alpha_1 = 2 = \frac{\sqrt{\pi}}{\sqrt{\pi}/2}$$

satisfying the formula. Now suppose this formula holds for $k \leq n$. Then from the induction hypothesis, 27.20, 27.19, 27.17 and 27.18,

$$\begin{aligned}
 \alpha_{n+1} &= 2\alpha_n \int_0^{\pi/2} \sin^{n+1}(\theta) d\theta \\
 &= 2\alpha_n \frac{n}{n+1} \int_0^{\pi/2} \sin^{n-1}(\theta) d\theta \\
 &= 2\alpha_n \frac{n}{n+1} \frac{\alpha_{n-1}}{2\alpha_{n-2}} \\
 &= \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} \frac{n}{n+1} \pi^{1/2} \frac{\Gamma(\frac{n-2}{2}+1)}{\Gamma(\frac{n-1}{2}+1)} \\
 &= \frac{\pi^{n/2}}{\Gamma(\frac{n-2}{2}+1) \Gamma(\frac{n}{2})} \frac{n}{n+1} \pi^{1/2} \frac{\Gamma(\frac{n-2}{2}+1)}{\Gamma(\frac{n-1}{2}+1)} \\
 &= 2\pi^{(n+1)/2} \frac{1}{n+1} \frac{1}{\Gamma(\frac{n-1}{2}+1)} \\
 &= \pi^{(n+1)/2} \frac{1}{\Gamma(\frac{n+1}{2})} \frac{1}{\Gamma(\frac{n-1}{2}+1)} \\
 &= \pi^{(n+1)/2} \frac{1}{\Gamma(\frac{n+1}{2}) \Gamma(\frac{n+1}{2})} = \frac{\pi^{(n+1)/2}}{\Gamma(\frac{n+1}{2}+1)}.
 \end{aligned}$$

This proves the theorem.

27.4.11 Electrostatics

Coloumb's law says that the electric field intensity at $\mathbf{x}$ of a charge q located at point, $\mathbf{x}_0$ is given by

$$\mathbf{E} = k \frac{q(\mathbf{x} - \mathbf{x}_0)}{|\mathbf{x} - \mathbf{x}_0|^3}$$

where the electric field intensity is defined to be the force experienced by a unit positive charge placed at the point, $\mathbf{x}$. Note that this is a vector and that its direction depends on the sign of q . It points away from $\mathbf{x}_0$ if q is positive and points toward $\mathbf{x}_0$ if q is negative. The constant, k is a physical constant like the gravitation constant. It has been computed through careful experiments similar to those used with the calculation of the gravitation constant.

The interesting thing about Coloumb's law is that $\mathbf{E}$ is the gradient of a function. In fact,

$$\mathbf{E} = \nabla \left(qk \frac{1}{|\mathbf{x} - \mathbf{x}_0|} \right).$$

The other thing which is significant about this is that in three dimensions and for $\mathbf{x} \neq \mathbf{x}_0$,

$$\nabla \cdot \nabla \left(qk \frac{1}{|\mathbf{x} - \mathbf{x}_0|} \right) = \nabla \cdot \mathbf{E} = 0. \quad (27.21)$$

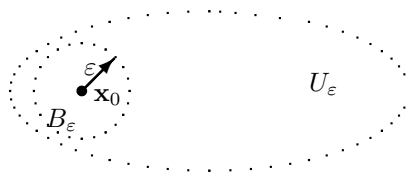
This is left as an exercise for you to verify.

These observations will be used to derive a very important formula for the integral,

$$\int_{\partial U} \mathbf{E} \cdot \mathbf{n} dS$$

where $\mathbf{E}$ is the electric field intensity due to a charge, q located at the point, $\mathbf{x}_0 \in U$, a bounded open set for which the divergence theorem holds.

Let U_ε denote the open set obtained by removing the open ball centered at $\mathbf{x}_0$ which has radius ε where ε is small enough that the following picture is a correct representation of the situation.



Then on the boundary of B_ε the unit outer normal to U_ε is $-\frac{\mathbf{x} - \mathbf{x}_0}{|\mathbf{x} - \mathbf{x}_0|}$. Therefore,

$$\begin{aligned} \int_{\partial B_\varepsilon} \mathbf{E} \cdot \mathbf{n} dS &= - \int_{\partial B_\varepsilon} k \frac{q(\mathbf{x} - \mathbf{x}_0)}{|\mathbf{x} - \mathbf{x}_0|^3} \cdot \frac{\mathbf{x} - \mathbf{x}_0}{|\mathbf{x} - \mathbf{x}_0|} dS \\ &= -kq \int_{\partial B_\varepsilon} \frac{1}{|\mathbf{x} - \mathbf{x}_0|^2} dS = \frac{-kq}{\varepsilon^2} \int_{\partial B_\varepsilon} dS \\ &= \frac{-kq}{\varepsilon^2} 4\pi\varepsilon^2 = -4\pi kq. \end{aligned}$$

Therefore, from the divergence theorem and observation 27.21,

$$-4\pi kq + \int_{\partial U} \mathbf{E} \cdot \mathbf{n} dS = \int_{\partial U_\varepsilon} \mathbf{E} \cdot \mathbf{n} dS = \int_{U_\varepsilon} \nabla \cdot \mathbf{E} dV = 0.$$

It follows that

$$4\pi kq = \int_{\partial U} \mathbf{E} \cdot \mathbf{n} dS.$$

If there are several charges located inside U , say $q_1, q_2, \dots, q_n$, then letting $\mathbf{E}_i$ denote the electric field intensity of the i^{th} charge and $\mathbf{E}$ denoting the total resulting electric field intensity due to all these charges,

$$\begin{aligned} \int_{\partial U} \mathbf{E} \cdot \mathbf{n} dS &= \sum_{i=1}^n \int_{\partial U} \mathbf{E}_i \cdot \mathbf{n} dS \\ &= \sum_{i=1}^n 4\pi k q_i = 4\pi k \sum_{i=1}^n q_i. \end{aligned}$$

This is known as Gauss's law and it is the fundamental result in electrostatics.

27.5 Exercises

1. To prove the divergence theorem, it was shown first that the spacial partial derivative in the volume integral could be exchanged for multiplication by an appropriate component of the exterior normal. This problem starts with the divergence theorem and goes the other direction. Assuming the divergence theorem, holds for a region, V , show that $\int_{\partial V} \mathbf{n} u dA = \int_V \nabla u dV$. Note this implies $\int_V \frac{\partial u}{\partial x} dV = \int_{\partial V} n_1 u dA$.
2. Fick's law for diffusion states the flux of a diffusing species, $\mathbf{J}$ is proportional to the gradient of the concentration, c . Write this law getting the sign right for the constant of proportionality and derive an equation similar to the heat equation for the concentration, c . Typically, c is the concentration of some sort of pollutant or a chemical.
3. Sometimes people consider diffusion in materials which are not homogeneous. This means that

$$\mathbf{J} = -K \nabla c$$

where K is a 3×3 matrix. Thus in terms of components,

$$J_i = - \sum_j K_{ij} \frac{\partial c}{\partial x_j}$$

Here c is the concentration which means the amount of pollutant or whatever is diffusing in a volume is obtained by integrating c over the volume. Derive a formula for a nonhomogeneous model of diffusion basied on the above.

4. Let V be such that the divergence theorem holds. Show that $\int_V \nabla \cdot (u \nabla v) dV = \int_{\partial V} u \frac{\partial v}{\partial \mathbf{n}} dA$ where $\mathbf{n}$ is the exterior normal and $\frac{\partial v}{\partial \mathbf{n}}$ denotes the directional derivative of v in the direction $\mathbf{n}$.
5. Let V be such that the divergence theorem holds. Show that

$$\int_V (v \nabla^2 u - u \nabla^2 v) dV = \int_{\partial V} \left(v \frac{\partial u}{\partial \mathbf{n}} - u \frac{\partial v}{\partial \mathbf{n}} \right) dA$$

where $\mathbf{n}$ is the exterior normal and $\frac{\partial u}{\partial \mathbf{n}}$ is defined in Problem 4.

6. Let V be a ball and suppose $\nabla^2 u = f$ in V while $u = g$ on ∂V . Show that there is at most one solution to this boundary value problem which is C^2 in V and continuous on V with its boundary. **Hint:** You might consider $w = u - v$ where

u and v are solutions to the problem. Then use the result of Problem 4 and the identity

$$w \nabla^2 w = \nabla \cdot (w \nabla w) - \nabla w \cdot \nabla w$$

to conclude $\nabla w = 0$. Then show this implies w must be a constant by considering $h(t) = w(t\mathbf{x} + (1-t)\mathbf{y})$ and showing h is a constant. Alternatively, you might consider the maximum principle.

7. Show that $\int_{\partial V} \nabla \times \mathbf{v} \cdot \mathbf{n} dA = 0$ where V is a region for which the divergence theorem holds and $\mathbf{v}$ is a C^2 vector field.
8. Let $\mathbf{F}(x, y, z) = (x, y, z)$ be a vector field in $\mathbb{R}^3$ and let V be a three dimensional shape and let $\mathbf{n} = (n_1, n_2, n_3)$. Show that $\int_{\partial V} (xn_1 + yn_2 + zn_3) dA = 3 \times \text{volume of } V$.
9. Let $\mathbf{F} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ and let V denote the tetrahedron formed by the planes, $x = 0, y = 0, z = 0$, and $\frac{1}{3}x + \frac{1}{3}y + \frac{1}{3}z = 1$. Verify the divergence theorem for this example.
10. Suppose $f : U \rightarrow \mathbb{R}$ is continuous where U is some open set and for all $B \subseteq U$ where B is a ball, $\int_B f(\mathbf{x}) dV = 0$. Show that this implies $f(\mathbf{x}) = 0$ for all $\mathbf{x} \in U$.
11. Let U denote the box centered at $(0, 0, 0)$ with sides parallel to the coordinate planes which has width 4, length 2 and height 3. Find the flux integral $\int_{\partial U} \mathbf{F} \cdot \mathbf{n} dS$ where $\mathbf{F} = (x + 3, 2y, 3z)$. **Hint:** If you like, you might want to use the divergence theorem.
12. Find the flux out of the cylinder whose base is $x^2 + y^2 \leq 1$ which has height 2 of the vector field $\mathbf{F} = (xy, zy, z^2 + x)$.
13. Find the flux out of the ball of radius 4 centered at $\mathbf{0}$ of the vector field $\mathbf{F} = (x, zy, z + x)$.
14. Verify 27.16 from 27.13 and the assumption that $S = kF$.
15. Show that if $u_k, k = 1, 2, \dots, n$ each satisfies 27.7 with $f = 0$ then for any choice of constants, $c_1, \dots, c_n$, so does

$$\sum_{k=1}^n c_k u_k.$$

16. Suppose $k(\mathbf{x}) = k$, a constant and $f = 0$. Then in one dimension, the heat equation is of the form $u_t = \alpha u_{xx}$. Show that $u(x, t) = e^{-\alpha n^2 t} \sin(nx)$ satisfies the heat equation³.
17. Let U be a three dimensional region for which the divergence theorem holds. Show that

$$\int_U \nabla \times \mathbf{F} dx = \int_{\partial U} \mathbf{n} \times \mathbf{F} dS$$

where $\mathbf{n}$ is the unit outer normal.

³Fourier, an officer in Napoleon's army studied solutions to the heat equation back in 1813. He was interested in heat flow in cannons. He sought to find solutions by adding up infinitely many solutions of this form. Actually, it was a little more complicated because cannons are not one dimensional but it was the beginning of the study of Fourier series, a topic which fascinated mathematicians for the next 150 years and motivated the development of analysis.

18. In a linear, viscous, incompressible fluid, the Cauchy stress is of the form

$$T_{ij}(t, \mathbf{y}) = \lambda \left(\frac{v_{i,j}(t, \mathbf{y}) + v_{j,i}(t, \mathbf{y})}{2} \right) - p \delta_{ij}$$

where p is the pressure, δ_{ij} equals 0 if $i \neq j$ and 1 if $i = j$, and the comma followed by an index indicates the partial derivative with respect to that variable and $\mathbf{v}$ is the velocity. Thus

$$v_{i,j} = \frac{\partial v_i}{\partial y_j}$$

Also, p denotes the pressure. Show, using the balance of mass equation that incompressible implies $\text{div } \mathbf{v} = 0$. Next show that the balance of momentum equation requires

$$\rho \dot{\mathbf{v}} - \frac{\lambda}{2} \Delta \mathbf{v} = \rho \left[\frac{\partial \mathbf{v}}{\partial t} + \frac{\partial \mathbf{v}}{\partial y_i} v_i \right] - \frac{\lambda}{2} \Delta \mathbf{v} = \mathbf{b} - \nabla p.$$

This is the famous Navier Stokes equation for incompressible viscous linear fluids. There are still open questions related to this equation, one of which is worth \$1,000,000 at this time.

Stokes And Green's Theorems

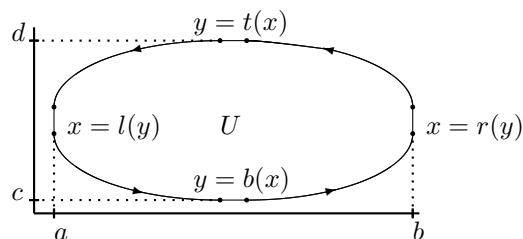
28.0.1 Outcomes

1. Recall and verify Green's theorem.
2. Apply Green's theorem to evaluate line integrals.
3. Apply Green's theorem to find the area of a region.
4. Explain what is meant by the curl of a vector field.
5. Evaluate the curl of a vector field.
6. Derive and apply formulas involving divergence, gradient and curl.
7. Recall and use Stoke's theorem.
8. Apply Stoke's theorem to calculate the circulation or work of a vector field around a simple closed curve.
9. Recall and apply the fundamental theorem for line integrals.
10. Determine whether a vector field is a gradient using the curl test.
11. Recover a function from its gradient when possible.

28.1 Green's Theorem

Green's theorem is an important theorem which relates line integrals to integrals over a surface in the plane. It can be used to establish the seemingly more general Stoke's theorem but is interesting for it's own sake. Historically, theorems like it were important in the development of complex analysis. I will first establish Green's theorem for regions of a particular sort and then show that the theorem holds for many other regions also. Suppose a region is of the form indicated in the following picture in which

$$\begin{aligned} U &= \{(x, y) : x \in (a, b) \text{ and } y \in (b(x), t(x))\} \\ &= \{(x, y) : y \in (c, d) \text{ and } x \in (l(y), r(y))\}. \end{aligned}$$



I will refer to such a region as being convex in both the x and y directions.

Lemma 28.1.1 *Let $\mathbf{F}(x, y) \equiv (P(x, y), Q(x, y))$ be a C^1 vector field defined near U where U is a region of the sort indicated in the above picture which is convex in both the x and y directions. Suppose also that the functions, r, l, t , and b in the above picture are all C^1 functions and denote by ∂U the boundary of U oriented such that the direction of motion is counter clockwise. (As you walk around U on ∂U , the points of U are on your left.) Then*

$$\begin{aligned} \int_{\partial U} P dx + Q dy &\equiv \\ \int_{\partial U} \mathbf{F} \cdot d\mathbf{R} &= \int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA. \end{aligned} \quad (28.1)$$

Proof: First consider the right side of 28.1.

$$\begin{aligned} &\int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA \\ &= \int_c^d \int_{l(y)}^{r(y)} \frac{\partial Q}{\partial x} dx dy - \int_a^b \int_{b(x)}^{t(x)} \frac{\partial P}{\partial y} dy dx \\ &= \int_c^d (Q(r(y), y) - Q(l(y), y)) dy + \int_a^b (P(x, b(x)) - P(x, t(x))) dx. \end{aligned} \quad (28.2)$$

Now consider the left side of 28.1. Denote by V the vertical parts of ∂U and by H the horizontal parts.

$$\begin{aligned} &\int_{\partial U} \mathbf{F} \cdot d\mathbf{R} = \\ &= \int_{\partial U} ((0, Q) + (P, 0)) \cdot d\mathbf{R} \\ &= \int_c^d (0, Q(r(s), s)) \cdot (r'(s), 1) ds + \int_H (0, Q(r(s), s)) \cdot (\pm 1, 0) ds \\ &\quad - \int_c^d (0, Q(l(s), s)) \cdot (l'(s), 1) ds + \int_a^b (P(s, b(s)), 0) \cdot (1, b'(s)) ds \\ &\quad + \int_V (P(s, b(s)), 0) \cdot (0, \pm 1) ds - \int_a^b (P(s, t(s)), 0) \cdot (1, t'(s)) ds \\ &= \int_c^d Q(r(s), s) ds - \int_c^d Q(l(s), s) ds + \int_a^b P(s, b(s)) ds - \int_a^b P(s, t(s)) ds \end{aligned}$$

which coincides with 28.2. This proves the lemma.

Corollary 28.1.2 *Let everything be the same as in Lemma 28.1.1 but only assume the functions r, l, t , and b are continuous and piecewise C^1 functions. Then the conclusion this lemma is still valid.*

Proof: The details are left for you. All you have to do is to break up the various line integrals into the sum of integrals over sub intervals on which the function of interest is C^1 .

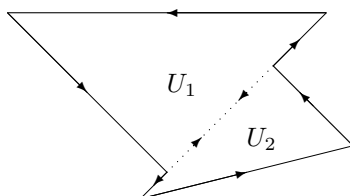
From this corollary, it follows 28.1 is valid for any triangle for example.

Now suppose 28.1 holds for $U_1, U_2, \dots, U_m$ and the open sets, U_k have the property that no two have nonempty intersection and their boundaries intersect only in a finite

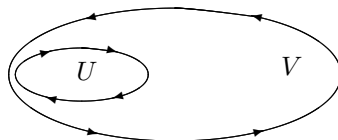
number of piecewise smooth curves. Then 28.1 must hold for $U \equiv \cup_{i=1}^m U_i$, the union of these sets. This is because

$$\begin{aligned} \int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA &= \\ &= \sum_{k=1}^m \int_{U_k} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA \\ &= \sum_{k=1}^m \int_{\partial U_k} \mathbf{F} \cdot d\mathbf{R} = \int_{\partial U} \mathbf{F} \cdot d\mathbf{R} \end{aligned}$$

because if $\Gamma = \partial U_k \cap \partial U_j$, then its orientation as a part of ∂U_k is opposite to its orientation as a part of ∂U_j and consequently the line integrals over Γ will cancel, points of Γ also not being in ∂U . As an illustration, consider the following picture for two such U_k .



Similarly, if $U \subseteq V$ and if also $\partial U \subseteq V$ and both U and V are open sets for which 28.1 holds, then the open set, $V \setminus (U \cup \partial U)$ consisting of what is left in V after deleting U along with its boundary also satisfies 28.1. Roughly speaking, you can drill holes in a region for which 28.1 holds and get another region for which this continues to hold provided 28.1 holds for the holes. To see why this is so, consider the following picture which typifies the situation just described.



Then

$$\begin{aligned} \int_{\partial V} \mathbf{F} \cdot d\mathbf{R} &= \int_V \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA \\ &= \int_U \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA + \int_{V \setminus U} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA \\ &= \int_{\partial U} \mathbf{F} \cdot d\mathbf{R} + \int_{V \setminus U} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA \end{aligned}$$

and so

$$\int_{V \setminus U} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA = \int_{\partial V} \mathbf{F} \cdot d\mathbf{R} - \int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$$

which equals

$$\int_{\partial(V \setminus U)} \mathbf{F} \cdot d\mathbf{R}$$

where ∂V is oriented as shown in the picture. (If you walk around the region, $V \setminus U$ with the area on the left, you get the indicated orientation for this curve.)

You can see that 28.1 is valid quite generally. This verifies the following theorem.

Theorem 28.1.3 (*Green's Theorem*) Let U be an open set in the plane and let ∂U be piecewise smooth and let $\mathbf{F}(x, y) = (P(x, y), Q(x, y))$ be a C^1 vector field defined near U . Then it is often¹ the case that

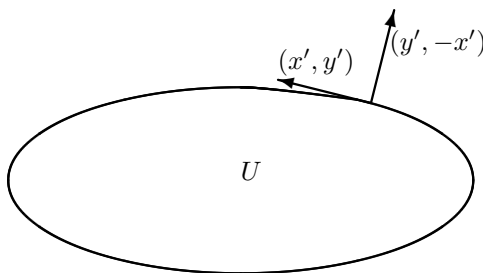
$$\int_{\partial U} \mathbf{F} \cdot d\mathbf{R} = \int_U \left(\frac{\partial Q}{\partial x}(x, y) - \frac{\partial P}{\partial y}(x, y) \right) dA.$$

Here is an alternate proof of Green's theorem from the divergence theorem.

Theorem 28.1.4 (*Green's Theorem*) Let U be an open set in the plane and let ∂U be piecewise smooth and let $\mathbf{F}(x, y) = (P(x, y), Q(x, y))$ be a C^1 vector field defined near U . Then it is often the case that

$$\int_{\partial U} \mathbf{F} \cdot d\mathbf{R} = \int_U \left(\frac{\partial Q}{\partial x}(x, y) - \frac{\partial P}{\partial y}(x, y) \right) dA.$$

Proof: Suppose the divergence theorem holds for U . Consider the following picture.



Since it is assumed that motion around U is counter clockwise, the tangent vector, (x', y') is as shown. The unit **exterior normal** is a multiple of

$$(x', y', 0) \times (0, 0, 1) = (y', -x', 0).$$

Use your right hand and the geometric description of the cross product to verify this. This would be the case at all the points where the unit exterior normal exists.

Now let $\mathbf{F}(x, y) = (Q(x, y), -P(x, y))$. Also note the area (length) element on the bounding curve ∂U is $\sqrt{(x')^2 + (y')^2} dt$. Suppose the boundary of U consists of m smooth curves, the i^{th} of which is parameterized by (x_i, y_i) with the parameter, $t \in [a_i, b_i]$. Then by the divergence theorem,

$$\int_U (Q_x - P_y) dA = \int_U \operatorname{div}(\mathbf{F}) dA = \int_{\partial U} \mathbf{F} \cdot \mathbf{n} dS$$

¹For a general version see the advanced calculus book by Apostol. The general versions involve the concept of a rectifiable (finite length) Jordan curve.

$$\begin{aligned}
&= \sum_{i=1}^m \int_{a_i}^{b_i} (Q(x_i(t), y_i(t)), -P(x_i(t), y_i(t))) \\
&\quad \cdot \frac{1}{\sqrt{(x'_i)^2 + (y'_i)^2}} (y'_i, -x'_i) \overbrace{\sqrt{(x'_i)^2 + (y'_i)^2} dt}^{dS} \\
&= \sum_{i=1}^m \int_{a_i}^{b_i} (Q(x_i(t), y_i(t)), -P(x_i(t), y_i(t))) \cdot (y'_i, -x'_i) dt \\
&= \sum_{i=1}^m \int_{a_i}^{b_i} Q(x_i(t), y_i(t)) y'_i(t) + P(x_i(t), y_i(t)) x'_i(t) dt \equiv \int_{\partial U} P dx + Q dy
\end{aligned}$$

This proves Green's theorem from the divergence theorem.

Proposition 28.1.5 *Let U be an open set in $\mathbb{R}^2$ for which Green's theorem holds. Then*

$$\text{Area of } U = \int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$$

where $\mathbf{F}(x, y) = \frac{1}{2}(-y, x)$, or $(-y, 0)$.

Proof: This follows immediately from Green's theorem.

Example 28.1.6 *Use Proposition 28.1.5 to find the area of the ellipse*

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1.$$

You can parameterize the boundary of this ellipse as

$$x = a \cos t, \quad y = b \sin t, \quad t \in [0, 2\pi].$$

Then from Proposition 28.1.5,

$$\begin{aligned}
\text{Area equals} &= \frac{1}{2} \int_0^{2\pi} (-b \sin t, a \cos t) \cdot (-a \sin t, b \cos t) dt \\
&= \frac{1}{2} \int_0^{2\pi} (ab) dt = \pi ab.
\end{aligned}$$

Example 28.1.7 *Find $\int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$ where U is the set,*

$$\{(x, y) : x^2 + 3y^2 \leq 9\}$$

and $\mathbf{F}(x, y) = (y, -x)$.

One way to do this is to parameterize the boundary of U and then compute the line integral directly. It is easier to use Green's theorem. The desired line integral equals

$$\int_U ((-1) - 1) dA = -2 \int_U dA.$$

Now U is an ellipse having area equal to $3\sqrt{3}$ and so the answer is $-6\sqrt{3}$.

Example 28.1.8 *Find $\int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$ where U is the set, $\{(x, y) : 2 \leq x \leq 4, 0 \leq y \leq 3\}$ and $\mathbf{F}(x, y) = (x \sin y, y^3 \cos x)$.*

From Green's theorem this line integral equals

$$\int_2^4 \int_0^3 (-y^3 \sin x - x \cos y) dy dx = \frac{81}{4} \cos 4 - 6 \sin 3 - \frac{81}{4} \cos 2.$$

This is much easier than computing the line integral because you don't have to break the boundary in pieces and consider each separately.

Example 28.1.9 Find $\int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$ where U is the set,

$$\{(x, y) : 2 \leq x \leq 4, x \leq y \leq 3\}$$

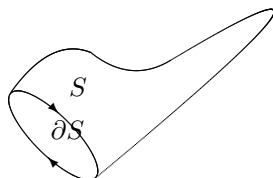
and $\mathbf{F}(x, y) = (x \sin y, y \sin x)$.

From Green's theorem this line integral equals

$$\int_2^4 \int_x^3 (y \cos x - x \cos y) dy dx = -\frac{3}{2} \sin 4 - 6 \sin 3 - 8 \cos 4 - \frac{9}{2} \sin 2 + 4 \cos 2.$$

28.2 Stoke's Theorem From Green's Theorem

Stoke's theorem is a generalization of Green's theorem which relates the integral over a surface to the integral around the boundary of the surface. These terms are a little different from what occurs in $\mathbb{R}^2$. To describe this, consider a sock. The surface is the sock and its boundary will be the edge of the opening of the sock in which you place your foot. Another way to think of this is to imagine a region in $\mathbb{R}^2$ of the sort discussed above for Green's theorem. Suppose it is on a sheet of rubber and the sheet of rubber is stretched in three dimensions. The boundary of the resulting surface is the result of the stretching applied to the boundary of the original region in $\mathbb{R}^2$. Here is a picture describing the situation.



Recall the following definition of the curl of a vector field.

Definition 28.2.1 Let

$$\mathbf{F}(x, y, z) = (F_1(x, y, z), F_2(x, y, z), F_3(x, y, z))$$

be a C^1 vector field defined on an open set, V in $\mathbb{R}^3$. Then

$$\nabla \times \mathbf{F} \equiv \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_1 & F_2 & F_3 \end{vmatrix} \equiv \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z} \right) \mathbf{i} + \left(\frac{\partial F_1}{\partial z} - \frac{\partial F_3}{\partial x} \right) \mathbf{j} + \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \right) \mathbf{k}.$$

This is also called $\text{curl}(\mathbf{F})$ and written as indicated, $\nabla \times \mathbf{F}$.

The following lemma gives the fundamental identity which will be used in the proof of Stoke's theorem.

Lemma 28.2.2 Let $\mathbf{R} : U \rightarrow V \subseteq \mathbb{R}^3$ where U is an open subset of $\mathbb{R}^2$ and V is an open subset of $\mathbb{R}^3$. Suppose $\mathbf{R}$ is C^2 and let $\mathbf{F}$ be a C^1 vector field defined in V .

$$(\mathbf{R}_u \times \mathbf{R}_v) \cdot (\nabla \times \mathbf{F})(\mathbf{R}(u, v)) = ((\mathbf{F} \circ \mathbf{R})_u \cdot \mathbf{R}_v - (\mathbf{F} \circ \mathbf{R})_v \cdot \mathbf{R}_u)(u, v). \quad (28.3)$$

Proof: Start with the left side and let $x_i = R_i(u, v)$ for short.

$$\begin{aligned} (\mathbf{R}_u \times \mathbf{R}_v) \cdot (\nabla \times \mathbf{F})(\mathbf{R}(u, v)) &= \varepsilon_{ijk} x_{ju} x_{kv} \varepsilon_{irs} \frac{\partial F_s}{\partial x_r} \\ &= (\delta_{jr} \delta_{ks} - \delta_{js} \delta_{kr}) x_{ju} x_{kv} \frac{\partial F_s}{\partial x_r} \\ &= x_{ju} x_{kv} \frac{\partial F_k}{\partial x_j} - x_{ju} x_{kv} \frac{\partial F_j}{\partial x_k} \\ &= \mathbf{R}_v \cdot \frac{\partial (\mathbf{F} \circ \mathbf{R})}{\partial u} - \mathbf{R}_u \cdot \frac{\partial (\mathbf{F} \circ \mathbf{R})}{\partial v} \end{aligned}$$

which proves 28.3.

The proof of Stoke's theorem given next follows [7]. First, it is convenient to give a definition.

Definition 28.2.3 A vector valued function, $\mathbf{R} : U \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^n$ is said to be in $C^k(\bar{U}, \mathbb{R}^n)$ if it is the restriction to $\bar{U}$ of a vector valued function which is defined on $\mathbb{R}^m$ and is C^k . That is, this function has continuous partial derivatives up to order k .

Theorem 28.2.4 (Stoke's Theorem) Let U be any region in $\mathbb{R}^2$ for which the conclusion of Green's theorem holds and let $\mathbf{R} \in C^2(\bar{U}, \mathbb{R}^3)$ be a one to one function satisfying $|(\mathbf{R}_u \times \mathbf{R}_v)(u, v)| \neq 0$ for all $(u, v) \in U$ and let S denote the surface,

$$\begin{aligned} S &\equiv \{\mathbf{R}(u, v) : (u, v) \in U\}, \\ \partial S &\equiv \{\mathbf{R}(u, v) : (u, v) \in \partial U\} \end{aligned}$$

where the orientation on ∂S is consistent with the counter clockwise orientation on ∂U (U is on the left as you walk around ∂U). Then for $\mathbf{F}$ a C^1 vector field defined near S ,

$$\int_{\partial S} \mathbf{F} \cdot d\mathbf{R} = \int_S \text{curl}(\mathbf{F}) \cdot \mathbf{n} dS$$

where $\mathbf{n}$ is the normal to S defined by

$$\mathbf{n} \equiv \frac{\mathbf{R}_u \times \mathbf{R}_v}{|\mathbf{R}_u \times \mathbf{R}_v|}.$$

Proof: Letting C be an oriented part of ∂U having parametrization,

$$\mathbf{r}(t) \equiv (u(t), v(t))$$

for $t \in [\alpha, \beta]$ and letting $\mathbf{R}(C)$ denote the oriented part of ∂S corresponding to C ,

$$\begin{aligned} \int_{\mathbf{R}(C)} \mathbf{F} \cdot d\mathbf{R} &= \\ &= \int_{\alpha}^{\beta} \mathbf{F}(\mathbf{R}(u(t), v(t))) \cdot (\mathbf{R}_u u'(t) + \mathbf{R}_v v'(t)) dt \\ &= \int_{\alpha}^{\beta} \mathbf{F}(\mathbf{R}(u(t), v(t))) \mathbf{R}_u(u(t), v(t)) u'(t) dt \\ &\quad + \int_{\alpha}^{\beta} \mathbf{F}(\mathbf{R}(u(t), v(t))) \mathbf{R}_v(u(t), v(t)) v'(t) dt \\ &= \int_C ((\mathbf{F} \circ \mathbf{R}) \cdot \mathbf{R}_u, (\mathbf{F} \circ \mathbf{R}) \cdot \mathbf{R}_v) \cdot d\mathbf{r}. \end{aligned}$$

Since this holds for each such piece of ∂U , it follows

$$\int_{\partial S} \mathbf{F} \cdot d\mathbf{R} = \int_{\partial U} ((\mathbf{F} \circ \mathbf{R}) \cdot \mathbf{R}_u, (\mathbf{F} \circ \mathbf{R}) \cdot \mathbf{R}_v) \cdot d\mathbf{r}.$$

By the assumption that the conclusion of Green's theorem holds for U , this equals

$$\begin{aligned} & \int_U [((\mathbf{F} \circ \mathbf{R}) \cdot \mathbf{R}_v)_u - ((\mathbf{F} \circ \mathbf{R}) \cdot \mathbf{R}_u)_v] dA \\ &= \int_U [(\mathbf{F} \circ \mathbf{R})_u \cdot \mathbf{R}_v + (\mathbf{F} \circ \mathbf{R}) \cdot \mathbf{R}_{vu} - (\mathbf{F} \circ \mathbf{R}) \cdot \mathbf{R}_{uv} - (\mathbf{F} \circ \mathbf{R})_v \cdot \mathbf{R}_u] dA \\ &= \int_U [(\mathbf{F} \circ \mathbf{R})_u \cdot \mathbf{R}_v - (\mathbf{F} \circ \mathbf{R})_v \cdot \mathbf{R}_u] dA \end{aligned}$$

the last step holding by equality of mixed partial derivatives, a result of the assumption that $\mathbf{R}$ is C^2 . Now by Lemma 28.2.2, this equals

$$\begin{aligned} & \int_U (\mathbf{R}_u \times \mathbf{R}_v) \cdot (\nabla \times \mathbf{F}) dA \\ &= \int_U \nabla \times \mathbf{F} \cdot (\mathbf{R}_u \times \mathbf{R}_v) dA \\ &= \int_S \nabla \times \mathbf{F} \cdot \mathbf{n} dS \end{aligned}$$

because $dS = |(\mathbf{R}_u \times \mathbf{R}_v)| dA$ and $\mathbf{n} = \frac{(\mathbf{R}_u \times \mathbf{R}_v)}{|(\mathbf{R}_u \times \mathbf{R}_v)|}$. Thus

$$\begin{aligned} (\mathbf{R}_u \times \mathbf{R}_v) dA &= \frac{(\mathbf{R}_u \times \mathbf{R}_v)}{|(\mathbf{R}_u \times \mathbf{R}_v)|} |(\mathbf{R}_u \times \mathbf{R}_v)| dA \\ &= \mathbf{n} dS. \end{aligned}$$

This proves Stoke's theorem.

Note that there is no mention made in the final result that $\mathbf{R}$ is C^2 . Therefore, it is not surprising that versions of this theorem are valid in which this assumption is not present. It is possible to obtain extremely general versions of Stoke's theorem if you use the Lebesgue integral.

28.2.1 The Normal And The Orientation

Stoke's theorem as just presented needs no apology. However, it is helpful in applications to have some additional geometric insight.

To begin with, suppose the surface S of interest is a parallelogram in $\mathbb{R}^3$ determined by the two vectors $\mathbf{a}, \mathbf{b}$. Thus $S = \mathbf{R}(Q)$ where $Q = [0, 1] \times [0, 1]$ is the unit square and for $(u, v) \in Q$,

$$\mathbf{R}(u, v) \equiv u\mathbf{a} + v\mathbf{b} + \mathbf{p},$$

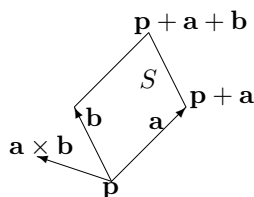
the point $\mathbf{p}$ being a corner of the parallelogram S . Then orient ∂S consistent with the counter clockwise orientation on ∂Q . Thus, following this orientation on S you go from $\mathbf{p}$ to $\mathbf{p} + \mathbf{a}$ to $\mathbf{p} + \mathbf{a} + \mathbf{b}$ to $\mathbf{p} + \mathbf{b}$ to $\mathbf{p}$. Then Stoke's theorem implies that with this orientation on ∂S ,

$$\int_{\partial S} \mathbf{F} \cdot d\mathbf{R} = \int_S \nabla \times \mathbf{F} \cdot \mathbf{n} ds$$

where

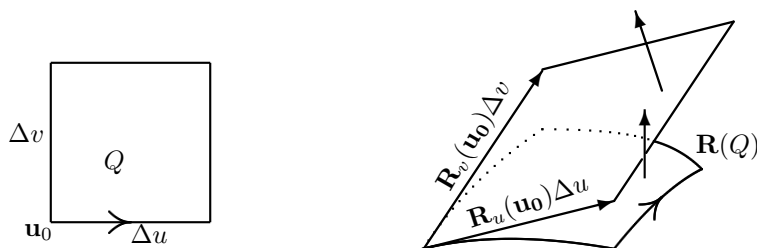
$$\mathbf{n} = \mathbf{R}_u \times \mathbf{R}_v / |\mathbf{R}_u \times \mathbf{R}_v| = \mathbf{a} \times \mathbf{b} / |\mathbf{a} \times \mathbf{b}|.$$

Now recall $\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}$ forms a right hand system.



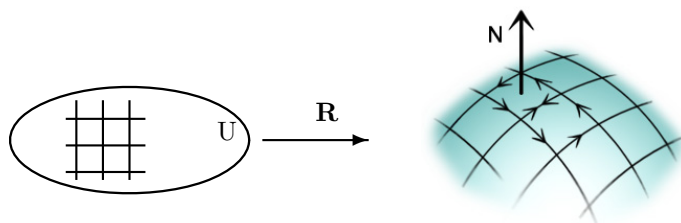
Thus, if you were walking around ∂S in the direction of the orientation with your left hand over the surface S , the normal vector $\mathbf{a} \times \mathbf{b}$ would be pointing in the direction of your head.

More generally, if S is a surface which is not necessarily a parallelogram but is instead as described in Theorem 28.2.4, you could consider a **small** rectangle Q contained in U and orient the boundary of $\mathbf{R}(Q)$ consistent with the counter clockwise orientation on ∂Q . Then if Q is small enough, as you walk around $\partial \mathbf{R}(Q)$ in the direction of the described orientation with your left hand over $\mathbf{R}(Q)$, your head points roughly in the direction of $\mathbf{R}_u \times \mathbf{R}_v$.

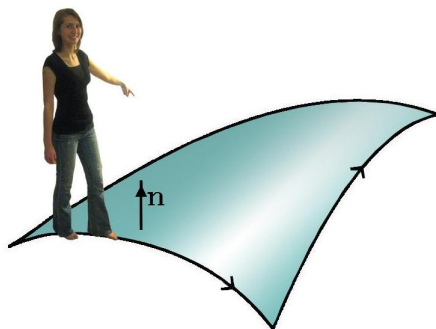


As explained above, this is true of the tangent parallelogram, and by continuity of $\mathbf{R}_v, \mathbf{R}_u$, the normals to the surface $\mathbf{R}(Q)$ $\mathbf{R}_u \times \mathbf{R}_v(\mathbf{u})$ for $\mathbf{u} \in Q$ will still point roughly in the same direction as your head if you walk in the indicated direction over $\partial \mathbf{R}(Q)$, meaning the angle between the vector from your feet to your head and the vector $\mathbf{R}_u \times \mathbf{R}_v(\mathbf{u})$ is less than $\pi/2$.

You can imagine filling U with such non-overlapping regions Q_i . Then orienting $\partial \mathbf{R}(Q_i)$ consistent with the counter clockwise orientation on Q_i , and adding the resulting line integrals, the line integrals over the common sides cancel as indicated in the following picture and the result is the line integral over ∂S .



Thus there is a simple relation between the field of normal vectors on S and the orientation of ∂S . It is simply this. If you walk along ∂S in the direction mandated by the orientation, with your left hand over the surface, the nearby normal vectors in Stoke's theorem will point roughly in the direction of your head.

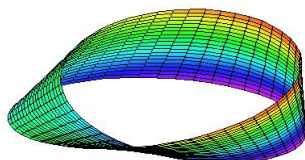


This also illustrates that you can **define** an orientation for ∂S by specifying a field of unit normal vectors for the surface, which varies continuously over the surface, and require that the motion over the boundary of the surface is such that your head points roughly in the direction of nearby normal vectors as you walk along the boundary with your left hand over S . The existence of such a continuous field of normal vectors is what constitutes an **orientable** surface.

28.2.2 The Mobeus Band

It turns out there are more general formulations of Stoke's theorem than what is presented above. However, it is always necessary for the surface, S to be **orientable**. This means it is possible to obtain a vector field of unit normals to the surface which is a continuous function of position on S .

An example of a surface which is not orientable is the famous Mobeus band, obtained by taking a long rectangular piece of paper and glueing the ends together after putting a twist in it. Here is a picture of one.



There is something quite interesting about this Mobeus band and this is that it can be written parametrically with a simple parameter domain. The picture above is a maple graph of the parametrically defined surface

$$\mathbf{R}(\theta, v) \equiv \begin{cases} x = 4 \cos \theta + v \cos \frac{\theta}{2} \\ y = 4 \sin \theta + v \sin \frac{\theta}{2} \\ z = v \sin \frac{\theta}{2} \end{cases}, \theta \in [0, 2\pi], v \in [-1, 1].$$

An obvious question is why the normal vector, $\mathbf{R}_{,\theta} \times \mathbf{R}_{,v} / |\mathbf{R}_{,\theta} \times \mathbf{R}_{,v}|$ is not a continuous function of position on S . You can see easily that it is a continuous function of both θ and v . However, the map, $\mathbf{R}$ is not one to one. In fact, $\mathbf{R}(0, 0) = \mathbf{R}(2\pi, 0)$. Therefore, near this point on S , there are two different values for the above normal vector. In fact, a tedious computation will show that this normal vector is

$$\frac{(4 \sin \frac{1}{2}\theta \cos \theta - \frac{1}{2}v, 4 \sin \frac{1}{2}\theta \sin \theta + \frac{1}{2}v, -8 \cos^2 \frac{1}{2}\theta \sin \frac{1}{2}\theta - 8 \cos^3 \frac{1}{2}\theta + 4 \cos \frac{1}{2}\theta)}{D}$$

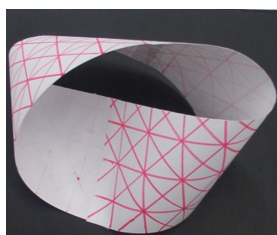
where

$$D = \left(16 \sin^2 \left(\frac{\theta}{2} \right) + \frac{v^2}{2} + 4 \sin \left(\frac{\theta}{2} \right) v (\sin \theta - \cos \theta) \right. \\ \left. + 4^3 \cos^2 \left(\frac{\theta}{2} \right) \left(\cos \left(\frac{1}{2} \theta \right) \sin \left(\frac{1}{2} \theta \right) + \cos^2 \left(\frac{1}{2} \theta \right) - \frac{1}{2} \right)^2 \right)$$

and you can verify that the denominator will not vanish. Letting $v = 0$ and $\theta = 0$ and 2π yields the two vectors, $(0, 0, -1)$, $(0, 0, 1)$ so there is a discontinuity. This is why I was careful to say in the statement of Stoke's theorem given above that $\mathbf{R}$ is one to one.

The Mobeus band has some usefulness. In old machine shops the equipment was run by a belt which was given a twist to spread the surface wear on the belt over twice the area.

The above explanation shows that $\mathbf{R}_\theta \times \mathbf{R}_v / |\mathbf{R}_\theta \times \mathbf{R}_v|$ fails to deliver an orientation for the Mobeus band. However, this does not answer the question whether there is some orientation for it other than this one. In fact there is none. You can see this by looking at the first of the two pictures below or by making one and tracing it with a pencil. There is only one side to the Mobeus band. An oriented surface must have two sides, one side identified by the given unit normal which varies continuously over the surface and the other side identified by the negative of this normal. The second picture below was taken by Ouyang when he was at meetings in Paris and saw it at a museum.



28.2.3 Conservative Vector Fields

Definition 28.2.5 A vector field, $\mathbf{F}$ defined in a three dimensional region is said to be **conservative**² if for every piecewise smooth closed curve, C , it follows $\int_C \mathbf{F} \cdot d\mathbf{R} = 0$.

Definition 28.2.6 Let $(\mathbf{x}, \mathbf{p}_1, \dots, \mathbf{p}_n, \mathbf{y})$ be an ordered list of points in $\mathbb{R}^p$. Let

$$\mathbf{p}(\mathbf{x}, \mathbf{p}_1, \dots, \mathbf{p}_n, \mathbf{y})$$

denote the piecewise smooth curve consisting of a straight line segment from $\mathbf{x}$ to $\mathbf{p}_1$ and then the straight line segment from $\mathbf{p}_1$ to $\mathbf{p}_2 \dots$ and finally the straight line segment from $\mathbf{p}_n$ to $\mathbf{y}$. This is called a **polygonal curve**. An open set in $\mathbb{R}^p$, U , is said to be a **region** if it has the property that for any two points, $\mathbf{x}, \mathbf{y} \in U$, there exists a polygonal curve joining the two points.

Conservative vector fields are important because of the following theorem, sometimes called the fundamental theorem for line integrals.

Theorem 28.2.7 Let U be a region in $\mathbb{R}^p$ and let $\mathbf{F} : U \rightarrow \mathbb{R}^p$ be a continuous vector field. Then $\mathbf{F}$ is conservative if and only if there exists a scalar valued function

²There is no such thing as a liberal vector field.

of p variables, ϕ such that $\mathbf{F} = \nabla\phi$. Furthermore, if C is an oriented curve which goes from $\mathbf{x}$ to $\mathbf{y}$ in U , then

$$\int_C \mathbf{F} \cdot d\mathbf{R} = \phi(\mathbf{y}) - \phi(\mathbf{x}). \quad (28.4)$$

Thus the line integral is path independent in this case. This function, ϕ is called a **scalar potential** for $\mathbf{F}$.

Proof: To save space and fussing over things which are unimportant, denote by $\mathbf{p}(\mathbf{x}_0, \mathbf{x})$ a polygonal curve from $\mathbf{x}_0$ to $\mathbf{x}$. Thus the orientation is such that it goes from $\mathbf{x}_0$ to $\mathbf{x}$. The curve $\mathbf{p}(\mathbf{x}, \mathbf{x}_0)$ denotes the same set of points but in the opposite order. Suppose first $\mathbf{F}$ is conservative. Fix $\mathbf{x}_0 \in U$ and let

$$\phi(\mathbf{x}) \equiv \int_{\mathbf{p}(\mathbf{x}_0, \mathbf{x})} \mathbf{F} \cdot d\mathbf{R}.$$

This is well defined because if $\mathbf{q}(\mathbf{x}_0, \mathbf{x})$ is another polygonal curve joining $\mathbf{x}_0$ to $\mathbf{x}$, Then the curve obtained by following $\mathbf{p}(\mathbf{x}_0, \mathbf{x})$ from $\mathbf{x}_0$ to $\mathbf{x}$ and then from $\mathbf{x}$ to $\mathbf{x}_0$ along $\mathbf{q}(\mathbf{x}, \mathbf{x}_0)$ is a closed piecewise smooth curve and so by assumption, the line integral along this closed curve equals 0. However, this integral is just

$$\int_{\mathbf{p}(\mathbf{x}_0, \mathbf{x})} \mathbf{F} \cdot d\mathbf{R} + \int_{\mathbf{q}(\mathbf{x}, \mathbf{x}_0)} \mathbf{F} \cdot d\mathbf{R} = \int_{\mathbf{p}(\mathbf{x}_0, \mathbf{x})} \mathbf{F} \cdot d\mathbf{R} - \int_{\mathbf{q}(\mathbf{x}_0, \mathbf{x})} \mathbf{F} \cdot d\mathbf{R}$$

which shows

$$\int_{\mathbf{p}(\mathbf{x}_0, \mathbf{x})} \mathbf{F} \cdot d\mathbf{R} = \int_{\mathbf{q}(\mathbf{x}_0, \mathbf{x})} \mathbf{F} \cdot d\mathbf{R}$$

and that ϕ is well defined. For small t ,

$$\begin{aligned} \frac{\phi(\mathbf{x} + t\mathbf{e}_i) - \phi(\mathbf{x})}{t} &= \frac{\int_{\mathbf{p}(\mathbf{x}_0, \mathbf{x} + t\mathbf{e}_i)} \mathbf{F} \cdot d\mathbf{R} - \int_{\mathbf{p}(\mathbf{x}_0, \mathbf{x})} \mathbf{F} \cdot d\mathbf{R}}{t} \\ &= \frac{\int_{\mathbf{p}(\mathbf{x}_0, \mathbf{x})} \mathbf{F} \cdot d\mathbf{R} + \int_{\mathbf{p}(\mathbf{x}, \mathbf{x} + t\mathbf{e}_i)} \mathbf{F} \cdot d\mathbf{R} - \int_{\mathbf{p}(\mathbf{x}_0, \mathbf{x})} \mathbf{F} \cdot d\mathbf{R}}{t}. \end{aligned}$$

Since U is open, for small t , the ball of radius $|t|$ centered at $\mathbf{x}$ is contained in U . Therefore, the line segment from $\mathbf{x}$ to $\mathbf{x} + t\mathbf{e}_i$ is also contained in U and so one can take $\mathbf{p}(\mathbf{x}, \mathbf{x} + t\mathbf{e}_i)(s) = \mathbf{x} + s(t\mathbf{e}_i)$ for $s \in [0, 1]$. Therefore, the above difference quotient reduces to

$$\begin{aligned} \frac{1}{t} \int_0^1 \mathbf{F}(\mathbf{x} + s(t\mathbf{e}_i)) \cdot t\mathbf{e}_i ds &= \int_0^1 F_i(\mathbf{x} + s(t\mathbf{e}_i)) ds \\ &= F_i(\mathbf{x} + s_t(t\mathbf{e}_i)) \end{aligned}$$

by the mean value theorem for integrals. Here s_t is some number between 0 and 1. By continuity of $\mathbf{F}$, this converges to $F_i(\mathbf{x})$ as $t \rightarrow 0$. Therefore, $\nabla\phi = \mathbf{F}$ as claimed.

Conversely, if $\nabla\phi = \mathbf{F}$, then if $\mathbf{R} : [a, b] \rightarrow \mathbb{R}^p$ is any C^1 curve joining $\mathbf{x}$ to $\mathbf{y}$,

$$\begin{aligned} \int_a^b \mathbf{F}(\mathbf{R}(t)) \cdot \mathbf{R}'(t) dt &= \int_a^b \nabla\phi(\mathbf{R}(t)) \cdot \mathbf{R}'(t) dt \\ &= \int_a^b \frac{d}{dt} (\phi(\mathbf{R}(t))) dt \\ &= \phi(\mathbf{R}(b)) - \phi(\mathbf{R}(a)) \\ &= \phi(\mathbf{y}) - \phi(\mathbf{x}) \end{aligned}$$

and this verifies 28.4 in the case where the curve joining the two points is smooth. The general case follows immediately from this by using this result on each of the pieces of the piecewise smooth curve. For example if the curve goes from $\mathbf{x}$ to $\mathbf{p}$ and then from $\mathbf{p}$ to $\mathbf{y}$, the above would imply the integral over the curve from $\mathbf{x}$ to $\mathbf{p}$ is $\phi(\mathbf{p}) - \phi(\mathbf{x})$ while from $\mathbf{p}$ to $\mathbf{y}$ the integral would yield $\phi(\mathbf{y}) - \phi(\mathbf{p})$. Adding these gives $\phi(\mathbf{y}) - \phi(\mathbf{x})$. The formula 28.4 implies the line integral over any closed curve equals zero because the starting and ending points of such a curve are the same. This proves the theorem.

Example 28.2.8 Let $\mathbf{F}(x, y, z) = (\cos x - yz \sin(xz), \cos(xz), -yx \sin(xz))$. Let C be a piecewise smooth curve which goes from $(\pi, 1, 1)$ to $(\frac{\pi}{2}, 3, 2)$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$.

The specifics of the curve are not given so the problem is nonsense unless the vector field is conservative. Therefore, it is reasonable to look for the function, ϕ satisfying $\nabla\phi = \mathbf{F}$. Such a function satisfies

$$\phi_x = \cos x - y(\sin xz)z$$

and so, assuming ϕ exists,

$$\phi(x, y, z) = \sin x + y \cos(xz) + \psi(y, z).$$

I have to add in the most general thing possible, $\psi(y, z)$ to ensure possible solutions are not being thrown out. It wouldn't be good at this point to add in a constant since the answer could involve a function of either or both of the other variables. Now from what was just obtained,

$$\phi_y = \cos(xz) + \psi_y = \cos xz$$

and so it is possible to take $\psi_y = 0$. Consequently, ϕ , if it exists is of the form

$$\phi(x, y, z) = \sin x + y \cos(xz) + \psi(z).$$

Now differentiating this with respect to z gives

$$\phi_z = -yx \sin(xz) + \psi_z = -yx \sin(xz)$$

and this shows ψ does not depend on z either. Therefore, it suffices to take $\psi = 0$ and

$$\phi(x, y, z) = \sin(x) + y \cos(xz).$$

Therefore, the desired line integral equals

$$\sin\left(\frac{\pi}{2}\right) + 3 \cos(\pi) - (\sin(\pi) + \cos(\pi)) = -1.$$

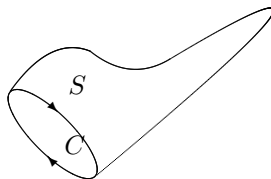
The above process for finding ϕ will not lead you astray in the case where there does not exist a scalar potential. As an example, consider the following.

Example 28.2.9 Let $\mathbf{F}(x, y, z) = (x, y^2x, z)$. Find a scalar potential for $\mathbf{F}$ if it exists.

If ϕ exists, then $\phi_x = x$ and so $\phi = \frac{x^2}{2} + \psi(y, z)$. Then $\phi_y = \psi_y(y, z) = xy^2$ but this is impossible because the left side depends only on y and z while the right side depends also on x . Therefore, this vector field is not conservative and there does not exist a scalar potential.

Definition 28.2.10 A set of points in three dimensional space, V is simply connected if every piecewise smooth closed curve, C is the edge of a surface, S which is

contained entirely within V in such a way that Stokes theorem holds for the surface, S and its edge, C .



This is like a sock. The surface is the sock and the curve, C goes around the opening of the sock.

As an application of Stoke's theorem, here is a useful theorem which gives a way to check whether a vector field is conservative.

Theorem 28.2.11 For a three dimensional simply connected open set, V and $\mathbf{F}$ a C^1 vector field defined in V , $\mathbf{F}$ is conservative if $\nabla \times \mathbf{F} = \mathbf{0}$ in V .

Proof: If $\nabla \times \mathbf{F} = \mathbf{0}$ then taking an arbitrary closed curve, C , and letting S be a surface bounded by C which is contained in V , Stoke's theorem implies

$$0 = \int_S \nabla \times \mathbf{F} \cdot \mathbf{n} dA = \int_C \mathbf{F} \cdot d\mathbf{R}.$$

Thus $\mathbf{F}$ is conservative.

Example 28.2.12 Determine whether the vector field,

$$(4x^3 + 2(\cos(x^2 + z^2))x, 1, 2(\cos(x^2 + z^2))z)$$

is conservative.

Since this vector field is defined on all of $\mathbb{R}^3$, it only remains to take its curl and see if it is the zero vector.

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \partial_x & \partial_y & \partial_z \\ 4x^3 + 2(\cos(x^2 + z^2))x & 1 & 2(\cos(x^2 + z^2))z \end{vmatrix}.$$

This is obviously equal to zero. Therefore, the given vector field is conservative. Can you find a potential function for it? Let ϕ be the potential function. Then $\phi_z = 2(\cos(x^2 + z^2))z$ and so $\phi(x, y, z) = \sin(x^2 + z^2) + g(x, y)$. Now taking the derivative of ϕ with respect to y , you see $g_y = 1$ so $g(x, y) = y + h(x)$. Hence $\phi(x, y, z) = y + g(x) + \sin(x^2 + z^2)$. Taking the derivative with respect to x , you get $4x^3 + 2(\cos(x^2 + z^2))x = g'(x) + 2x \cos(x^2 + z^2)$ and so it suffices to take $g(x) = x^4$. Hence $\phi(x, y, z) = y + x^4 + \sin(x^2 + z^2)$.

28.2.4 Some Terminology

If $\mathbf{F} = (P, Q, R)$ is a vector field. Then the statement that $\mathbf{F}$ is conservative is the same as saying the differential form $Pdx + Qdy + Rdz$ is exact. Some people like to say things in terms of vector fields and some say it in terms of differential forms. In Example 28.2.12, the differential form $(4x^3 + 2(\cos(x^2 + z^2))x)dx + dy + (2(\cos(x^2 + z^2))z)dz$ is exact.

28.3 Maxwell's Equations And The Wave Equation

Many of the ideas presented above are useful in analyzing Maxwell's equations. These equations are derived in advanced physics courses. They are

$$\nabla \times \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} = \mathbf{0} \quad (28.5)$$

$$\nabla \cdot \mathbf{E} = 4\pi\rho \quad (28.6)$$

$$\nabla \times \mathbf{B} - \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} = \frac{4\pi}{c} \mathbf{f} \quad (28.7)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (28.8)$$

and it is assumed these hold on all of $\mathbb{R}^3$ to eliminate technical considerations having to do with whether something is simply connected.

In these equations, $\mathbf{E}$ is the electrostatic field and $\mathbf{B}$ is the magnetic field while ρ and $\mathbf{f}$ are sources. By 28.8 $\mathbf{B}$ has a vector potential, $\mathbf{A}_1$ such that $\mathbf{B} = \nabla \times \mathbf{A}_1$. Now go to 28.5 and write

$$\nabla \times \mathbf{E} + \frac{1}{c} \nabla \times \frac{\partial \mathbf{A}_1}{\partial t} = \mathbf{0}$$

showing that

$$\nabla \times \left(\mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}_1}{\partial t} \right) = \mathbf{0}$$

It follows $\mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}_1}{\partial t}$ has a scalar potential, ψ_1 satisfying

$$\nabla \psi_1 = \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}_1}{\partial t}. \quad (28.9)$$

Now suppose ϕ is a time dependent scalar field satisfying

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = \frac{1}{c} \frac{\partial \psi_1}{\partial t} - \nabla \cdot \mathbf{A}_1. \quad (28.10)$$

Next define

$$\mathbf{A} \equiv \mathbf{A}_1 + \nabla \phi, \quad \psi \equiv \psi_1 + \frac{1}{c} \frac{\partial \phi}{\partial t}. \quad (28.11)$$

Therefore, in terms of the new variables, 28.10 becomes

$$\nabla^2 \phi - \frac{1}{c^2} \frac{\partial^2 \phi}{\partial t^2} = \frac{1}{c} \left(\frac{\partial \psi}{\partial t} - \frac{1}{c} \frac{\partial^2 \phi}{\partial t^2} \right) - \nabla \cdot \mathbf{A} + \nabla^2 \phi$$

which yields

$$0 = \frac{\partial \psi}{\partial t} - c \nabla \cdot \mathbf{A}. \quad (28.12)$$

Then it follows from Theorem 27.1.3 on Page 588 that $\mathbf{A}$ is also a vector potential for $\mathbf{B}$. That is

$$\nabla \times \mathbf{A} = \mathbf{B}. \quad (28.13)$$

From 28.9

$$\nabla \left(\psi - \frac{1}{c} \frac{\partial \phi}{\partial t} \right) = \mathbf{E} + \frac{1}{c} \left(\frac{\partial \mathbf{A}}{\partial t} - \nabla \frac{\partial \phi}{\partial t} \right)$$

and so

$$\nabla \psi = \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t}. \quad (28.14)$$

Using 28.7 and 28.14,

$$\nabla \times (\nabla \times \mathbf{A}) - \frac{1}{c} \frac{\partial}{\partial t} \left(\nabla \psi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \right) = \frac{4\pi}{c} \mathbf{f}. \quad (28.15)$$

Now from Theorem 27.1.3 on Page 588 this implies

$$\nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} - \nabla \left(\frac{1}{c} \frac{\partial \psi}{\partial t} \right) + \frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} = \frac{4\pi}{c} \mathbf{f}$$

and using 28.12, this gives

$$\frac{1}{c^2} \frac{\partial^2 \mathbf{A}}{\partial t^2} - \nabla^2 \mathbf{A} = \frac{4\pi}{c} \mathbf{f}. \quad (28.16)$$

Also from 28.14, 28.6, and 28.12,

$$\begin{aligned} \nabla^2 \psi &= \nabla \cdot \mathbf{E} + \frac{1}{c} \frac{\partial}{\partial t} (\nabla \cdot \mathbf{A}) \\ &= 4\pi\rho + \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} \end{aligned}$$

and so

$$\frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} - \nabla^2 \psi = -4\pi\rho. \quad (28.17)$$

This is very interesting. If a solution to the wave equations, 28.17, and 28.16 can be found along with a solution to 28.12, then letting the magnetic field be given by 28.13 and letting $\mathbf{E}$ be given by 28.14 the result is a solution to Maxwell's equations. This is significant because wave equations are easier to think of than Maxwell's equations. Note the above argument also showed that it is always possible, by solving another wave equation, to get 28.12 to hold.

28.4 Exercises

1. Determine whether the vector field,

$$(2xy^3 \sin z^4, 3x^2y^2 \sin z^4 + 1, 4x^2y^3 (\cos z^4) z^3 + 1)$$

is conservative. If it is conservative, find a potential function.

2. Determine whether the vector field,

$$(2xy^3 \sin z + y^2 + z, 3x^2y^2 \sin z + 2xy, x^2y^3 \cos z + x)$$

is conservative. If it is conservative, find a potential function.

3. Determine whether the vector field, $(2xy^3 \sin z + z, 3x^2y^2 \sin z + 2xy, x^2y^3 \cos z + x)$ is conservative. If it is conservative, find a potential function.
4. Find scalar potentials for the following vector fields if it is possible to do so. If it is not possible to do so, explain why.

- (a) $(y^2, 2xy + \sin z, 2z + y \cos z)$

- (b) $(2z (\cos (x^2 + y^2)) x, 2z (\cos (x^2 + y^2)) y, \sin (x^2 + y^2) + 2z)$

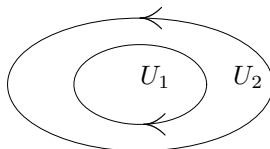
- (c) $(f(x), g(y), h(z))$

- (d) (xy, z^2, y^3)

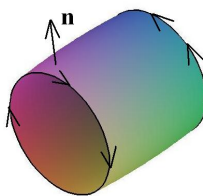
- (e) $\left(z + 2\frac{x}{x^2+y^2+1}, 2\frac{y}{x^2+y^2+1}, x + 3z^2\right)$
5. If a vector field is not conservative on the set U , is it possible the same vector field could be conservative on some subset of U ? Explain and give examples if it is possible. If it is not possible also explain why.
 6. Prove that if a vector field, $\mathbf{F}$ has a scalar potential, then it has infinitely many scalar potentials.
 7. Here is a vector field: $\mathbf{F} \equiv (2xy, x^2 - 5y^4, 3z^2)$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$ where C is a curve which goes from $(1, 2, 3)$ to $(4, -2, 1)$.
 8. Here is a vector field: $\mathbf{F} \equiv (2xy, x^2 - 5y^4, 3(\cos z^3)z^2)$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$ where C is a curve which goes from $(1, 0, 1)$ to $(-4, -2, 1)$.
 9. Find $\int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$ where U is the set, $\{(x, y) : 2 \leq x \leq 4, 0 \leq y \leq x\}$ and $\mathbf{F}(x, y) = (x \sin y, y \sin x)$.
 10. Find $\int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$ where U is the set, $\{(x, y) : 2 \leq x \leq 3, 0 \leq y \leq x^2\}$ and $\mathbf{F}(x, y) = (x \cos y, y + x)$.
 11. Find $\int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$ where U is the set, $\{(x, y) : 1 \leq x \leq 2, x \leq y \leq 3\}$ and $\mathbf{F}(x, y) = (x \sin y, y \sin x)$.
 12. Find $\int_{\partial U} \mathbf{F} \cdot d\mathbf{R}$ where U is the set, $\{(x, y) : x^2 + y^2 \leq 2\}$ and $\mathbf{F}(x, y) = (-y^3, x^3)$.
 13. Show that for many open sets in $\mathbb{R}^2$, Area of $U = \int_{\partial U} x dy$, and Area of $U = \int_{\partial U} -y dx$ and Area of $U = \frac{1}{2} \int_{\partial U} -y dx + x dy$. **Hint:** Use Green's theorem.
 14. Two smooth oriented surfaces, S_1 and S_2 intersect in a piecewise smooth oriented closed curve, C . Let $\mathbf{F}$ be a C^1 vector field defined on $\mathbb{R}^3$. Explain why $\int_{S_1} \text{curl}(\mathbf{F}) \cdot \mathbf{n} dS = \int_{S_2} \text{curl}(\mathbf{F}) \cdot \mathbf{n} dS$. Here $\mathbf{n}$ is the normal to the surface which corresponds to the given orientation of the curve, C .
 15. Show that $\text{curl}(\psi \nabla \phi) = \nabla \psi \times \nabla \phi$ and explain why $\int_S \nabla \psi \times \nabla \phi \cdot \mathbf{n} dS = \int_{\partial S} (\psi \nabla \phi) \cdot d\mathbf{r}$.
 16. Find a simple formula for $\text{div}(\nabla(u^\alpha))$ where $\alpha \in \mathbb{R}$.
 17. Parametric equations for one arch of a cycloid are given by $x = a(t - \sin t)$ and $y = a(1 - \cos t)$ where here $t \in [0, 2\pi]$. Sketch a rough graph of this arch of a cycloid and then find the area between this arch and the x axis. **Hint:** This is very easy using Green's theorem and the vector field, $\mathbf{F} = (-y, x)$.
 18. Let $\mathbf{r}(t) = (\cos^3(t), \sin^3(t))$ where $t \in [0, 2\pi]$. Sketch this curve and find the area enclosed by it using Green's theorem.
 19. Verify that Green's theorem can be considered a special case of Stoke's theorem.
 20. Consider the vector field, $\left(\frac{-y}{(x^2+y^2)}, \frac{x}{(x^2+y^2)}, 0\right) = \mathbf{F}$. Show that $\nabla \times \mathbf{F} = \mathbf{0}$ but that for the closed curve, whose parametrization is $\mathbf{R}(t) = (\cos t, \sin t, 0)$ for $t \in [0, 2\pi]$, $\int_C \mathbf{F} \cdot d\mathbf{R} \neq 0$. Therefore, the vector field is not conservative. Does this contradict Theorem 28.2.11? Explain.

21. Let $\mathbf{x}$ be a point of $\mathbb{R}^3$ and let $\mathbf{n}$ be a unit vector. Let D_r be the circular disk of radius r containing $\mathbf{x}$ which is perpendicular to $\mathbf{n}$. Placing the tail of $\mathbf{n}$ at $\mathbf{x}$ and viewing D_r from the point of $\mathbf{n}$, orient ∂D_r in the counter clockwise direction. Now suppose $\mathbf{F}$ is a vector field defined near $\mathbf{x}$. Show that $\text{curl}(\mathbf{F}) \cdot \mathbf{n} = \lim_{r \rightarrow 0} \frac{1}{\pi r^2} \int_{\partial D_r} \mathbf{F} \cdot d\mathbf{R}$. This last integral is sometimes called the circulation density of $\mathbf{F}$. Explain how this shows that $\text{curl}(\mathbf{F}) \cdot \mathbf{n}$ measures the tendency for the vector field to “curl” around the point, the vector $\mathbf{n}$ at the point $\mathbf{x}$.
22. The cylinder $x^2 + y^2 = 4$ is intersected with the plane $x + y + z = 2$. This yields a closed curve, C . Orient this curve in the counter clockwise direction when viewed from a point high on the z axis. Let $\mathbf{F} = (x^2y, z + y, x^2)$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$.
23. The cylinder $x^2 + 4y^2 = 4$ is intersected with the plane $x + 3y + 2z = 1$. This yields a closed curve, C . Orient this curve in the counter clockwise direction when viewed from a point high on the z axis. Let $\mathbf{F} = (y, z + y, x^2)$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$.
24. The cylinder $x^2 + y^2 = 4$ is intersected with the plane $x + 3y + 2z = 1$. This yields a closed curve, C . Orient this curve in the clockwise direction when viewed from a point high on the z axis. Let $\mathbf{F} = (y, z + y, x)$. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$.
25. Let $\mathbf{F} = (xz, z^2(y + \sin x), z^3y)$. Find the surface integral, $\int_S \text{curl}(\mathbf{F}) \cdot \mathbf{n} dA$ where S is the surface, $z = 4 - (x^2 + y^2)$, $z \geq 0$.
26. Let $\mathbf{F} = (xz, (y^3 + x), z^3y)$. Find the surface integral, $\int_S \text{curl}(\mathbf{F}) \cdot \mathbf{n} dA$ where S is the surface, $z = 16 - (x^2 + y^2)$, $z \geq 0$.
27. The cylinder $z = y^2$ intersects the surface $z = 8 - x^2 - 4y^2$ in a curve, C which is oriented in the counter clockwise direction when viewed high on the z axis. Find $\int_C \mathbf{F} \cdot d\mathbf{R}$ if $\mathbf{F} = \left(\frac{z^2}{2}, xy, xz\right)$.
28. Suppose solutions have been found to 28.17, 28.16, and 28.12. Then define $\mathbf{E}$ and $\mathbf{B}$ using 28.14 and 28.13. Verify Maxwell's equations hold for $\mathbf{E}$ and $\mathbf{B}$.
29. Suppose now you have found solutions to 28.17 and 28.16, ψ_1 and A_1 . Then go show again that if ϕ satisfies 28.10 and $\psi \equiv \psi_1 + \frac{1}{c} \frac{\partial \phi}{\partial t}$, while $\mathbf{A} \equiv \mathbf{A}_1 + \nabla \phi$, then 28.12 holds for $\mathbf{A}$ and ψ .
30. Why consider Maxwell's equations? Why not just consider 28.17, 28.16, and 28.12?
31. Tell which open sets are simply connected.
 - (a) The inside of a car radiator.
 - (b) A donut.
 - (c) The solid part of a cannon ball which contains a void on the interior.
 - (d) The inside of a donut which has had a large bite taken out of it.
 - (e) All of $\mathbb{R}^3$ except the z axis.
 - (f) All of $\mathbb{R}^3$ except the xy plane.
32. Let P be a polygon with vertices $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), (x_1, y_1)$ encountered as you move over the boundary of the polygon which is assumed a simple closed curve in the counter clockwise direction. Using Problem 13, find a nice formula for the area of the polygon in terms of the vertices.

33. Here is a picture of two regions in the plane, U_1 and U_2 . Suppose Green's theorem holds for each of these regions. Explain why Green's theorem must also hold for the region which lies between them if the boundary is oriented as shown in the picture.



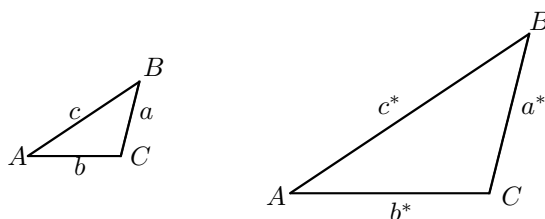
34. Here is a picture of a surface which has two bounding curves oriented as shown. Explain why Stoke's theorem will hold for such a surface and sketch a region in the plane which could serve as a parameter domain for this surface.



Basic Plane Geometry

A.1 Similar Triangles And Parallel Lines

Definition A.1.1 *Two triangles are similar if they have the same angles. For example, in the following picture, the two triangles are similar because the angles are the same.*



The fundamental axiom for similar triangles is the following.

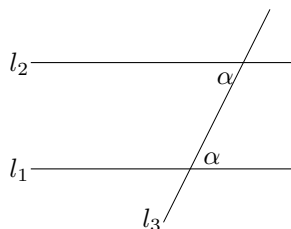
Axiom A.1.2 *If two triangles are similar then the ratios of corresponding parts are the same.*

For example in the above picture, this says that

$$\frac{a}{b} = \frac{a^*}{b^*}$$

Definition A.1.3 *Two lines in the plane are said to be parallel if no matter how far they are extended, they never intersect.*

Definition A.1.4 *If two lines l_1 and l_2 are parallel and if they are intersected by a line, l_3 , the alternate interior angles are shown in the following picture labeled as α .*



As suggested by the above picture, the following axiom will be used.

Axiom A.1.5 *If l_1 and l_2 are parallel lines intersected by l_3 , then alternate interior angles are equal.*

Definition A.1.6 An angle is a right angle if when either side is extended, the new angle formed by the extension equals the original angle.

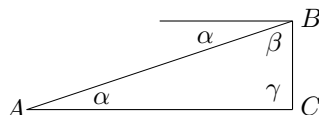
Axiom A.1.7 Suppose l_1 and l_2 both intersect a third line, l_3 in a right angle. Then l_1 and l_2 are parallel.

Definition A.1.8 A right triangle is one in which one of the angles is a right angle.

Axiom A.1.9 Given a straight line and a point, there exists a straight line which contains the point and intersects the given line in two right angles. This line is called perpendicular to the given line.

Theorem A.1.10 Let α, β , and γ be the angles of a right triangle with γ the right angle. Then if the angles, α and β are placed next to each other, the resulting angle is a right angle.

Proof: Consider the following picture.



In the picture the top horizontal line is obtained from Axiom A.1.9. It is a line perpendicular to the line determined by the line segment joining B and C which passes through the point, B . Thus from Axiom A.1.7 this line is parallel to the line joining A and B and by Axiom A.1.5 the angle between the line joining A and B and this new line is α as shown in the picture. Therefore, the angle formed by placing α and β together is a right angle as claimed.

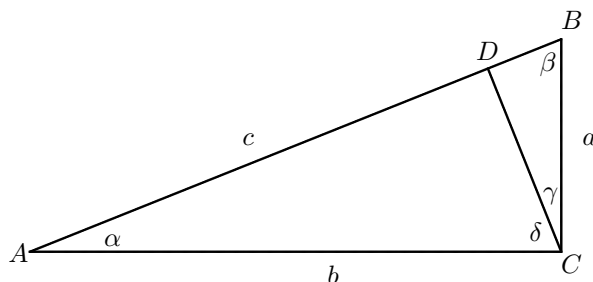
Definition A.1.11 When an angle α is placed next to an angle β as shown above, then the resulting angle is denoted by $\alpha + \beta$. A right angle is said to have 90° or to be a 90° angle.

With this definition, Theorem A.1.10 says the sum of the two non 90° angles in a right triangle is 90° .

In a right triangle the long side is called the hypotenuse. The similar triangles axiom can be used to prove the Pythagorean theorem.

Theorem A.1.12 (Pythagoras) In a right triangle the square of the length of the hypotenuse equals the sum of the squares of the lengths of the other two sides.

Proof: Consider the following picture in which the large triangle is a right triangle and D is the point where the line through C perpendicular to the line from A to B intersects the line from A to B . Then c is defined to be the length of the line from A to B , a is the length of the line from B to C , and b is the length of the line from A to C . Denote by $\overline{DB}$ the length of the line from D to B .



Then from Theorem A.1.10, $\delta + \gamma = 90^\circ$ and $\beta + \gamma = 90^\circ$. Therefore, $\delta = \beta$. Also from this same theorem, $\alpha + \delta = 90^\circ$ and so $\alpha = \gamma$. Therefore, the three triangles shown in the picture are all similar. By Axiom A.1.2,

$$\frac{c}{a} = \frac{a}{\overline{DB}}, \text{ and } \frac{c}{b} = \frac{b}{c - \overline{DB}}.$$

Therefore, $c\overline{DB} = a^2$ and

$$c(c - \overline{DB}) = b^2$$

so

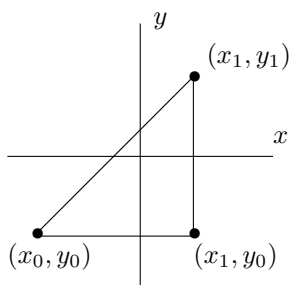
$$\begin{aligned} c^2 &= c\overline{DB} + b^2 \\ &= a^2 + b^2. \end{aligned}$$

This proves the Pythagorean theorem.¹

This theorem implies there should exist some such number which deserves to be called $\sqrt{a^2 + b^2}$.

A.2 Distance Formula And Trigonometric Functions

As just explained, points in the plane may be identified by giving a pair of numbers. Suppose there are two points in the plane and it is desired to find the distance between them. There are actually many ways used to measure this distance but the best way, and the only way used in this book is determined by the Pythagorean theorem. Consider the following picture.



In this picture, the distance between the points denoted by (x_0, y_0) and (x_1, y_1) should be the square root of the sum of the squares of the lengths of the two sides. The length of the side on the bottom is $|x_0 - x_1|$ while the length of the side on the right is $|y_0 - y_1|$. Therefore, by the Pythagorean theorem the distance between the two indicated points is $\sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2}$. Note you could write

$$\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}$$

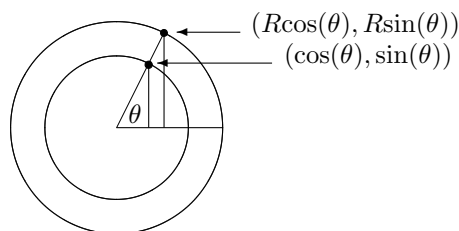
or even

$$\sqrt{(x_0 - x_1)^2 + (y_1 - y_0)^2}$$

¹This theorem is due to Pythagoras who lived about 572-497 B.C. This was during the Babylonian captivity of the Jews. Thus Pythagoras was probably a contemporary of the prophet Daniel, sometime before Ezra and Nehemiah. Alexander the great would not come along for more than 100 years. There was, however, an even earlier Greek mathematician named Thales, 624-547 B.C. who also did fundamental work in geometry. Greek geometry was organized and published by Euclid about 300 B.C.

and it would make no difference in the resulting number. The distance between the two points is written as $|(x_0, y_0) - (x_1, y_1)|$ or sometimes when P_0 is the point determined by (x_0, y_0) and P_1 is the point determined by (x_1, y_1) , as $d(P_0, P_1)$ or $|P_0P_1|$.

The trigonometric functions $\cos$ and $\sin$ are defined next. Consider the following picture in which the small circle has radius 1, the large circle has radius R , and the right side of each of the two triangles is perpendicular to the bottom side which lies on the x axis.



By Theorem A.1.10 on Page 636 the two triangles have the same angles and so they are similar. Now define by $(\cos \theta, \sin \theta)$ the coordinates of the top vertex of the smaller triangle. Therefore, it follows the coordinates of the top vertex of the larger triangle are as shown. This shows the following definition is well defined.

Definition A.2.1 For θ an angle, define $\cos \theta$ and $\sin \theta$ as follows. Place the vertex of the angle (The vertex is the point.) at the point whose coordinates are $(0, 0)$ in such a way that one side of the angle lies on the positive x axis and the other side extends upward. Extend this other side until it intersects a circle of radius R . Then the point of intersection, is given as $(R \cos \theta, R \sin \theta)$. In particular, this specifies $\cos \theta$ and $\sin \theta$ by simply letting $R = 1$.

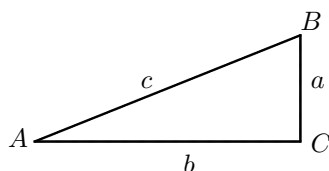
Proposition A.2.2 For any angle, θ , $\cos^2 \theta + \sin^2 \theta = 1$.

Proof: This follows immediately from the above definition and the distance formula. Since $(\cos \theta, \sin \theta)$ is a point on the circle which has radius 1, the distance of this point to $(0, 0)$ equals 1. Thus the above identity holds.

The other trigonometric functions are defined as follows.

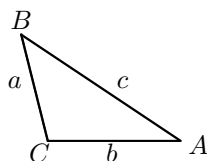
$$\tan \theta \equiv \frac{\sin \theta}{\cos \theta}, \cot \theta \equiv \frac{\cos \theta}{\sin \theta}, \sec \theta \equiv \frac{1}{\cos \theta}, \csc \theta \equiv \frac{1}{\sin \theta}. \quad (1.1)$$

It is also important to understand these functions in terms of a right triangle. Consider the following picture of a right triangle.



You should verify $\sin A \equiv a/c$, $\cos A \equiv b/c$, $\tan A \equiv a/b$, $\sec A \equiv c/b$, and $\csc A \equiv c/a$.

Having defined the $\cos$ and $\sin$ there is a very important generalization of the Pythagorean theorem known as the law of cosines. Consider the following picture of a triangle in which a, b and c are the lengths of the sides and A, B , and C denote the angles indicated.



The law of cosines is the following.

Theorem A.2.3 *Let ABC be a triangle as shown above. Then*

$$c^2 = a^2 + b^2 - 2ab \cos C$$

Proof: Situate the triangle so the vertex of the angle, C , is on the point whose coordinates are $(0, 0)$ and so the side opposite the vertex, B is on the positive x axis as shown in the above picture. Then from the definition of the $\cos C$, the coordinates of the vertex, B are $(a \cos C, a \sin C)$ while it is clear the coordinates of A are $(b, 0)$. Therefore, from the distance formula, and Proposition A.2.2,

$$\begin{aligned} c^2 &= (a \cos C - b)^2 + a^2 \sin^2 C \\ &= a^2 \cos^2 C - 2ab \cos C + b^2 + a^2 \sin^2 C \\ &= a^2 + b^2 - 2ab \cos C \end{aligned}$$

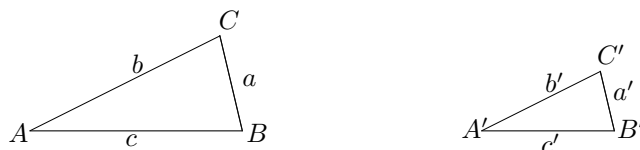
as claimed.

Corollary A.2.4 *Let ABC be any triangle as shown above. Then the length of any side is no longer than the sum of the lengths of the other two sides.*

Proof: This follows immediately from the law of cosines. From Proposition A.2.2, $|\cos \theta| \leq 1$ and so $c^2 = a^2 + b^2 - 2ab \cos C \leq a^2 + b^2 + 2ab = (a + b)^2$. This proves the corollary.

Corollary A.2.5 *Suppose T and T' are two triangles such that one angle is the same in the two triangles and in each triangle, the sides forming that angle are equal. Then the corresponding sides are proportional.*

Proof: Let $T = ABC$ with the two equal sides being AC and AB . Let T' be labeled in the same way but with primes on the letters. Thus the angle at A is equal to the angle at A' . The following picture is descriptive of the situation.



Denote by a, a', b, b', c and c' the sides indicated in the picture. Then by the law of cosines,

$$\begin{aligned} a^2 &= b^2 + c^2 - 2bc \cos A \\ &= 2b^2 - 2b^2 \cos A \end{aligned}$$

and so $a/b = \sqrt{2(1 - \cos A)}$. Similar reasoning shows $a'/b' = \sqrt{2(1 - \cos A)}$ and so

$$a/b = a'/b'.$$

Similarly, $a/c = a'/c'$. By assumption $c/b = 1 = c'/b'$.

Such triangles in which two sides are equal are called isosceles.

The Fundamental Theorem Of Algebra

The fundamental theorem of algebra states that every non constant polynomial having coefficients in $\mathbb{C}$ has a zero in $\mathbb{C}$. If $\mathbb{C}$ is replaced by $\mathbb{R}$, this is not true because of the example, $x^2 + 1 = 0$. This theorem is a very remarkable result and notwithstanding its title, all the best proofs of it depend on either analysis or topology. It was studied by many people in the eighteenth century. Gauss gave a proof in 1797 which had a loose end. The first correct proof was given by Argand in 1806. A good discussion of the history of this theorem as well as many different ways to prove it are found in the Wikipedia article on the web. Just google fundamental theorem of algebra. The proof given here follows Rudin [23]. See also Hardy [15] for a similar proof, more discussion and references. The best proof is found in the theory of complex analysis.

Recall Corollary 1.9.15 which says every nonzero complex number has k k^{th} roots.

Now here is a definition of what it means for a sequence of complex numbers to converge. You will note it is the **same** definition given for a sequence of real numbers only here the absolute value refers to the absolute value of a complex number.

Definition B.1.6 *Let $\{z_n\}$ be a sequence of complex numbers. Then*

$$\lim_{n \rightarrow \infty} z_n = z$$

means: For every $\varepsilon > 0$ there exists n_ε such that if $n \geq n_\varepsilon$, then

$$|z_n - z| < \varepsilon.$$

Now here is a useful observation.

Proposition B.1.7 *Suppose $\{a_n + ib_n\}$ is a sequence of complex numbers. Then it converges to $a + ib$ if and only if $\lim_{n \rightarrow \infty} a_n = a$ and $\lim_{n \rightarrow \infty} b_n = b$.*

Proof: I will leave the details to you. Recall

$$|(a_n + ib_n) - (a + ib)| \equiv \sqrt{(a - a_n)^2 + (b - b_n)^2}.$$

This implies

$$\max(|a_n - a|, |b_n - b|) \leq |(a_n + ib_n) - (a + ib)| \leq |a_n - a| + |b_n - b|$$

Now the conclusion of the proposition follows right away. This proves the proposition.

Next is an important existence theorem.

Theorem B.1.8 Denote by $[a, b] + i[c, d]$ the set of complex numbers $x + iy$ where $x \in [a, b]$ and $y \in [c, d]$. Suppose $\{z_n\}$ is a sequence of complex numbers in this set. Then there exists a complex number $z \in [a, b] + i[c, d]$ and a subsequence $\{z_{n_k}\}$ such that

$$\lim_{k \rightarrow \infty} z_{n_k} = z.$$

Proof: Let $z_n = x_n + iy_n$ and consider first the sequence $\{x_n\} \subseteq [a, b]$. Consider the two intervals $[a, \frac{a+b}{2}]$, $[\frac{a+b}{2}, b]$, each having length $(b-a)/2$. Then in one of these intervals, perhaps both, there are x_n for infinitely many values of n . Pick one of the two intervals for which this is so. Then divide it in half. One of the two halves has x_n for infinitely many values of n . Pick the half for which this is so. Divide it in half and pick the half which has the property that it contains x_n for infinitely many n . Continue this way. Thus there is a sequence of intervals $\{I_n\}$ where the length of the n^{th} interval is no larger than $(b-a)2^{-n+1}$ and each of these intervals contains x_n for infinitely many values of n . Now pick $x_{k_1} \in I_1$. If $x_{k_1}, \dots, x_{k_n}$ have been chosen, let $x_{k_{n+1}}$ be in I_{n+1} such that $k_{n+1} > k_n$. This can be done because each interval contains x_n for infinitely many n . Then $\{x_{n_k}\}_{k=1}^\infty$ is a Cauchy sequence because for $k, l > n$,

$$|x_{n_k} - x_{n_l}| < (b-a)2^{-n+1}.$$

By Theorem 2.11.15 and Theorem 2.15.3, this sequence converges to some $x \in [a, b]$. Now consider the sequence $\{y_{n_k}\}_{k=1}^\infty \subseteq [c, d]$. By the same reasoning there is a subsequence $\{y_{n_{k_l}}\}_{l=1}^\infty$ which converges to $y \in [c, d]$. It follows from Theorem 2.11.10 that $\{x_{n_{k_l}}\}_{l=1}^\infty$ also converges to x . By Proposition B.1.7

$$\lim_{l \rightarrow \infty} x_{n_{k_l}} + iy_{n_{k_l}} = x + iy \in [a, b] + i[c, d].$$

This proves the theorem.

Lemma B.1.9 If $p(z) = \sum_{k=0}^n a_k z^k$ is a polynomial and if $\lim_{n \rightarrow \infty} z_n = z$, then

$$\lim_{n \rightarrow \infty} |p(z_n)| = |p(z)|.$$

Proof: $z_n = z_n - z + z$ and so by the binomial theorem,

$$\begin{aligned} p(z_n) &= a_0 + \sum_{k=1}^n a_k \sum_{j=0}^k \binom{k}{j} z^{k-j} (z_n - z)^j \\ &= a_0 + \sum_{k=1}^n a_k \left(z^k + \sum_{j=1}^k \binom{k}{j} z^{k-j} (z_n - z)^j \right) \\ &= p(z) + \sum_{k=1}^n a_k \sum_{j=1}^k \binom{k}{j} z^{k-j} (z_n - z)^j \equiv p(z) + e(z_n - z) \end{aligned}$$

where $\lim_{n \rightarrow \infty} e(z_n - z) = 0$. It follows from the triangle inequality

$$||p(z_n)| - |p(z)|| \leq |p(z_n) - p(z)| \leq |e(z_n - z)|$$

and so $\lim_{n \rightarrow \infty} |p(z_n)| = |p(z)|$ as claimed. This proves the lemma.

Lemma B.1.10 Let $p(z)$ be a polynomial with complex coefficients as above. Then $|p(z)|$ achieves its minimum value on any set of the form $[-R, R] + i[-R, R]$.

Proof: $|p(z)|$ is bounded below by 0 and so

$$\lambda \equiv \inf \{|p(z)| : z \in [-R, R] + i[-R, R]\}$$

exists. By Proposition 9.1.6 there exists a sequence $\{z_n\}$ of points of $[-R, R] + i[-R, R]$ which satisfy

$$\lambda = \lim_{n \rightarrow \infty} |p(z_n)|$$

This is called a minimizing sequence. Then by Theorem B.1.8 there is a subsequence $\{z_{n_k}\}$ which converges to a point $z \in [-R, R] + i[-R, R]$. Then by Lemma B.1.9,

$$\lambda = \lim_{k \rightarrow \infty} |p(z_{n_k})| = |p(z)|.$$

This proves the lemma.

Theorem B.1.11 (*Fundamental theorem of Algebra*) Let $p(z)$ be a nonconstant polynomial. Then there exists $z \in \mathbb{C}$ such that $p(z) = 0$.

Proof: Let

$$p(z) = \sum_{k=0}^n a_k z^k$$

where $a_n \neq 0$, $n > 0$. Then

$$|p(z)| \geq |a_n| |z|^n - \sum_{k=0}^{n-1} |a_k| |z|^k$$

and so

$$\lim_{|z| \rightarrow \infty} |p(z)| = \infty. \quad (2.1)$$

Now let

$$\lambda \equiv \inf \{|p(z)| : z \in \mathbb{C}\}.$$

By 2.1, there exists an $R > 0$ such that if $z \notin [-R, R] + i[-R, R]$, it follows that $|p(z)| > \lambda + 1$. Therefore,

$$\lambda \equiv \inf \{|p(z)| : z \in \mathbb{C}\} = \inf \{|p(z)| : z \in [-R, R] + i[-R, R]\}.$$

By Lemma B.1.10, there exists $w \in [-R, R] + i[-R, R]$ such that

$$\lambda = |p(w)|$$

I want to argue $\lambda = 0$. Suppose it is greater than 0. Then consider

$$q(z) \equiv \frac{p(z+w)}{p(w)}.$$

It follows $q(z)$ is of the form

$$q(z) = 1 + c_k z^k + \cdots + c_n z^n$$

where $c_k \neq 0$, because $q(0) = 1$. It is also true that

$$|q(z)| = \frac{|p(z+w)|}{|p(w)|} \geq 1$$

by the assumption that $|p(w)|$ is the smallest value of $|p(z)|$. Now let $\theta \in \mathbb{C}$ be a complex number with $|\theta| = 1$ and

$$\theta c_k w^k = -|w|^k |c_k|.$$

(If

$$w \neq 0, \theta = \frac{-|w|^k |c_k|}{w^k c_k}$$

and if $w = 0$, $\theta = 1$ will work.) Next let $\eta^k = \theta$ and let t be a small positive number.

$$q(t\eta w) \equiv 1 - t^k |w|^k |c_k| + \cdots + c_n t^n (\eta w)^n$$

which is of the form

$$1 - t^k |w|^k |c_k| + t^k (g(t, w))$$

where $\lim_{t \rightarrow 0} g(t, w) = 0$. Letting t be small enough,

$$|g(t, w)| < |w|^k |c_k| / 2$$

and so for such t ,

$$|q(t\eta w)| < 1 - t^k |w|^k |c_k| + t^k |w|^k |c_k| / 2 < 1,$$

a contradiction to $|q(z)| \geq 1$. This proves the theorem.

The Mathematical Theory Of Determinants*



C.1 The Function sgn_n

It is easiest to give a different definition of the determinant which is clearly well defined and then prove the earlier one in terms of Laplace expansion. Let $(i_1, \dots, i_n)$ be an ordered list of numbers from $\{1, \dots, n\}$. This means the order is important so $(1, 2, 3)$ and $(2, 1, 3)$ are different. There will be some repetition between this section and the earlier section on determinants. The main purpose is to give all the missing proofs. Two books which give a good introduction to determinants are Apostol [2] and Rudin [23]. A recent book which also has a good introduction is Baker [4].

The following Lemma will be essential in the definition of the determinant.

Lemma C.1.1 *There exists a unique function, sgn_n which maps each list of numbers from $\{1, \dots, n\}$ to one of the three numbers, 0, 1, or -1 which also has the following properties.*

$$\text{sgn}_n(1, \dots, n) = 1 \quad (3.1)$$

$$\text{sgn}_n(i_1, \dots, p, \dots, q, \dots, i_n) = -\text{sgn}_n(i_1, \dots, q, \dots, p, \dots, i_n) \quad (3.2)$$

In words, the second property states that if two of the numbers are switched, the value of the function is multiplied by -1 . Also, in the case where $n > 1$ and $\{i_1, \dots, i_n\} = \{1, \dots, n\}$ so that every number from $\{1, \dots, n\}$ appears in the ordered list, $(i_1, \dots, i_n)$,

$$\begin{aligned} \text{sgn}_n(i_1, \dots, i_{\theta-1}, n, i_{\theta+1}, \dots, i_n) &\equiv \\ (-1)^{n-\theta} \text{sgn}_{n-1}(i_1, \dots, i_{\theta-1}, i_{\theta+1}, \dots, i_n) &\quad (3.3) \end{aligned}$$

where $n = i_{\theta}$ in the ordered list, $(i_1, \dots, i_n)$.

Proof: To begin with, it is necessary to show the existence of such a function. This is clearly true if $n = 1$. Define $\text{sgn}_1(1) \equiv 1$ and observe that it works. No switching is possible. In the case where $n = 2$, it is also clearly true. Let $\text{sgn}_2(1, 2) = 1$ and $\text{sgn}_2(2, 1) = -1$ while $\text{sgn}_2(2, 2) = \text{sgn}_2(1, 1) = 0$ and verify it works. Assuming such a function exists for n , sgn_{n+1} will be defined in terms of sgn_n . If there are any repeated numbers in $(i_1, \dots, i_{n+1})$, $\text{sgn}_{n+1}(i_1, \dots, i_{n+1}) \equiv 0$. If there are no repeats, then $n+1$ appears somewhere in the ordered list. Let θ be the position of the number $n+1$ in the list. Thus, the list is of the form $(i_1, \dots, i_{\theta-1}, n+1, i_{\theta+1}, \dots, i_{n+1})$. From 3.3 it must be that

$$\begin{aligned} \text{sgn}_{n+1}(i_1, \dots, i_{\theta-1}, n+1, i_{\theta+1}, \dots, i_{n+1}) &\equiv \\ (-1)^{n+1-\theta} \text{sgn}_n(i_1, \dots, i_{\theta-1}, i_{\theta+1}, \dots, i_{n+1}). \end{aligned}$$

It is necessary to verify this satisfies 3.1 and 3.2 with n replaced with $n+1$. The first of these is obviously true because

$$\text{sgn}_{n+1}(1, \dots, n, n+1) \equiv (-1)^{n+1-(n+1)} \text{sgn}_n(1, \dots, n) = 1.$$

If there are repeated numbers in $(i_1, \dots, i_{n+1})$, then it is obvious 3.2 holds because both sides would equal zero from the above definition. It remains to verify 3.2 in the case where there are no numbers repeated in $(i_1, \dots, i_{n+1})$. Consider

$$\text{sgn}_{n+1}(i_1, \dots, \overset{r}{p}, \dots, \overset{s}{q}, \dots, i_{n+1}),$$

where the r above the p indicates the number, p is in the r^{th} position and the s above the q indicates that the number, q is in the s^{th} position. Suppose first that $r < \theta < s$. Then

$$\begin{aligned} \text{sgn}_{n+1}(i_1, \dots, \overset{r}{p}, \dots, n+1, \dots, \overset{s}{q}, \dots, i_{n+1}) &\equiv \\ (-1)^{n+1-\theta} \text{sgn}_n(i_1, \dots, \overset{r}{p}, \dots, \overset{s-1}{q}, \dots, i_{n+1}) \end{aligned}$$

while

$$\begin{aligned} \text{sgn}_{n+1}(i_1, \dots, \overset{r}{q}, \dots, n+1, \dots, \overset{s}{p}, \dots, i_{n+1}) &= \\ (-1)^{n+1-\theta} \text{sgn}_n(i_1, \dots, \overset{r}{q}, \dots, \overset{s-1}{p}, \dots, i_{n+1}) \end{aligned}$$

and so, by induction, a switch of p and q introduces a minus sign in the result. Similarly, if $\theta > s$ or if $\theta < r$ it also follows that 3.2 holds. The interesting case is when $\theta = r$ or $\theta = s$. Consider the case where $\theta = r$ and note the other case is entirely similar.

$$\begin{aligned} \text{sgn}_{n+1}(i_1, \dots, n+1, \dots, \overset{s}{q}, \dots, i_{n+1}) &= \\ (-1)^{n+1-r} \text{sgn}_n(i_1, \dots, \overset{s-1}{q}, \dots, i_{n+1}) \end{aligned} \quad (3.4)$$

while

$$\begin{aligned} \text{sgn}_{n+1}(i_1, \dots, \overset{r}{q}, \dots, n+1, \dots, i_{n+1}) &= \\ (-1)^{n+1-s} \text{sgn}_n(i_1, \dots, \overset{r}{q}, \dots, i_{n+1}). \end{aligned} \quad (3.5)$$

By making $s-1-r$ switches, move the q which is in the $s-1^{\text{th}}$ position in 3.4 to the r^{th} position in 3.5. By induction, each of these switches introduces a factor of -1 and so

$$\text{sgn}_n(i_1, \dots, \overset{s-1}{q}, \dots, i_{n+1}) = (-1)^{s-1-r} \text{sgn}_n(i_1, \dots, \overset{r}{q}, \dots, i_{n+1}).$$

Therefore,

$$\begin{aligned}
 \operatorname{sgn}_{n+1} \left(i_1, \dots, n+1, \dots, \overset{r}{q}, \dots, i_{n+1} \right) &= (-1)^{n+1-r} \operatorname{sgn}_n \left(i_1, \dots, \overset{s-1}{q}, \dots, i_{n+1} \right) \\
 &= (-1)^{n+1-r} (-1)^{s-1-r} \operatorname{sgn}_n \left(i_1, \dots, \overset{r}{q}, \dots, i_{n+1} \right) \\
 &= (-1)^{n+s} \operatorname{sgn}_n \left(i_1, \dots, \overset{r}{q}, \dots, i_{n+1} \right) = (-1)^{2s-1} (-1)^{n+1-s} \operatorname{sgn}_n \left(i_1, \dots, \overset{r}{q}, \dots, i_{n+1} \right) \\
 &= -\operatorname{sgn}_{n+1} \left(i_1, \dots, \overset{r}{q}, \dots, n+1, \dots, i_{n+1} \right).
 \end{aligned}$$

This proves the existence of the desired function.

To see this function is unique, note that you can obtain any ordered list of distinct numbers from a sequence of switches. If there exist two functions, f and g both satisfying 3.1 and 3.2, you could start with $f(1, \dots, n) = g(1, \dots, n)$ and applying the same sequence of switches, eventually arrive at $f(i_1, \dots, i_n) = g(i_1, \dots, i_n)$. If any numbers are repeated, then 3.2 gives both functions are equal to zero for that ordered list. This proves the lemma.

C.2 The Determinant

C.2.1 The Definition

In what follows sgn will often be used rather than sgn_n because the context supplies the appropriate n .

Definition C.2.1 Let f be a real valued function which has the set of ordered lists of numbers from $\{1, \dots, n\}$ as its domain. Define

$$\sum_{(k_1, \dots, k_n)} f(k_1 \cdots k_n)$$

to be the sum of all the $f(k_1 \cdots k_n)$ for all possible choices of ordered lists $(k_1, \dots, k_n)$ of numbers of $\{1, \dots, n\}$. For example,

$$\sum_{(k_1, k_2)} f(k_1, k_2) = f(1, 2) + f(2, 1) + f(1, 1) + f(2, 2).$$

Definition C.2.2 Let $(a_{ij}) = A$ denote an $n \times n$ matrix. The determinant of A , denoted by $\det(A)$ is defined by

$$\det(A) \equiv \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(k_1, \dots, k_n) a_{1k_1} \cdots a_{nk_n}$$

where the sum is taken over all ordered lists of numbers from $\{1, \dots, n\}$. Note it suffices to take the sum over only those ordered lists in which there are no repeats because if there are, $\operatorname{sgn}(k_1, \dots, k_n) = 0$ and so that term contributes 0 to the sum.

Let A be an $n \times n$ matrix, $A = (a_{ij})$ and let $(r_1, \dots, r_n)$ denote an ordered list of n numbers from $\{1, \dots, n\}$. Let $A(r_1, \dots, r_n)$ denote the matrix whose k^{th} row is the r_k row of the matrix, A . Thus

$$\det(A(r_1, \dots, r_n)) = \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(k_1, \dots, k_n) a_{r_1 k_1} \cdots a_{r_n k_n} \quad (3.6)$$

and

$$A(1, \dots, n) = A.$$

C.2.2 Permuting Rows Or Columns

Proposition C.2.3 *Let*

$$(r_1, \dots, r_n)$$

be an ordered list of numbers from $\{1, \dots, n\}$. Then

$$\begin{aligned} & \operatorname{sgn}(r_1, \dots, r_n) \det(A) \\ &= \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(k_1, \dots, k_n) a_{r_1 k_1} \cdots a_{r_n k_n} \\ &= \det(A(r_1, \dots, r_n)). \end{aligned} \quad (3.7)$$

$$(3.8)$$

Proof: Let $(1, \dots, n) = (1, \dots, r, \dots, s, \dots, n)$ so $r < s$.

$$\det(A(1, \dots, r, \dots, s, \dots, n)) = \quad (3.9)$$

$$\sum_{(k_1, \dots, k_n)} \operatorname{sgn}(k_1, \dots, k_r, \dots, k_s, \dots, k_n) a_{1k_1} \cdots a_{rk_r} \cdots a_{sk_s} \cdots a_{nk_n},$$

and renaming the variables, calling k_s, k_r and k_r, k_s , this equals

$$\begin{aligned} &= \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(k_1, \dots, k_s, \dots, k_r, \dots, k_n) a_{1k_1} \cdots a_{rk_s} \cdots a_{sk_r} \cdots a_{nk_n} \\ &= \sum_{(k_1, \dots, k_n)} -\operatorname{sgn} \left(k_1, \dots, \overbrace{k_r, \dots, k_s}^{\text{These got switched}}, \dots, k_n \right) a_{1k_1} \cdots a_{sk_r} \cdots a_{rk_s} \cdots a_{nk_n} \\ &= -\det(A(1, \dots, s, \dots, r, \dots, n)). \end{aligned} \quad (3.10)$$

Consequently,

$$\begin{aligned} & \det(A(1, \dots, s, \dots, r, \dots, n)) = \\ & -\det(A(1, \dots, r, \dots, s, \dots, n)) = -\det(A) \end{aligned}$$

Now letting $A(1, \dots, s, \dots, r, \dots, n)$ play the role of A , and continuing in this way, switching pairs of numbers,

$$\det(A(r_1, \dots, r_n)) = (-1)^p \det(A)$$

where it took p switches to obtain $(r_1, \dots, r_n)$ from $(1, \dots, n)$. By Lemma C.1.1, this implies

$$\det(A(r_1, \dots, r_n)) = (-1)^p \det(A) = \operatorname{sgn}(r_1, \dots, r_n) \det(A)$$

and proves the proposition in the case when there are no repeated numbers in the ordered list, $(r_1, \dots, r_n)$. However, if there is a repeat, say the r^{th} row equals the s^{th} row, then the reasoning of 3.9-3.10 shows that $A(r_1, \dots, r_n) = 0$ and also $\operatorname{sgn}(r_1, \dots, r_n) = 0$ so the formula holds in this case also.

Observation C.2.4 *There are $n!$ ordered lists of distinct numbers from $\{1, \dots, n\}$.*

To see this, consider n slots placed in order. There are n choices for the first slot. For each of these choices, there are $n-1$ choices for the second. Thus there are $n(n-1)$ ways to fill the first two slots. Then for each of these ways there are $n-2$ choices left for the third slot. Continuing this way, there are $n!$ ordered lists of distinct numbers from $\{1, \dots, n\}$ as stated in the observation.

C.2.3 A Symmetric Definition

With the above, it is possible to give a more symmetric description of the determinant from which it will follow that $\det(A) = \det(A^T)$.

Corollary C.2.5 *The following formula for $\det(A)$ is valid.*

$$\det(A) = \frac{1}{n!} \sum_{(r_1, \dots, r_n)} \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(r_1, \dots, r_n) \operatorname{sgn}(k_1, \dots, k_n) a_{r_1 k_1} \cdots a_{r_n k_n}. \quad (3.11)$$

And also $\det(A^T) = \det(A)$ where A^T is the transpose of A . (Recall that for $A^T = (a_{ij}^T)$, $a_{ij}^T = a_{ji}$.)

Proof: From Proposition C.2.3, if the r_i are distinct,

$$\det(A) = \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(r_1, \dots, r_n) \operatorname{sgn}(k_1, \dots, k_n) a_{r_1 k_1} \cdots a_{r_n k_n}.$$

Summing over all ordered lists, $(r_1, \dots, r_n)$ where the r_i are distinct, (If the r_i are not distinct, $\operatorname{sgn}(r_1, \dots, r_n) = 0$ and so there is no contribution to the sum.)

$$n! \det(A) = \sum_{(r_1, \dots, r_n)} \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(r_1, \dots, r_n) \operatorname{sgn}(k_1, \dots, k_n) a_{r_1 k_1} \cdots a_{r_n k_n}.$$

This proves the corollary since the formula gives the same number for A as it does for A^T .

C.2.4 The Alternating Property Of The Determinant

Corollary C.2.6 *If two rows or two columns in an $n \times n$ matrix, A , are switched, the determinant of the resulting matrix equals (-1) times the determinant of the original matrix. If A is an $n \times n$ matrix in which two rows are equal or two columns are equal then $\det(A) = 0$. Suppose the i^{th} row of A equals $(xa_1 + yb_1, \dots, xa_n + yb_n)$. Then*

$$\det(A) = x \det(A_1) + y \det(A_2)$$

where the i^{th} row of A_1 is $(a_1, \dots, a_n)$ and the i^{th} row of A_2 is $(b_1, \dots, b_n)$, all other rows of A_1 and A_2 coinciding with those of A . In other words, $\det$ is a linear function of each row A . The same is true with the word "row" replaced with the word "column".

Proof: By Proposition C.2.3 when two rows are switched, the determinant of the resulting matrix is (-1) times the determinant of the original matrix. By Corollary C.2.5 the same holds for columns because the columns of the matrix equal the rows of the transposed matrix. Thus if A_1 is the matrix obtained from A by switching two columns,

$$\det(A) = \det(A^T) = -\det(A_1^T) = -\det(A_1).$$

If A has two equal columns or two equal rows, then switching them results in the same matrix. Therefore, $\det(A) = -\det(A)$ and so $\det(A) = 0$.

It remains to verify the last assertion.

$$\det(A) \equiv \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(k_1, \dots, k_n) a_{1k_1} \cdots (xa_{ik_i} + yb_{ik_i}) \cdots a_{nk_n}$$

$$\begin{aligned}
&= x \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(k_1, \dots, k_n) a_{1k_1} \cdots a_{k_i} \cdots a_{nk_n} \\
&\quad + y \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(k_1, \dots, k_n) a_{1k_1} \cdots b_{k_i} \cdots a_{nk_n} \\
&\equiv x \det(A_1) + y \det(A_2).
\end{aligned}$$

The same is true of columns because $\det(A^T) = \det(A)$ and the rows of A^T are the columns of A .

C.2.5 Linear Combinations And Determinants

Definition C.2.7 A vector, $\mathbf{w}$, is a linear combination of the vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ if there exists scalars, $c_1, \dots, c_r$ such that $\mathbf{w} = \sum_{k=1}^r c_k \mathbf{v}_k$. This is the same as saying $\mathbf{w} \in \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$.

The following corollary is also of great use.

Corollary C.2.8 Suppose A is an $n \times n$ matrix and some column (row) is a linear combination of r other columns (rows). Then $\det(A) = 0$.

Proof: Let $A = (\mathbf{a}_1 \cdots \mathbf{a}_n)$ be the columns of A and suppose the condition that one column is a linear combination of r of the others is satisfied. Then by using Corollary C.2.6 you may rearrange the columns to have the n^{th} column a linear combination of the first r columns. Thus $\mathbf{a}_n = \sum_{k=1}^r c_k \mathbf{a}_k$ and so

$$\det(A) = \det(\mathbf{a}_1 \cdots \mathbf{a}_r \cdots \mathbf{a}_{n-1} \sum_{k=1}^r c_k \mathbf{a}_k).$$

By Corollary C.2.6

$$\det(A) = \sum_{k=1}^r c_k \det(\mathbf{a}_1 \cdots \mathbf{a}_r \cdots \mathbf{a}_{n-1} \mathbf{a}_k) = 0.$$

The case for rows follows from the fact that $\det(A) = \det(A^T)$. This proves the corollary.

C.2.6 The Determinant Of A Product

Recall the following definition of matrix multiplication.

Definition C.2.9 If A and B are $n \times n$ matrices, $A = (a_{ij})$ and $B = (b_{ij})$, $AB = (c_{ij})$ where

$$c_{ij} \equiv \sum_{k=1}^n a_{ik} b_{kj}.$$

One of the most important rules about determinants is that the determinant of a product equals the product of the determinants.

Theorem C.2.10 Let A and B be $n \times n$ matrices. Then

$$\det(AB) = \det(A) \det(B).$$

Proof: Let c_{ij} be the ij^{th} entry of AB . Then by Proposition C.2.3,

$$\begin{aligned} \det(AB) &= \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(k_1, \dots, k_n) c_{1k_1} \cdots c_{nk_n} \\ &= \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(k_1, \dots, k_n) \left(\sum_{r_1} a_{1r_1} b_{r_1 k_1} \right) \cdots \left(\sum_{r_n} a_{nr_n} b_{r_n k_n} \right) \\ &= \sum_{(r_1, \dots, r_n)} \sum_{(k_1, \dots, k_n)} \operatorname{sgn}(k_1, \dots, k_n) b_{r_1 k_1} \cdots b_{r_n k_n} (a_{1r_1} \cdots a_{nr_n}) \\ &= \sum_{(r_1, \dots, r_n)} \operatorname{sgn}(r_1 \cdots r_n) a_{1r_1} \cdots a_{nr_n} \det(B) = \det(A) \det(B). \end{aligned}$$

This proves the theorem.

C.2.7 Cofactor Expansions

Lemma C.2.11 Suppose a matrix is of the form

$$M = \begin{pmatrix} A & * \\ \mathbf{0} & a \end{pmatrix} \quad (3.12)$$

or

$$M = \begin{pmatrix} A & \mathbf{0} \\ * & a \end{pmatrix} \quad (3.13)$$

where a is a number and A is an $(n-1) \times (n-1)$ matrix and $*$ denotes either a column or a row having length $n-1$ and the $\mathbf{0}$ denotes either a column or a row of length $n-1$ consisting entirely of zeros. Then

$$\det(M) = a \det(A).$$

Proof: Denote M by (m_{ij}) . Thus in the first case, $m_{nn} = a$ and $m_{ni} = 0$ if $i \neq n$ while in the second case, $m_{nn} = a$ and $m_{in} = 0$ if $i \neq n$. From the definition of the determinant,

$$\det(M) \equiv \sum_{(k_1, \dots, k_n)} \operatorname{sgn}_n(k_1, \dots, k_n) m_{1k_1} \cdots m_{nk_n}$$

Letting θ denote the position of n in the ordered list, $(k_1, \dots, k_n)$ then using the earlier conventions used to prove Lemma C.1.1, $\det(M)$ equals

$$\sum_{(k_1, \dots, k_n)} (-1)^{n-\theta} \operatorname{sgn}_{n-1} \left(k_1, \dots, k_{\theta-1}, k_{\theta+1}, \dots, k_n \right) m_{1k_1} \cdots m_{nk_n}$$

Now suppose 3.13. Then if $k_n \neq n$, the term involving m_{nk_n} in the above expression equals zero. Therefore, the only terms which survive are those for which $\theta = n$ or in other words, those for which $k_n = n$. Therefore, the above expression reduces to

$$a \sum_{(k_1, \dots, k_{n-1})} \operatorname{sgn}_{n-1}(k_1, \dots, k_{n-1}) m_{1k_1} \cdots m_{(n-1)k_{n-1}} = a \det(A).$$

To get the assertion in the situation of 3.12 use Corollary C.2.5 and 3.13 to write

$$\det(M) = \det(M^T) = \det \left(\begin{pmatrix} A^T & \mathbf{0} \\ * & a \end{pmatrix} \right) = a \det(A^T) = a \det(A).$$

This proves the lemma.

In terms of the theory of determinants, arguably the most important idea is that of Laplace expansion along a row or a column. This will follow from the above definition of a determinant.

Definition C.2.12 Let $A = (a_{ij})$ be an $n \times n$ matrix. Then a new matrix called the cofactor matrix, $\text{cof}(A)$ is defined by $\text{cof}(A) = (c_{ij})$ where to obtain c_{ij} delete the i^{th} row and the j^{th} column of A , take the determinant of the $(n-1) \times (n-1)$ matrix which results, (This is called the ij^{th} minor of A .) and then multiply this number by $(-1)^{i+j}$. To make the formulas easier to remember, $\text{cof}(A)_{ij}$ will denote the ij^{th} entry of the cofactor matrix.

The following is the main result. Earlier this was given as a definition and the outrageous totally unjustified assertion was made that the same number would be obtained by expanding the determinant along any row or column. The following theorem proves this assertion.

Theorem C.2.13 Let A be an $n \times n$ matrix where $n \geq 2$. Then

$$\det(A) = \sum_{j=1}^n a_{ij} \text{cof}(A)_{ij} = \sum_{i=1}^n a_{ij} \text{cof}(A)_{ij}.$$

The first formula consists of expanding the determinant along the i^{th} row and the second expands the determinant along the j^{th} column.

Proof: Let $(a_{i1}, \dots, a_{in})$ be the i^{th} row of A . Let B_j be the matrix obtained from A by leaving every row the same except the i^{th} row which in B_j equals $(0, \dots, 0, a_{ij}, 0, \dots, 0)$. Then by Corollary C.2.6,

$$\det(A) = \sum_{j=1}^n \det(B_j)$$

Denote by A^{ij} the $(n-1) \times (n-1)$ matrix obtained by deleting the i^{th} row and the j^{th} column of A . Thus $\text{cof}(A)_{ij} \equiv (-1)^{i+j} \det(A^{ij})$. At this point, recall that from Proposition C.2.3, when two rows or two columns in a matrix, M , are switched, this results in multiplying the determinant of the old matrix by -1 to get the determinant of the new matrix. Therefore, by Lemma C.2.11,

$$\begin{aligned} \det(B_j) &= (-1)^{n-j} (-1)^{n-i} \det \left(\begin{pmatrix} A^{ij} & * \\ \mathbf{0} & a_{ij} \end{pmatrix} \right) \\ &= (-1)^{i+j} \det \left(\begin{pmatrix} A^{ij} & * \\ \mathbf{0} & a_{ij} \end{pmatrix} \right) = a_{ij} \text{cof}(A)_{ij}. \end{aligned}$$

Therefore,

$$\det(A) = \sum_{j=1}^n a_{ij} \text{cof}(A)_{ij}$$

which is the formula for expanding $\det(A)$ along the i^{th} row. Also,

$$\begin{aligned} \det(A) &= \det(A^T) = \sum_{j=1}^n a_{ij}^T \text{cof}(A^T)_{ij} \\ &= \sum_{j=1}^n a_{ji} \text{cof}(A)_{ji} \end{aligned}$$

which is the formula for expanding $\det(A)$ along the i^{th} column. This proves the theorem. Note that this gives an easy way to write a formula for the inverse of an $n \times n$ matrix. Recall the definition of the inverse of a matrix in Definition 12.1.28 on Page 317.

C.2.8 Formula For The Inverse

Theorem C.2.14 A^{-1} exists if and only if $\det(A) \neq 0$. If $\det(A) \neq 0$, then $A^{-1} = (a_{ij}^{-1})$ where

$$a_{ij}^{-1} = \det(A)^{-1} \operatorname{cof}(A)_{ji}$$

for $\operatorname{cof}(A)_{ij}$ the ij^{th} cofactor of A .

Proof: By Theorem C.2.13 and letting $(a_{ir}) = A$, if $\det(A) \neq 0$,

$$\sum_{i=1}^n a_{ir} \operatorname{cof}(A)_{ir} \det(A)^{-1} = \det(A) \det(A)^{-1} = 1.$$

Now consider

$$\sum_{i=1}^n a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

when $k \neq r$. Replace the k^{th} column with the r^{th} column to obtain a matrix, B_k whose determinant equals zero by Corollary C.2.6. However, expanding this matrix along the k^{th} column yields

$$0 = \det(B_k) \det(A)^{-1} = \sum_{i=1}^n a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

Summarizing,

$$\sum_{i=1}^n a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1} = \delta_{rk}.$$

Using the other formula in Theorem C.2.13, and similar reasoning,

$$\sum_{j=1}^n a_{rj} \operatorname{cof}(A)_{kj} \det(A)^{-1} = \delta_{rk}$$

This proves that if $\det(A) \neq 0$, then A^{-1} exists with $A^{-1} = (a_{ij}^{-1})$, where

$$a_{ij}^{-1} = \operatorname{cof}(A)_{ji} \det(A)^{-1}.$$

Now suppose A^{-1} exists. Then by Theorem C.2.10,

$$1 = \det(I) = \det(AA^{-1}) = \det(A) \det(A^{-1})$$

so $\det(A) \neq 0$. This proves the theorem.

The next corollary points out that if an $n \times n$ matrix, A has a right or a left inverse, then it has an inverse.

Corollary C.2.15 Let A be an $n \times n$ matrix and suppose there exists an $n \times n$ matrix, B such that $BA = I$. Then A^{-1} exists and $A^{-1} = B$. Also, if there exists C an $n \times n$ matrix such that $AC = I$, then A^{-1} exists and $A^{-1} = C$.

Proof: Since $BA = I$, Theorem C.2.10 implies

$$\det B \det A = 1$$

and so $\det A \neq 0$. Therefore from Theorem C.2.14, A^{-1} exists. Therefore,

$$A^{-1} = (BA) A^{-1} = B (AA^{-1}) = BI = B.$$

The case where $CA = I$ is handled similarly.

The conclusion of this corollary is that left inverses, right inverses and inverses are all the same in the context of $n \times n$ matrices.

Theorem C.2.14 says that to find the inverse, take the transpose of the cofactor matrix and divide by the determinant. The transpose of the cofactor matrix is called the adjugate or sometimes the classical adjoint of the matrix A . It is an abomination to call it the adjoint although you do sometimes see it referred to in this way. In words, A^{-1} is equal to one over the determinant of A times the adjugate matrix of A .

C.2.9 Cramer's Rule

In case you are solving a system of equations, $A\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$, it follows that if A^{-1} exists,

$$\mathbf{x} = (A^{-1}A)\mathbf{x} = A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{y}$$

thus solving the system. Now in the case that A^{-1} exists, there is a formula for A^{-1} given above. Using this formula,

$$x_i = \sum_{j=1}^n a_{ij}^{-1} y_j = \sum_{j=1}^n \frac{1}{\det(A)} \operatorname{cof}(A)_{ji} y_j.$$

By the formula for the expansion of a determinant along a column,

$$x_i = \frac{1}{\det(A)} \det \begin{pmatrix} * & \cdots & y_1 & \cdots & * \\ \vdots & & \vdots & & \vdots \\ * & \cdots & y_n & \cdots & * \end{pmatrix},$$

where here the i^{th} column of A is replaced with the column vector, $(y_1, \dots, y_n)^T$, and the determinant of this modified matrix is taken and divided by $\det(A)$. This formula is known as Cramer's rule.

C.2.10 Upper Triangular Matrices

Definition C.2.16 A matrix M , is upper triangular if $M_{ij} = 0$ whenever $i > j$. Thus such a matrix equals zero below the main diagonal, the entries of the form M_{ii} as shown.

$$\begin{pmatrix} * & * & \cdots & * \\ 0 & * & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & * \end{pmatrix}$$

A lower triangular matrix is defined similarly as a matrix for which all entries above the main diagonal are equal to zero.

With this definition, here is a simple corollary of Theorem C.2.13.

Corollary C.2.17 Let M be an upper (lower) triangular matrix. Then $\det(M)$ is obtained by taking the product of the entries on the main diagonal.

C.2.11 The Determinant Rank

Definition C.2.18 A submatrix of a matrix A is the rectangular array of numbers obtained by deleting some rows and columns of A . Let A be an $m \times n$ matrix. The **determinant rank** of the matrix equals r where r is the largest number such that some $r \times r$ submatrix of A has a non zero determinant.

Theorem C.2.19 If A , an $m \times n$ matrix has determinant rank, r , then there exist r rows (columns) of the matrix such that every other row (column) is a linear combination of these r rows (columns).

Proof: Suppose the determinant rank of $A = (a_{ij})$ equals r . Thus some $r \times r$ submatrix has non zero determinant and there is no larger square submatrix which has non zero determinant. Suppose such a submatrix is determined by the r columns whose indices are

$$j_1 < \cdots < j_r$$

and the r rows whose indices are

$$i_1 < \cdots < i_r$$

I want to show that every row is a linear combination of these rows. Consider the l^{th} row and let p be an index between 1 and n . Form the following $(r+1) \times (r+1)$ matrix

$$\begin{pmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_r} & a_{i_1 p} \\ \vdots & & \vdots & \vdots \\ a_{i_r j_1} & \cdots & a_{i_r j_r} & a_{i_r p} \\ a_{l j_1} & \cdots & a_{l j_r} & a_{l p} \end{pmatrix}$$

Of course you can assume $l \notin \{i_1, \dots, i_r\}$ because there is nothing to prove if the l^{th} row is one of the chosen ones. The above matrix has determinant 0. This is because if $p \notin \{j_1, \dots, j_r\}$ then the above would be a submatrix of A which is too large to have non zero determinant. On the other hand, if $p \in \{j_1, \dots, j_r\}$ then the above matrix has two columns which are equal so its determinant is still 0.

Expand the determinant of the above matrix along the last column. Let C_k denote the cofactor associated with the entry $a_{i_k p}$. This is not dependent on the choice of p . Remember, you delete the column and the row the entry is in and take the determinant of what is left and multiply by -1 raised to an appropriate power. Let C denote the cofactor associated with $a_{l p}$. This is given to be nonzero, it being the determinant of the matrix

$$\begin{pmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_r} \\ \vdots & & \vdots \\ a_{i_r j_1} & \cdots & a_{i_r j_r} \end{pmatrix}$$

Thus

$$0 = a_{l p} C + \sum_{k=1}^r C_k a_{i_k p}$$

which implies

$$a_{l p} = \sum_{k=1}^r \frac{-C_k}{C} a_{i_k p} \equiv \sum_{k=1}^r m_k a_{i_k p}$$

Since this is true for every p and since m_k does not depend on p , this has shown the l^{th} row is a linear combination of the $i_1, i_2, \dots, i_r$ rows. The determinant rank does not change when you replace A with A^T . Therefore, the same conclusion holds for the columns. This proves the theorem.

C.2.12 Telling Whether A Is One To One Or Onto

The following theorem is of fundamental importance and ties together many of the ideas presented above.

Theorem C.2.20 *Let A be an $n \times n$ matrix. Then the following are equivalent.*

1. $\det(A) = 0$.
2. A, A^T are not one to one.
3. A is not onto.

Proof: Suppose $\det(A) = 0$. Then the determinant rank of $A = r < n$. Therefore, there exist r columns such that every other column is a linear combination of these columns by Theorem C.2.19. In particular, it follows that for some m , the m^{th} column is a linear combination of all the others. Thus letting $A = (\mathbf{a}_1 \cdots \mathbf{a}_m \cdots \mathbf{a}_n)$ where the columns are denoted by $\mathbf{a}_i$, there exists scalars, α_i such that

$$\mathbf{a}_m = \sum_{k \neq m} \alpha_k \mathbf{a}_k.$$

Now consider the column vector, $\mathbf{x} \equiv (\alpha_1 \cdots -1 \cdots \alpha_n)^T$. Then

$$A\mathbf{x} = -\mathbf{a}_m + \sum_{k \neq m} \alpha_k \mathbf{a}_k = \mathbf{0}.$$

Since also $A\mathbf{0} = \mathbf{0}$, it follows A is not one to one. Similarly, A^T is not one to one by the same argument applied to A^T . This verifies that 1.) implies 2.).

Now suppose 2.). Then since A^T is not one to one, it follows there exists $\mathbf{x} \neq \mathbf{0}$ such that

$$A^T \mathbf{x} = \mathbf{0}.$$

Taking the transpose of both sides yields

$$\mathbf{x}^T A = \mathbf{0}^T$$

where the $\mathbf{0}^T$ is a $1 \times n$ matrix or row vector. Now if $A\mathbf{y} = \mathbf{x}$, then

$$|\mathbf{x}|^2 = \mathbf{x}^T (A\mathbf{y}) = (\mathbf{x}^T A) \mathbf{y} = \mathbf{0}^T \mathbf{y} = 0$$

contrary to $\mathbf{x} \neq \mathbf{0}$. Consequently there can be no $\mathbf{y}$ such that $A\mathbf{y} = \mathbf{x}$ and so A is not onto. This shows that 2.) implies 3.).

Finally, suppose 3.). If 1.) does not hold, then $\det(A) \neq 0$ but then from Theorem C.2.14 A^{-1} exists and so for every $\mathbf{y} \in \mathbb{F}^n$ there exists a unique $\mathbf{x} \in \mathbb{F}^n$ such that $A\mathbf{x} = \mathbf{y}$. In fact $\mathbf{x} = A^{-1}\mathbf{y}$. Thus A would be onto contrary to 3.). This shows 3.) implies 1.) and proves the theorem.

Corollary C.2.21 *Let A be an $n \times n$ matrix. Then the following are equivalent.*

1. $\det(A) \neq 0$.
2. A and A^T are one to one.
3. A is onto.

Proof: This follows immediately from the above theorem.

C.2.13 Schur's Theorem

Consider the following system of equations for $x_1, x_2, \dots, x_n$

$$\sum_{j=1}^n a_{ij}x_j = 0, \quad i = 1, 2, \dots, m \quad (3.14)$$

where $m < n$. Then the following theorem is a fundamental observation.

Theorem C.2.22 *Let the system of equations be as just described in 3.14 where $m < n$. Then letting*

$$\mathbf{x}^T \equiv (x_1, x_2, \dots, x_n) \in \mathbb{R}^n,$$

there exists $\mathbf{x} \neq \mathbf{0}$ such that the components satisfy each of the equations of 3.14.

Proof: The above system is of the form

$$A\mathbf{x} = \mathbf{0}$$

where A is an $m \times n$ matrix with $m < n$. Therefore, if you form the matrix

$$\begin{pmatrix} A \\ 0 \end{pmatrix},$$

an $n \times n$ matrix having $n - m$ rows of zeros on the bottom, it follows this matrix has determinant equal to 0. Therefore, from Theorem C.2.20, there exists $\mathbf{x} \neq \mathbf{0}$ such that $A\mathbf{x} = \mathbf{0}$. This proves the theorem.

Definition C.2.23 *A set of vectors in $\mathbb{R}^n$ $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ is called an **orthonormal set of vectors** if*

$$\mathbf{x}_i \cdot \mathbf{x}_j = \delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Theorem C.2.24 *Let $\mathbf{v}_1$ be a unit vector ($|\mathbf{v}_1| = 1$) in $\mathbb{R}^n$, $n > 1$. Then there exist vectors $\{\mathbf{v}_2, \dots, \mathbf{v}_n\}$ such that*

$$\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$$

is an orthonormal set of vectors.

Proof: The equation for $\mathbf{x}$

$$\mathbf{v}_1 \cdot \mathbf{x} = 0$$

has a nonzero solution $\mathbf{x}$ by Theorem C.2.22. Pick such a solution and divide by its magnitude to get $\mathbf{v}_2$ a unit vector such that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. Now suppose $\mathbf{v}_1, \dots, \mathbf{v}_k$ have been chosen such that $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is an orthonormal set of vectors. Then consider the equations

$$\mathbf{v}_j \cdot \mathbf{x} = 0 \quad j = 1, 2, \dots, k$$

This amounts to the situation of Theorem C.2.22 in which there are more variables than equations. Therefore, by this theorem, there exists a nonzero $\mathbf{x}$ solving all these equations. Divide by its magnitude and this gives $\mathbf{v}_{k+1}$. This proves the theorem.

Definition C.2.25 *If U is an $n \times n$ matrix whose columns form an orthonormal set of vectors, then U is called an **orthogonal matrix**. Note that from the way we multiply matrices,*

$$U^T U = U U^T = I.$$

Thus $U^{-1} = U^T$.

Note the product of orthogonal matrices is orthogonal because

$$(U_1 U_2)^T (U_1 U_2) = U_2^T U_1^T U_1 U_2 = I.$$

Two matrices A and B are similar if there is some invertible matrix S such that $A = S^{-1}BS$. Note that similar matrices have the same characteristic equation because by Theorem C.2.10 which says the determinant of a product is the product of the determinants,

$$\begin{aligned} \det(\lambda I - A) &= \det(\lambda I - S^{-1}BS) = \det(S^{-1}(\lambda I - B)S) \\ &= \det(S^{-1}) \det(\lambda I - B) \det(S) = \det(S^{-1}S) \det(\lambda I - B) = \det(\lambda I - B) \end{aligned}$$

With this preparation, here is a case of Schur's theorem.

Theorem C.2.26 *Let A be a real $n \times n$ matrix which has all real eigenvalues. Then there exists an orthogonal matrix, U such that*

$$U^T A U = T, \quad (3.15)$$

where T is an upper triangular matrix having the eigenvalues of A on the main diagonal listed according to multiplicity as zeros of the characteristic equation.

Proof: The theorem is clearly true if A is a 1×1 matrix. Just let $U = 1$ the 1×1 matrix which has 1 down the main diagonal and zeros elsewhere. Suppose it is true for $(n-1) \times (n-1)$ matrices and let A be an $n \times n$ matrix. Then let $\mathbf{v}_1$ be a unit eigenvector for A . Then there exists λ_1 such that

$$A\mathbf{v}_1 = \lambda_1 \mathbf{v}_1, \quad |\mathbf{v}_1| = 1.$$

By Theorem C.2.24 there exists $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, an orthonormal set in $\mathbb{R}^n$. Let U_0 be a matrix whose i^{th} column is $\mathbf{v}_i$. Then from the above, it follows U_0 is orthogonal. Then from the way you multiply matrices $U_0^T A U_0$ is of the form

$$\begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}$$

where A_1 is an $(n-1) \times (n-1)$ matrix. The above matrix is similar to A so it has the same eigenvalues and indeed the same characteristic equation. Also the eigenvalues of A_1 are all real because each of these eigenvalues is an eigenvalue of the above matrix and is therefore an eigenvalue of A . Now by induction there exists an $(n-1) \times (n-1)$ orthogonal matrix $\tilde{U}_1$ such that

$$\tilde{U}_1^* A_1 \tilde{U}_1 = T_{n-1},$$

an upper triangular matrix. Consider

$$U_1 \equiv \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \tilde{U}_1 \end{pmatrix}$$

This is a orthogonal matrix and

$$\begin{aligned} U_1^T U_0^T A U_0 U_1 &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \tilde{U}_1^* \end{pmatrix} \begin{pmatrix} \lambda_1 & * \\ \mathbf{0} & A_1 \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \tilde{U}_1 \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & * \\ \mathbf{0} & T_{n-1} \end{pmatrix} \equiv T \end{aligned}$$

where T is upper triangular. Then let $U = U_0 U_1$. Since $(U_0 U_1)^T = U_1^T U_0^T$, it follows A is similar to T and that $U_0 U_1$ is orthogonal. Hence A and T have the same characteristic polynomials and since the eigenvalues of T are the diagonal entries listed according to algebraic multiplicity, this proves the theorem.

C.2.14 Symmetric Matrices

Recall a real matrix A is symmetric if $A = A^T$.

Lemma C.2.27 *A real symmetric matrix has all real eigenvalues.*

Proof: Recall the eigenvalues are solutions λ to

$$\det(\lambda I - A) = 0$$

and so by Theorem C.2.20, there exists $\mathbf{x}$ a vector such that

$$A\mathbf{x} = \lambda\mathbf{x}, \mathbf{x} \neq \mathbf{0}$$

Of course if A is real, it is still possible that the eigenvalue could be complex and if this is the case, then the vector $\mathbf{x}$ will also end up being complex. I wish to show that the eigenvalues are all real. Suppose then that λ is an eigenvalue and let $\mathbf{x}$ be the corresponding eigenvector described above. Then letting $\bar{\mathbf{x}}$ denote the complex conjugate of $\mathbf{x}$,

$$\lambda \mathbf{x}^T \bar{\mathbf{x}} = (A\mathbf{x})^T \bar{\mathbf{x}} = \mathbf{x}^T A^T \bar{\mathbf{x}} = \mathbf{x}^T A \bar{\mathbf{x}} = \mathbf{x}^T \overline{A\mathbf{x}} = \mathbf{x}^T \bar{\mathbf{x}} \bar{\lambda}$$

and so, canceling $\mathbf{x}^T \bar{\mathbf{x}}$, it follows $\lambda = \bar{\lambda}$ showing λ is real. This proves the lemma.

Theorem C.2.28 *Let A be a real symmetric matrix. Then there exists a diagonal matrix D consisting of the eigenvalues of A down the main diagonal and an orthogonal matrix U such that*

$$U^T A U = D.$$

Proof: Since A has all real eigenvalues, it follows from Theorem C.2.26, there exists an orthogonal matrix U such that

$$U^T A U = T$$

where T is upper triangular. Now

$$T^T = U^T A^T U = U^T A U = T$$

and so in fact T is a diagonal matrix having the eigenvalues of A down the diagonal. This proves the theorem.

Theorem C.2.29 *Let A be a real symmetric matrix which has all positive eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. Then*

$$(A\mathbf{x} \cdot \mathbf{x}) \equiv \mathbf{x}^T A \mathbf{x} \geq \lambda_1 |\mathbf{x}|^2$$

Proof: Let U be the orthogonal matrix of Theorem C.2.28. Then

$$\begin{aligned} (A\mathbf{x} \cdot \mathbf{x}) &= \mathbf{x}^T A \mathbf{x} = (\mathbf{x}^T U) D (U^T \mathbf{x}) \\ &= (U^T \mathbf{x}) D (U^T \mathbf{x}) = \sum_i \lambda_i |(U^T \mathbf{x})_i|^2 \\ &\geq \lambda_1 \sum_i |(U^T \mathbf{x})_i|^2 = \lambda_1 (U^T \mathbf{x} \cdot U^T \mathbf{x}) \\ &= \lambda_1 (U^T \mathbf{x})^T U^T \mathbf{x} = \lambda_1 \mathbf{x}^T U U^T \mathbf{x} = \lambda_1 \mathbf{x}^T I \mathbf{x} = \lambda_1 |\mathbf{x}|^2. \end{aligned}$$

C.3 Exercises

1. Let $m < n$ and let A be an $m \times n$ matrix. Show that A is **not** one to one. **Hint:** Consider the $n \times n$ matrix, A_1 which is of the form

$$A_1 \equiv \begin{pmatrix} A \\ 0 \end{pmatrix}$$

where the 0 denotes an $(n - m) \times n$ matrix of zeros. Thus $\det A_1 = 0$ and so A_1 is not one to one. Now observe that $A_1 \mathbf{x}$ is the vector,

$$A_1 \mathbf{x} = \begin{pmatrix} A\mathbf{x} \\ \mathbf{0} \end{pmatrix}$$

which equals zero if and only if $A\mathbf{x} = \mathbf{0}$.

2. Show that matrix multiplication is associative. That is, $(AB)C = A(BC)$.
3. Show that the inverse of a matrix, if it exists, is unique. Thus if $AB = BA = I$, then $B = A^{-1}$.
4. In the proof of Theorem C.2.14 it was claimed that $\det(I) = 1$. Here $I = (\delta_{ij})$. Prove this assertion. Also prove Corollary C.2.17.
5. Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be vectors in $\mathbb{F}^n$ and let $M(\mathbf{v}_1, \dots, \mathbf{v}_n)$ denote the matrix whose i^{th} column equals $\mathbf{v}_i$. Define

$$d(\mathbf{v}_1, \dots, \mathbf{v}_n) \equiv \det(M(\mathbf{v}_1, \dots, \mathbf{v}_n)).$$

Prove that d is linear in each variable, (multilinear), that

$$d(\mathbf{v}_1, \dots, \mathbf{v}_i, \dots, \mathbf{v}_j, \dots, \mathbf{v}_n) = -d(\mathbf{v}_1, \dots, \mathbf{v}_j, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n), \quad (3.16)$$

and

$$d(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1 \quad (3.17)$$

where here $\mathbf{e}_j$ is the vector in $\mathbb{F}^n$ which has a zero in every position except the j^{th} position in which it has a one.

6. Suppose $f : \mathbb{F}^n \times \dots \times \mathbb{F}^n \rightarrow \mathbb{F}$ satisfies 3.16 and 3.17 and is linear in each variable. Show that $f = d$.
7. Show that if you replace a row (column) of an $n \times n$ matrix A with itself added to some multiple of another row (column) then the new matrix has the same determinant as the original one.
8. If $A = (a_{ij})$, show that $\det(A) = \sum_{(k_1, \dots, k_n)} \text{sgn}(k_1, \dots, k_n) a_{k_1 1} \dots a_{k_n n}$.
9. Use the result of Problem 7 to evaluate by hand the determinant

$$\det \begin{pmatrix} 1 & 2 & 3 & 2 \\ -6 & 3 & 2 & 3 \\ 5 & 2 & 2 & 3 \\ 3 & 4 & 6 & 4 \end{pmatrix}.$$

10. Find the inverse if it exists of the matrix,

$$\begin{pmatrix} e^t & \cos t & \sin t \\ e^t & -\sin t & \cos t \\ e^t & -\cos t & -\sin t \end{pmatrix}.$$

11. Let $Ly = y^{(n)} + a_{n-1}(x)y^{(n-1)} + \cdots + a_1(x)y' + a_0(x)y$ where the a_i are given continuous functions defined on a closed interval, (a, b) and y is some function which has n derivatives so it makes sense to write Ly . Suppose $Ly_k = 0$ for $k = 1, 2, \dots, n$. The Wronskian of these functions, y_i is defined as

$$W(y_1, \dots, y_n)(x) \equiv \det \begin{pmatrix} y_1(x) & \cdots & y_n(x) \\ y_1'(x) & \cdots & y_n'(x) \\ \vdots & & \vdots \\ y_1^{(n-1)}(x) & \cdots & y_n^{(n-1)}(x) \end{pmatrix}$$

Show that for $W(x) = W(y_1, \dots, y_n)(x)$ to save space,

$$W'(x) = \det \begin{pmatrix} y_1(x) & \cdots & y_n(x) \\ y_1'(x) & \cdots & y_n'(x) \\ \vdots & & \vdots \\ y_1^{(n)}(x) & \cdots & y_n^{(n)}(x) \end{pmatrix}.$$

Now use the differential equation, $Ly = 0$ which is satisfied by each of these functions, y_i and properties of determinants presented above to verify that $W' + a_{n-1}(x)W = 0$. Give an explicit solution of this linear differential equation, Abel's formula, and use your answer to verify that the Wronskian of these solutions to the equation, $Ly = 0$ either vanishes identically on (a, b) or never.

12. Two $n \times n$ matrices, A and B , are similar if $B = S^{-1}AS$ for some invertible $n \times n$ matrix, S . Show that if two matrices are similar, they have the same characteristic polynomials.
13. Suppose the characteristic polynomial of an $n \times n$ matrix, A is of the form

$$t^n + a_{n-1}t^{n-1} + \cdots + a_1t + a_0$$

and that $a_0 \neq 0$. Find a formula A^{-1} in terms of powers of the matrix, A . Show that A^{-1} exists if and only if $a_0 \neq 0$.

14. In constitutive modelling of the stress and strain tensors, one sometimes considers sums of the form $\sum_{k=0}^{\infty} a_k A^k$ where A is a 3×3 matrix. Show using the Cayley Hamilton theorem that if such a thing makes any sense, you can always obtain it as a finite sum having no more than n terms.

Implicit Function Theorem*



The implicit function theorem is one of the greatest theorems in mathematics. There are many versions of this theorem which are of far greater generality than the one given here. The proof given here is like one found in one of Caratheodory's books on the calculus of variations. It is not as elegant as some of the others which are based on a contraction mapping principle but it may be more accessible. However, it is an advanced topic. Don't waste your time with it unless you have first read and understood the material on rank and determinants found in the chapter on the mathematical theory of determinants. You will also need to use the extreme value theorem for a function of n variables and the chain rule as well as everything about matrix multiplication.

Definition D.1.1 Suppose U is an open set in $\mathbb{R}^n \times \mathbb{R}^m$ and $(\mathbf{x}, \mathbf{y})$ will denote a typical point of $\mathbb{R}^n \times \mathbb{R}^m$ with $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. Let $\mathbf{f} : U \rightarrow \mathbb{R}^p$ be in $C^1(U)$. Then define

$$\begin{aligned} D_1 \mathbf{f}(\mathbf{x}, \mathbf{y}) &\equiv \begin{pmatrix} f_{1,x_1}(\mathbf{x}, \mathbf{y}) & \cdots & f_{1,x_n}(\mathbf{x}, \mathbf{y}) \\ \vdots & & \vdots \\ f_{p,x_1}(\mathbf{x}, \mathbf{y}) & \cdots & f_{p,x_n}(\mathbf{x}, \mathbf{y}) \end{pmatrix}, \\ D_2 \mathbf{f}(\mathbf{x}, \mathbf{y}) &\equiv \begin{pmatrix} f_{1,y_1}(\mathbf{x}, \mathbf{y}) & \cdots & f_{1,y_m}(\mathbf{x}, \mathbf{y}) \\ \vdots & & \vdots \\ f_{p,y_1}(\mathbf{x}, \mathbf{y}) & \cdots & f_{p,y_m}(\mathbf{x}, \mathbf{y}) \end{pmatrix}. \end{aligned}$$

Thus $D\mathbf{f}(\mathbf{x}, \mathbf{y})$ is a $p \times (n + m)$ matrix of the form

$$D\mathbf{f}(\mathbf{x}, \mathbf{y}) = \left(D_1 \mathbf{f}(\mathbf{x}, \mathbf{y}) \mid D_2 \mathbf{f}(\mathbf{x}, \mathbf{y}) \right).$$

Note that $D_1 \mathbf{f}(\mathbf{x}, \mathbf{y})$ is an $p \times n$ matrix and $D_2 \mathbf{f}(\mathbf{x}, \mathbf{y})$ is a $p \times m$ matrix.

Theorem D.1.2 (implicit function theorem) Suppose U is an open set in $\mathbb{R}^n \times \mathbb{R}^m$. Let $\mathbf{f} : U \rightarrow \mathbb{R}^n$ be in $C^1(U)$ and suppose

$$\mathbf{f}(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}, \quad D_1 \mathbf{f}(\mathbf{x}_0, \mathbf{y}_0)^{-1} \text{ exists.} \quad (4.1)$$

Then there exist positive constants, δ, η , such that for every $\mathbf{y} \in B(\mathbf{y}_0, \eta)$ there exists a unique $\mathbf{x}(\mathbf{y}) \in B(\mathbf{x}_0, \delta)$ such that

$$\mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) = \mathbf{0}. \quad (4.2)$$

Furthermore, the mapping, $\mathbf{y} \rightarrow \mathbf{x}(\mathbf{y})$ is in $C^1(B(\mathbf{y}_0, \eta))$.

Proof: Let

$$\mathbf{f}(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} f_1(\mathbf{x}, \mathbf{y}) \\ f_2(\mathbf{x}, \mathbf{y}) \\ \vdots \\ f_n(\mathbf{x}, \mathbf{y}) \end{pmatrix}.$$

Define for $(\mathbf{x}^1, \dots, \mathbf{x}^n) \in \overline{B(\mathbf{x}_0, \delta)}^n$ and $\mathbf{y} \in B(\mathbf{y}_0, \eta)$ the following matrix.

$$J(\mathbf{x}^1, \dots, \mathbf{x}^n, \mathbf{y}) \equiv \begin{pmatrix} f_{1,x_1}(\mathbf{x}^1, \mathbf{y}) & \cdots & f_{1,x_n}(\mathbf{x}^1, \mathbf{y}) \\ \vdots & & \vdots \\ f_{n,x_1}(\mathbf{x}^n, \mathbf{y}) & \cdots & f_{n,x_n}(\mathbf{x}^n, \mathbf{y}) \end{pmatrix}.$$

Then by the assumption of continuity of all the partial derivatives and the extreme value theorem, there exists $r > 0$ and $\delta_0, \eta_0 > 0$ such that if $\delta \leq \delta_0$ and $\eta \leq \eta_0$, it follows that for all $(\mathbf{x}^1, \dots, \mathbf{x}^n) \in \overline{B(\mathbf{x}_0, \delta)}^n$ and $\mathbf{y} \in \overline{B(\mathbf{y}_0, \eta)}$,

$$\det(J(\mathbf{x}^1, \dots, \mathbf{x}^n, \mathbf{y})) > r > 0. \quad (4.3)$$

and $\overline{B(\mathbf{x}_0, \delta_0)} \times \overline{B(\mathbf{y}_0, \eta_0)} \subseteq U$. By continuity of all the partial derivatives and the extreme value theorem, it can also be assumed there exists a constant K such that for all $(\mathbf{x}, \mathbf{y}) \in \overline{B(\mathbf{x}_0, \delta_0)} \times \overline{B(\mathbf{y}_0, \eta_0)}$ and $i = 1, 2, \dots, n$, the i^{th} row of $D_2 \mathbf{f}(\mathbf{x}, \mathbf{y})$, given by $D_2 f_i(\mathbf{x}, \mathbf{y})$ satisfies

$$|D_2 f_i(\mathbf{x}, \mathbf{y})| < K, \quad (4.4)$$

and for all $(\mathbf{x}^1, \dots, \mathbf{x}^n) \in \overline{B(\mathbf{x}_0, \delta_0)}^n$ and $\mathbf{y} \in \overline{B(\mathbf{y}_0, \eta_0)}$ the i^{th} row of the matrix,

$$J(\mathbf{x}^1, \dots, \mathbf{x}^n, \mathbf{y})^{-1}$$

which equals $\mathbf{e}_i^T (J(\mathbf{x}^1, \dots, \mathbf{x}^n, \mathbf{y})^{-1})$ satisfies

$$\left| \mathbf{e}_i^T (J(\mathbf{x}^1, \dots, \mathbf{x}^n, \mathbf{y})^{-1}) \right| < K. \quad (4.5)$$

(Recall that $\mathbf{e}_i$ is the column vector consisting of all zeros except for a 1 in the i^{th} position.)

To begin with, it is shown that for a given $\mathbf{y} \in B(\mathbf{y}_0, \eta)$ there is at most one $\mathbf{x} \in B(\mathbf{x}_0, \delta)$ such that $\mathbf{f}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$.

Pick $\mathbf{y} \in B(\mathbf{y}_0, \eta)$ and suppose there exist $\mathbf{x}, \mathbf{z} \in \overline{B(\mathbf{x}_0, \delta)}$ such that $\mathbf{f}(\mathbf{x}, \mathbf{y}) = \mathbf{f}(\mathbf{z}, \mathbf{y}) = \mathbf{0}$. Consider f_i and let

$$h(t) \equiv f_i(\mathbf{x} + t(\mathbf{z} - \mathbf{x}), \mathbf{y}).$$

Then $h(1) = h(0)$ and so by Rolle's theorem, $h'(t_i) = 0$ for some $t_i \in (0, 1)$. Therefore, from the chain rule and for this value of t_i ,

$$h'(t_i) = Df_i(\mathbf{x} + t_i(\mathbf{z} - \mathbf{x}), \mathbf{y})(\mathbf{z} - \mathbf{x}) = 0. \quad (4.6)$$

Then denote by $\mathbf{x}^i$ the vector $\mathbf{x} + t_i(\mathbf{z} - \mathbf{x})$. It follows from 4.6 that

$$J(\mathbf{x}^1, \dots, \mathbf{x}^n, \mathbf{y})(\mathbf{z} - \mathbf{x}) = \mathbf{0}$$

and so from 4.3 $\mathbf{z} - \mathbf{x} = \mathbf{0}$. (The matrix, in the above is invertible since its determinant is nonzero.) Now it will be shown that if η is chosen sufficiently small, then for all $\mathbf{y} \in B(\mathbf{y}_0, \eta)$, there exists a unique $\mathbf{x}(\mathbf{y}) \in B(\mathbf{x}_0, \delta)$ such that $\mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) = \mathbf{0}$.

Claim: If η is small enough, then the function $h_{\mathbf{y}}(\mathbf{x}) \equiv |\mathbf{f}(\mathbf{x}, \mathbf{y})|^2$ achieves its minimum value on $\overline{B(\mathbf{x}_0, \delta)}$ at a point of $B(\mathbf{x}_0, \delta)$. (The existence of a point in $\overline{B(\mathbf{x}_0, \delta)}$ at which $h_{\mathbf{y}}$ achieves its minimum follows from the extreme value theorem.)

Proof of claim: Suppose this is not the case. Then there exists a sequence $\eta_k \rightarrow 0$ and for some $\mathbf{y}_k$ having $|\mathbf{y}_k - \mathbf{y}_0| < \eta_k$, the minimum of $h_{\mathbf{y}_k}$ on $\overline{B(\mathbf{x}_0, \delta)}$ occurs on a point of $\overline{B(\mathbf{x}_0, \delta)}$, $\mathbf{x}_k$ such that $|\mathbf{x}_0 - \mathbf{x}_k| = \delta$. Now taking a subsequence, still denoted by k , it can be assumed that $\mathbf{x}_k \rightarrow \mathbf{x}$ with $|\mathbf{x} - \mathbf{x}_0| = \delta$ and $\mathbf{y}_k \rightarrow \mathbf{y}_0$. This follows from the fact that $\{\mathbf{x} \in \overline{B(\mathbf{x}_0, \delta)} : |\mathbf{x} - \mathbf{x}_0| = \delta\}$ is a closed and bounded set and is therefore sequentially compact. Let $\varepsilon > 0$. Then for k large enough, the continuity of $\mathbf{y} \rightarrow h_{\mathbf{y}}(\mathbf{x}_0)$ implies $h_{\mathbf{y}_k}(\mathbf{x}_0) < \varepsilon$ because $h_{\mathbf{y}_0}(\mathbf{x}_0) = 0$ since $\mathbf{f}(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$. Therefore, from the definition of $\mathbf{x}_k$, it is also the case that $h_{\mathbf{y}_k}(\mathbf{x}_k) < \varepsilon$. Passing to the limit yields $h_{\mathbf{y}_0}(\mathbf{x}) \leq \varepsilon$. Since $\varepsilon > 0$ is arbitrary, it follows that $h_{\mathbf{y}_0}(\mathbf{x}) = 0$ which contradicts the first part of the argument in which it was shown that for $\mathbf{y} \in B(\mathbf{y}_0, \eta)$ there is at most one point $\mathbf{x}$ of $\overline{B(\mathbf{x}_0, \delta)}$ where $\mathbf{f}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$. Here two have been obtained, $\mathbf{x}_0$ and $\mathbf{x}$. This proves the claim.

Choose $\eta < \eta_0$ and also small enough that the above claim holds and let $\mathbf{x}(\mathbf{y})$ denote a point of $B(\mathbf{x}_0, \delta)$ at which the minimum of $h_{\mathbf{y}}$ on $\overline{B(\mathbf{x}_0, \delta)}$ is achieved. Since $\mathbf{x}(\mathbf{y})$ is an interior point, you can consider $h_{\mathbf{y}}(\mathbf{x}(\mathbf{y}) + t\mathbf{v})$ for $|t|$ small and conclude this function of t has a zero derivative at $t = 0$. Now

$$h_{\mathbf{y}}(\mathbf{x}(\mathbf{y}) + t\mathbf{v}) = \sum_{i=1}^n f_i^2(\mathbf{x}(\mathbf{y}) + t\mathbf{v}, \mathbf{y})$$

and so from the chain rule,

$$\frac{d}{dt} h_{\mathbf{y}}(\mathbf{x}(\mathbf{y}) + t\mathbf{v}) = \sum_{i=1}^n 2f_i(\mathbf{x}(\mathbf{y}) + t\mathbf{v}, \mathbf{y}) \frac{\partial f_i(\mathbf{x}(\mathbf{y}) + t\mathbf{v}, \mathbf{y})}{\partial x_j} v_j.$$

Therefore, letting $t = 0$, it is required that for every $\mathbf{v}$,

$$\sum_{i=1}^n 2f_i(\mathbf{x}(\mathbf{y}), \mathbf{y}) \frac{\partial f_i(\mathbf{x}(\mathbf{y}), \mathbf{y})}{\partial x_j} v_j = 0.$$

In terms of matrices this reduces to

$$0 = 2\mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y})^T D_1 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) \mathbf{v}$$

for every vector $\mathbf{v}$. Therefore,

$$\mathbf{0} = \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y})^T D_1 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y})$$

From 4.3, it follows $\mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) = \mathbf{0}$. This proves the existence of the function $\mathbf{y} \rightarrow \mathbf{x}(\mathbf{y})$ such that $\mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) = \mathbf{0}$ for all $\mathbf{y} \in B(\mathbf{y}_0, \eta)$.

It remains to verify this function is a C^1 function. To do this, let $\mathbf{y}_1$ and $\mathbf{y}_2$ be points of $B(\mathbf{y}_0, \eta)$. Then as before, consider the i^{th} component of $\mathbf{f}$ and consider the same argument using the mean value theorem to write

$$\begin{aligned} 0 &= f_i(\mathbf{x}(\mathbf{y}_1), \mathbf{y}_1) - f_i(\mathbf{x}(\mathbf{y}_2), \mathbf{y}_2) \\ &= f_i(\mathbf{x}(\mathbf{y}_1), \mathbf{y}_1) - f_i(\mathbf{x}(\mathbf{y}_2), \mathbf{y}_1) + f_i(\mathbf{x}(\mathbf{y}_2), \mathbf{y}_1) - f_i(\mathbf{x}(\mathbf{y}_2), \mathbf{y}_2) \\ &= D_1 f_i(\mathbf{x}^i, \mathbf{y}_1)(\mathbf{x}(\mathbf{y}_1) - \mathbf{x}(\mathbf{y}_2)) + D_2 f_i(\mathbf{x}(\mathbf{y}_2), \mathbf{y}^i)(\mathbf{y}_1 - \mathbf{y}_2). \end{aligned} \quad (4.7)$$

where $\mathbf{y}^i$ is a point on the line segment joining $\mathbf{y}_1$ and $\mathbf{y}_2$. Thus from 4.4 and the Cauchy Schwarz inequality,

$$|D_2 f_i(\mathbf{x}(\mathbf{y}_2), \mathbf{y}^i)(\mathbf{y}_1 - \mathbf{y}_2)| \leq K |\mathbf{y}_1 - \mathbf{y}_2|.$$

Therefore, letting $M(\mathbf{y}^1, \dots, \mathbf{y}^n) \equiv M$ denote the matrix having the i^{th} row equal to $D_2 f_i(\mathbf{x}(\mathbf{y}_2), \mathbf{y}^i)$, it follows

$$|M(\mathbf{y}_1 - \mathbf{y}_2)| \leq \left(\sum_i K^2 |\mathbf{y}_1 - \mathbf{y}_2|^2 \right)^{1/2} = \sqrt{m} K |\mathbf{y}_1 - \mathbf{y}_2|. \quad (4.8)$$

Also, from 4.7,

$$J(\mathbf{x}^1, \dots, \mathbf{x}^n, \mathbf{y}_1)(\mathbf{x}(\mathbf{y}_1) - \mathbf{x}(\mathbf{y}_2)) = -M(\mathbf{y}_1 - \mathbf{y}_2) \quad (4.9)$$

and so from 4.8 and 4.10,

$$|\mathbf{x}(\mathbf{y}_1) - \mathbf{x}(\mathbf{y}_2)| = \left| J(\mathbf{x}^1, \dots, \mathbf{x}^n, \mathbf{y}_1)^{-1} M(\mathbf{y}_1 - \mathbf{y}_2) \right| \quad (4.10)$$

$$\begin{aligned} &= \left(\sum_{i=1}^n \left| \mathbf{e}_i^T J(\mathbf{x}^1, \dots, \mathbf{x}^n, \mathbf{y}_1)^{-1} M(\mathbf{y}_1 - \mathbf{y}_2) \right|^2 \right)^{1/2} \\ &\leq \left(\sum_{i=1}^n K^2 |M(\mathbf{y}_1 - \mathbf{y}_2)|^2 \right)^{1/2} \leq \left(\sum_{i=1}^n K^2 (\sqrt{m} K |\mathbf{y}_1 - \mathbf{y}_2|)^2 \right)^{1/2} \\ &= K^2 \sqrt{mn} |\mathbf{y}_1 - \mathbf{y}_2| \end{aligned} \quad (4.11)$$

It follows as in the proof of the chain rule that

$$\mathbf{o}(\mathbf{x}(\mathbf{y} + \mathbf{v}) - \mathbf{x}(\mathbf{y})) = \mathbf{o}(\mathbf{v}). \quad (4.12)$$

Now let $\mathbf{y} \in B(\mathbf{y}_0, \eta)$ and let $|\mathbf{v}|$ be sufficiently small that $\mathbf{y} + \mathbf{v} \in B(\mathbf{y}_0, \eta)$. Then

$$\begin{aligned} \mathbf{0} &= \mathbf{f}(\mathbf{x}(\mathbf{y} + \mathbf{v}), \mathbf{y} + \mathbf{v}) - \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) \\ &= \mathbf{f}(\mathbf{x}(\mathbf{y} + \mathbf{v}), \mathbf{y} + \mathbf{v}) - \mathbf{f}(\mathbf{x}(\mathbf{y} + \mathbf{v}), \mathbf{y}) + \mathbf{f}(\mathbf{x}(\mathbf{y} + \mathbf{v}), \mathbf{y}) - \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) \\ &= D_2 \mathbf{f}(\mathbf{x}(\mathbf{y} + \mathbf{v}), \mathbf{y}) \mathbf{v} + D_1 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y})(\mathbf{x}(\mathbf{y} + \mathbf{v}) - \mathbf{x}(\mathbf{y})) + \mathbf{o}(|\mathbf{x}(\mathbf{y} + \mathbf{v}) - \mathbf{x}(\mathbf{y})|) \\ &= D_2 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) \mathbf{v} + D_1 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y})(\mathbf{x}(\mathbf{y} + \mathbf{v}) - \mathbf{x}(\mathbf{y})) + \\ &\quad \mathbf{o}(|\mathbf{x}(\mathbf{y} + \mathbf{v}) - \mathbf{x}(\mathbf{y})|) + (D_2 \mathbf{f}(\mathbf{x}(\mathbf{y} + \mathbf{v}), \mathbf{y}) \mathbf{v} - D_2 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) \mathbf{v}) \\ &= D_2 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) \mathbf{v} + D_1 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y})(\mathbf{x}(\mathbf{y} + \mathbf{v}) - \mathbf{x}(\mathbf{y})) + \mathbf{o}(\mathbf{v}). \end{aligned}$$

Therefore,

$$\mathbf{x}(\mathbf{y} + \mathbf{v}) - \mathbf{x}(\mathbf{y}) = -D_1 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y})^{-1} D_2 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) \mathbf{v} + \mathbf{o}(\mathbf{v})$$

which shows that $D\mathbf{x}(\mathbf{y}) = -D_1 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y})^{-1} D_2 \mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y})$ and $\mathbf{y} \rightarrow D\mathbf{x}(\mathbf{y})$ is continuous. ■

D.2 More Continuous Partial Derivatives

The implicit function theorem will now be improved slightly. If $\mathbf{f}$ is C^k , it follows that the function which is implicitly defined is also C^k , not just C^1 , meaning all mixed partial

derivatives of $\mathbf{f}$ up to order k are continuous. Since the inverse function theorem comes as a case of the implicit function theorem, this shows that the inverse function also inherits the property of being C^k . First some notation is convenient. Let $\alpha = (\alpha_1, \dots, \alpha_n)$ where each α_i is a nonnegative integer. Then letting $|\alpha| = \sum_i \alpha_i$,

$$D^\alpha \mathbf{f}(\mathbf{x}) \equiv \frac{\partial^{|\alpha|} \mathbf{f}}{\partial \alpha_1 \partial \alpha_2 \dots \partial \alpha_n}(\mathbf{x}), \quad D^0 \mathbf{f}(\mathbf{x}) \equiv \mathbf{f}(\mathbf{x})$$

Theorem D.2.1 (*implicit function theorem*) Suppose U is an open set in $\mathbb{F}^n \times \mathbb{F}^m$. Let $\mathbf{f} : U \rightarrow \mathbb{F}^n$ be in $C^k(U)$ and suppose

$$\mathbf{f}(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}, \quad D_1 \mathbf{f}(\mathbf{x}_0, \mathbf{y}_0)^{-1} \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^n). \quad (4.13)$$

Then there exist positive constants, δ, η , such that for every $\mathbf{y} \in B(\mathbf{y}_0, \eta)$ there exists a unique $\mathbf{x}(\mathbf{y}) \in B(\mathbf{x}_0, \delta)$ such that

$$\mathbf{f}(\mathbf{x}(\mathbf{y}), \mathbf{y}) = \mathbf{0}. \quad (4.14)$$

Furthermore, the mapping, $\mathbf{y} \rightarrow \mathbf{x}(\mathbf{y})$ is in $C^k(B(\mathbf{y}_0, \eta))$.

Proof: From the implicit function theorem $\mathbf{y} \rightarrow \mathbf{x}(\mathbf{y})$ is C^1 . It remains to show that it is C^k for $k > 1$ assuming that $\mathbf{f}$ is C^k . From 4.14

$$\frac{\partial \mathbf{x}}{\partial y^l} = -D_1 \mathbf{f}(\mathbf{x}, \mathbf{y})^{-1} \frac{\partial \mathbf{f}}{\partial y^l}.$$

Thus the following formula holds for $q = 1$ and $|\alpha| = q$.

$$D^\alpha \mathbf{x}(\mathbf{y}) = \sum_{|\beta| \leq q} M_\beta(\mathbf{x}, \mathbf{y}) D^\beta \mathbf{f}(\mathbf{x}, \mathbf{y}) \quad (4.15)$$

where M_β is a matrix whose entries are differentiable functions of $D^\gamma \mathbf{x}$ for $|\gamma| < q$ and $D^\tau \mathbf{f}(\mathbf{x}, \mathbf{y})$ for $|\tau| \leq q$. This follows easily from the description of $D_1 \mathbf{f}(\mathbf{x}, \mathbf{y})^{-1}$ in terms of the cofactor matrix and the determinant of $D_1 \mathbf{f}(\mathbf{x}, \mathbf{y})$. Suppose 4.15 holds for $|\alpha| = q < k$. Then by induction, this yields $\mathbf{x}$ is C^q . Then

$$\frac{\partial D^\alpha \mathbf{x}(\mathbf{y})}{\partial y^p} = \sum_{|\beta| \leq |\alpha|} \frac{\partial M_\beta(\mathbf{x}, \mathbf{y})}{\partial y^p} D^\beta \mathbf{f}(\mathbf{x}, \mathbf{y}) + M_\beta(\mathbf{x}, \mathbf{y}) \frac{\partial D^\beta \mathbf{f}(\mathbf{x}, \mathbf{y})}{\partial y^p}.$$

By the chain rule $\frac{\partial M_\beta(\mathbf{x}, \mathbf{y})}{\partial y^p}$ is a matrix whose entries are differentiable functions of $D^\tau \mathbf{f}(\mathbf{x}, \mathbf{y})$ for $|\tau| \leq q + 1$ and $D^\gamma \mathbf{x}$ for $|\gamma| < q + 1$. It follows, since y^p was arbitrary, that for any $|\alpha| = q + 1$, a formula like 4.15 holds with q being replaced by $q + 1$. By induction, $\mathbf{x}$ is C^k . ■

As a simple corollary, this yields the inverse function theorem.

Theorem D.2.2 (*inverse function theorem*) Let $\mathbf{x}_0 \in U \subseteq \mathbb{F}^n$ and let $\mathbf{f} : U \rightarrow \mathbb{F}^n$. Suppose for k a positive integer,

$$\mathbf{f} \text{ is } C^k(U), \quad \text{and} \quad D\mathbf{f}(\mathbf{x}_0)^{-1} \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^n). \quad (4.16)$$

Then there exist open sets, W , and V such that

$$\mathbf{x}_0 \in W \subseteq U, \quad (4.17)$$

$$\mathbf{f} : W \rightarrow V \text{ is one to one and onto,} \quad (4.18)$$

$$\mathbf{f}^{-1} \text{ is } C^k. \quad (4.19)$$

D.3 The Method Of Lagrange Multipliers

As an application of the implicit function theorem, consider the method of Lagrange multipliers. Recall the problem is to maximize or minimize a function subject to equality constraints. Let $f : U \rightarrow \mathbb{R}$ be a C^1 function where $U \subseteq \mathbb{R}^n$ and let

$$g_i(\mathbf{x}) = 0, \quad i = 1, \dots, m \quad (4.20)$$

be a collection of equality constraints with $m < n$. Now consider the system of nonlinear equations

$$\begin{aligned} f(\mathbf{x}) &= a \\ g_i(\mathbf{x}) &= 0, \quad i = 1, \dots, m. \end{aligned}$$

Recall $\mathbf{x}_0$ is a local maximum if $f(\mathbf{x}_0) \geq f(\mathbf{x})$ for all $\mathbf{x}$ near $\mathbf{x}_0$ which also satisfies the constraints 4.20. A local minimum is defined similarly. Let $\mathbf{F} : U \times \mathbb{R} \rightarrow \mathbb{R}^{m+1}$ be defined by

$$\mathbf{F}(\mathbf{x}, a) \equiv \begin{pmatrix} f(\mathbf{x}) - a \\ g_1(\mathbf{x}) \\ \vdots \\ g_m(\mathbf{x}) \end{pmatrix}. \quad (4.21)$$

Now consider the $m+1 \times n$ matrix,

$$\begin{pmatrix} f_{x_1}(\mathbf{x}_0) & \cdots & f_{x_n}(\mathbf{x}_0) \\ g_{1x_1}(\mathbf{x}_0) & \cdots & g_{1x_n}(\mathbf{x}_0) \\ \vdots & & \vdots \\ g_{mx_1}(\mathbf{x}_0) & \cdots & g_{mx_n}(\mathbf{x}_0) \end{pmatrix}.$$

If this matrix has rank $m+1$ then some $m+1 \times m+1$ submatrix has nonzero determinant. It follows from the implicit function theorem there exists $m+1$ variables, $x_{i_1}, \dots, x_{i_{m+1}}$ such that the system

$$\mathbf{F}(\mathbf{x}, a) = \mathbf{0} \quad (4.22)$$

specifies these $m+1$ variables as a function of the remaining $n - (m+1)$ variables and a in an open set of $\mathbb{R}^{n-m}$. Thus there is a solution $(\mathbf{x}, a)$ to 4.22 for some $\mathbf{x}$ close to $\mathbf{x}_0$ whenever a is in some open interval. Therefore, $\mathbf{x}_0$ cannot be either a local minimum or a local maximum. It follows that if $\mathbf{x}_0$ is either a local maximum or a local minimum, then the above matrix must have rank less than $m+1$ which requires the rows to be linearly dependent. Thus, there exist m scalars,

$$\lambda_1, \dots, \lambda_m,$$

and a scalar μ , not all zero such that

$$\mu \begin{pmatrix} f_{x_1}(\mathbf{x}_0) \\ \vdots \\ f_{x_n}(\mathbf{x}_0) \end{pmatrix} = \lambda_1 \begin{pmatrix} g_{1x_1}(\mathbf{x}_0) \\ \vdots \\ g_{1x_n}(\mathbf{x}_0) \end{pmatrix} + \cdots + \lambda_m \begin{pmatrix} g_{mx_1}(\mathbf{x}_0) \\ \vdots \\ g_{mx_n}(\mathbf{x}_0) \end{pmatrix}. \quad (4.23)$$

If the column vectors

$$\begin{pmatrix} g_{1x_1}(\mathbf{x}_0) \\ \vdots \\ g_{1x_n}(\mathbf{x}_0) \end{pmatrix}, \dots, \begin{pmatrix} g_{mx_1}(\mathbf{x}_0) \\ \vdots \\ g_{mx_n}(\mathbf{x}_0) \end{pmatrix} \quad (4.24)$$

are linearly independent, then, $\mu \neq 0$ and dividing by μ yields an expression of the form

$$\begin{pmatrix} f_{x_1}(\mathbf{x}_0) \\ \vdots \\ f_{x_n}(\mathbf{x}_0) \end{pmatrix} = \lambda_1 \begin{pmatrix} g_{1x_1}(\mathbf{x}_0) \\ \vdots \\ g_{1x_n}(\mathbf{x}_0) \end{pmatrix} + \cdots + \lambda_m \begin{pmatrix} g_{mx_1}(\mathbf{x}_0) \\ \vdots \\ g_{mx_n}(\mathbf{x}_0) \end{pmatrix} \quad (4.25)$$

at every point $\mathbf{x}_0$ which is either a local maximum or a local minimum. This proves the following theorem.

Theorem D.3.1 *Let U be an open subset of $\mathbb{R}^n$ and let $f : U \rightarrow \mathbb{R}$ be a C^1 function. Then if $\mathbf{x}_0 \in U$ is either a local maximum or local minimum of f subject to the constraints 4.20, then 4.23 must hold for some scalars $\mu, \lambda_1, \dots, \lambda_m$ not all equal to zero. If the vectors in 4.24 are linearly independent, it follows that an equation of the form 4.25 holds.*

D.4 The Local Structure Of C^1 Mappings

In linear algebra it is shown that every invertible matrix can be written as a product of elementary matrices, those matrices which are obtained from doing a row operation to the identity matrix. Two of the row operations produce a matrix which will change exactly one entry of a vector when it is multiplied by the elementary matrix. The other row operation involves switching two rows and this has the effect of switching two entries in a vector when multiplied on the left by the elementary matrix. Thus, in terms of the effect on a vector, the mapping determined by the given matrix can be considered as a composition of mappings which either flip two entries of the vector or change exactly one. A similar local result is available for nonlinear mappings.

Definition D.4.1 *Let U be an open set in $\mathbb{R}^n$ and let $\mathbf{G} : U \rightarrow \mathbb{R}^n$. Then $\mathbf{G}$ is called primitive if it is of the form*

$$\mathbf{G}(\mathbf{x}) = \begin{pmatrix} x_1 & \cdots & \alpha(\mathbf{x}) & \cdots & x_n \end{pmatrix}^T.$$

Thus, $\mathbf{G}$ is primitive if it only changes one of the variables. A function $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is called a flip if

$$\mathbf{F}(x_1, \dots, x_k, \dots, x_l, \dots, x_n) = (x_1, \dots, x_l, \dots, x_k, \dots, x_n)^T.$$

Thus a function is a flip if it interchanges two coordinates. Also, for $m = 1, 2, \dots, n$, define

$$P_m(\mathbf{x}) \equiv \begin{pmatrix} x_1 & x_2 & \cdots & x_m & 0 & \cdots & 0 \end{pmatrix}^T$$

It turns out that if $\mathbf{h}(\mathbf{0}) = \mathbf{0}$, $D\mathbf{h}(\mathbf{0})^{-1}$ exists, and $\mathbf{h}$ is C^1 on U , then $\mathbf{h}$ can be written as a composition of primitive functions and flips. This is a very interesting application of the inverse function theorem.

Theorem D.4.2 *Let $\mathbf{h} : U \rightarrow \mathbb{R}^n$ be a C^1 function with $\mathbf{h}(\mathbf{0}) = \mathbf{0}$, $D\mathbf{h}(\mathbf{0})^{-1}$ exists. Then there is an open set $V \subseteq U$ containing $\mathbf{0}$, flips $\mathbf{F}_1, \dots, \mathbf{F}_{n-1}$, and primitive functions $\mathbf{G}_n, \mathbf{G}_{n-1}, \dots, \mathbf{G}_1$ such that for $\mathbf{x} \in V$,*

$$\mathbf{h}(\mathbf{x}) = \mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1} \circ \mathbf{G}_n \circ \mathbf{G}_{n-1} \circ \cdots \circ \mathbf{G}_1(\mathbf{x}).$$

The primitive function $\mathbf{G}_j$ leaves x_i unchanged for $i \neq j$.

Proof: Let

$$\mathbf{h}_1(\mathbf{x}) \equiv \mathbf{h}(\mathbf{x}) = \begin{pmatrix} \alpha_1(\mathbf{x}) & \cdots & \alpha_n(\mathbf{x}) \end{pmatrix}^T$$

$$D\mathbf{h}(\mathbf{0})\mathbf{e}_1 = \begin{pmatrix} \alpha_{1,1}(\mathbf{0}) & \cdots & \alpha_{n,1}(\mathbf{0}) \end{pmatrix}^T$$

where $\alpha_{k,1}$ denotes $\frac{\partial \alpha_k}{\partial x_1}$. Since $D\mathbf{h}(\mathbf{0})$ is one to one, the right side of this expression cannot be zero. Hence there exists some k such that $\alpha_{k,1}(\mathbf{0}) \neq 0$. Now define

$$\mathbf{G}_1(\mathbf{x}) \equiv \begin{pmatrix} \alpha_k(\mathbf{x}) & x_2 & \cdots & x_n \end{pmatrix}^T$$

Then the matrix of $D\mathbf{G}(\mathbf{0})$ is of the form

$$\begin{pmatrix} \alpha_{k,1}(\mathbf{0}) & \cdots & \cdots & \alpha_{k,n}(\mathbf{0}) \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

and its determinant equals $\alpha_{k,1}(\mathbf{0}) \neq 0$. Therefore, by the inverse function theorem, there exists an open set U_1 containing $\mathbf{0}$ and an open set V_2 containing $\mathbf{0}$ such that $\mathbf{G}_1(U_1) = V_2$ and $\mathbf{G}_1$ is one to one and onto such that it and its inverse are both C^1 . Let $\mathbf{F}_1$ denote the flip which interchanges x_k with x_1 . Now define

$$\mathbf{h}_2(\mathbf{y}) \equiv \mathbf{F}_1 \circ \mathbf{h}_1 \circ \mathbf{G}_1^{-1}(\mathbf{y})$$

Thus

$$\begin{aligned} \mathbf{h}_2(\mathbf{G}_1(\mathbf{x})) &\equiv \mathbf{F}_1 \circ \mathbf{h}_1(\mathbf{x}) \\ &= \begin{pmatrix} \alpha_k(\mathbf{x}) & \cdots & \alpha_1(\mathbf{x}) & \cdots & \alpha_n(\mathbf{x}) \end{pmatrix}^T \end{aligned} \quad (4.26)$$

Therefore,

$$P_1 \mathbf{h}_2(\mathbf{G}_1(\mathbf{x})) = \begin{pmatrix} \alpha_k(\mathbf{x}) & 0 & \cdots & 0 \end{pmatrix}^T.$$

Also

$$P_1(\mathbf{G}_1(\mathbf{x})) = \begin{pmatrix} \alpha_k(\mathbf{x}) & 0 & \cdots & 0 \end{pmatrix}^T$$

so $P_1 \mathbf{h}_2(\mathbf{y}) = P_1(\mathbf{y})$ for all $\mathbf{y} \in V_2$. Also, $\mathbf{h}_2(\mathbf{0}) = \mathbf{0}$ and $D\mathbf{h}_2(\mathbf{0})^{-1}$ exists because of the definition of $\mathbf{h}_2$ above and the chain rule. Also, since $\mathbf{F}_1^2 = I$, the identity map, it follows from 4.26 that

$$\mathbf{h}(\mathbf{x}) = \mathbf{h}_1(\mathbf{x}) = \mathbf{F}_1 \circ \mathbf{h}_2 \circ \mathbf{G}_1(\mathbf{x}). \quad (4.27)$$

Note that on an open set containing the origin, $\mathbf{h}_2$ leaves the first entry unchanged. This is what $P_1 \mathbf{h}_2(\mathbf{y}) = P_1(\mathbf{y})$ says. In contrast, $\mathbf{h}_1 = \mathbf{h}$ left possibly no entries unchanged.

Suppose then that for $m \geq 2$, $\mathbf{h}_m$ leaves the first $m-1$ entries unchanged,

$$P_{m-1} \mathbf{h}_m(\mathbf{x}) = P_{m-1}(\mathbf{x}) \quad (4.28)$$

for all $\mathbf{x} \in U_m$, an open subset of U containing $\mathbf{0}$ and $\mathbf{h}_m(\mathbf{0}) = \mathbf{0}$, $D\mathbf{h}_m(\mathbf{0})^{-1}$ exists. From 4.28, $\mathbf{h}_m(\mathbf{x})$ must be of the form

$$\mathbf{h}_m(\mathbf{x}) = \begin{pmatrix} x_1 & \cdots & x_{m-1} & \alpha_1(\mathbf{x}) & \cdots & \alpha_n(\mathbf{x}) \end{pmatrix}^T$$

where these α_k are different than the ones used earlier. Then

$$D\mathbf{h}_m(\mathbf{0})\mathbf{e}_m = \begin{pmatrix} 0 & \cdots & 0 & \alpha_{1,m}(\mathbf{0}) & \cdots & \alpha_{n,m}(\mathbf{0}) \end{pmatrix}^T \neq \mathbf{0}$$

because $D\mathbf{h}_m(\mathbf{0})^{-1}$ exists. Therefore, there exists a $k \geq m$ such that $\alpha_{k,m}(\mathbf{0}) \neq 0$, not the same k as before. Define

$$\mathbf{G}_m(\mathbf{x}) \equiv \begin{pmatrix} x_1 & \cdots & x_{m-1} & \alpha_k(\mathbf{x}) & \cdots & x_n \end{pmatrix}^T \quad (4.29)$$

so a change in $\mathbf{G}_m$ occurs only in the m^{th} slot. Then $\mathbf{G}_m(\mathbf{0}) = \mathbf{0}$ and $D\mathbf{G}_m(\mathbf{0})^{-1}$ exists similar to the above. In fact

$$\det(D\mathbf{G}_m(\mathbf{0})) = \alpha_{k,m}(\mathbf{0}).$$

Therefore, by the inverse function theorem, there exists an open set V_m containing $\mathbf{0}$ such that $V_m = \mathbf{G}_m(U_m)$ with $\mathbf{G}_m$ and its inverse being one to one, continuous and onto. Let $\mathbf{F}_m$ be the flip which flips x_m and x_k . Then define $\mathbf{h}_{m+1}$ on V_{m+1} by

$$\mathbf{h}_{m+1}(\mathbf{y}) = \mathbf{F}_m \circ \mathbf{h}_m \circ \mathbf{G}_m^{-1}(\mathbf{y}).$$

Thus for $\mathbf{x} \in U_m$,

$$\mathbf{h}_{m+1}(\mathbf{G}_m(\mathbf{x})) = (\mathbf{F}_m \circ \mathbf{h}_m)(\mathbf{x}). \quad (4.30)$$

and consequently, since $\mathbf{F}_m^2 = I$,

$$\mathbf{F}_m \circ \mathbf{h}_{m+1} \circ \mathbf{G}_m(\mathbf{x}) = \mathbf{h}_m(\mathbf{x}) \quad (4.31)$$

It follows

$$\begin{aligned} P_m \mathbf{h}_{m+1}(\mathbf{G}_m(\mathbf{x})) &= P_m(\mathbf{F}_m \circ \mathbf{h}_m)(\mathbf{x}) \\ &= \begin{pmatrix} x_1 & \cdots & x_{m-1} & \alpha_k(\mathbf{x}) & 0 & \cdots & 0 \end{pmatrix}^T \end{aligned}$$

and

$$P_m(\mathbf{G}_m(\mathbf{x})) = \begin{pmatrix} x_1 & \cdots & x_{m-1} & \alpha_k(\mathbf{x}) & 0 & \cdots & 0 \end{pmatrix}^T.$$

Therefore, for $\mathbf{y} \in V_m$,

$$P_m \mathbf{h}_{m+1}(\mathbf{y}) = P_m(\mathbf{y}).$$

As before, $\mathbf{h}_{m+1}(\mathbf{0}) = \mathbf{0}$ and $D\mathbf{h}_{m+1}(\mathbf{0})^{-1}$ exists. Therefore, we can apply 4.31 repeatedly, obtaining the following:

$$\begin{aligned} \mathbf{h}(\mathbf{x}) &= \mathbf{F}_1 \circ \mathbf{h}_2 \circ \mathbf{G}_1(\mathbf{x}) \\ &= \mathbf{F}_1 \circ \mathbf{F}_2 \circ \mathbf{h}_3 \circ \mathbf{G}_2 \circ \mathbf{G}_1(\mathbf{x}) \\ &\vdots \\ &= \mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1} \circ \mathbf{h}_n \circ \mathbf{G}_{n-1} \circ \cdots \circ \mathbf{G}_1(\mathbf{x}) \end{aligned}$$

where $\mathbf{h}_n$ fixes the first $n-1$ entries,

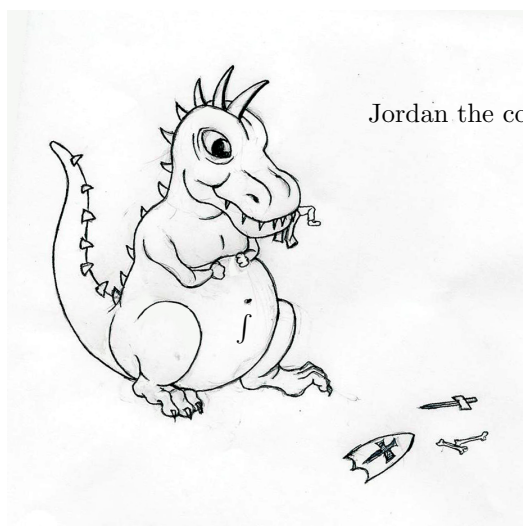
$$P_{n-1} \mathbf{h}_n(\mathbf{x}) = P_{n-1}(\mathbf{x}) = \begin{pmatrix} x_1 & \cdots & x_{n-1} & 0 \end{pmatrix}^T,$$

and so $\mathbf{h}_n(\mathbf{x})$ is a primitive mapping of the form

$$\mathbf{h}_n(\mathbf{x}) = \begin{pmatrix} x_1 & \cdots & x_{n-1} & \alpha(\mathbf{x}) \end{pmatrix}^T.$$

Therefore, define the primitive function $\mathbf{G}_n(\mathbf{x})$ to equal $\mathbf{h}_n(\mathbf{x})$.

The Theory Of The Riemann Integral*



Jordan the contented dragon

The definition of the Riemann integral of a function of n variables uses the following definition.

Definition E.0.3 For $i = 1, \dots, n$, let $\{\alpha_k^i\}_{k=-\infty}^{\infty}$ be points on $\mathbb{R}$ which satisfy

$$\lim_{k \rightarrow \infty} \alpha_k^i = \infty, \quad \lim_{k \rightarrow -\infty} \alpha_k^i = -\infty, \quad \alpha_k^i < \alpha_{k+1}^i. \quad (5.1)$$

For such sequences, define a grid on $\mathbb{R}^n$ denoted by $\mathcal{G}$ or $\mathcal{F}$ as the collection of boxes of the form

$$Q = \prod_{i=1}^n [\alpha_{j_i}^i, \alpha_{j_i+1}^i]. \quad (5.2)$$

If $\mathcal{G}$ is a grid, $\mathcal{F}$ is called a refinement of $\mathcal{G}$ if every box of $\mathcal{G}$ is the union of boxes of $\mathcal{F}$.

Lemma E.0.4 If $\mathcal{G}$ and $\mathcal{F}$ are two grids, they have a common refinement, denoted here by $\mathcal{G} \vee \mathcal{F}$.

Proof: Let $\{\alpha_k^i\}_{k=-\infty}^{\infty}$ be the sequences used to construct $\mathcal{G}$ and let $\{\beta_k^i\}_{k=-\infty}^{\infty}$ be the sequence used to construct $\mathcal{F}$. Now let $\{\gamma_k^i\}_{k=-\infty}^{\infty}$ denote the union of $\{\alpha_k^i\}_{k=-\infty}^{\infty}$

and $\{\beta_k^i\}_{k=-\infty}^\infty$. It is necessary to show that for each i these points can be arranged in order. To do so, let $\gamma_0^i \equiv \alpha_0^i$. Now if

$$\gamma_{-j}^i, \dots, \gamma_0^i, \dots, \gamma_j^i$$

have been chosen such that they are in order and all distinct, let γ_{j+1}^i be the first element of

$$\{\alpha_k^i\}_{k=-\infty}^\infty \cup \{\beta_k^i\}_{k=-\infty}^\infty \quad (5.3)$$

which is larger than γ_j^i and let $\gamma_{-(j+1)}^i$ be the last element of 5.3 which is strictly smaller than γ_{-j}^i . The assumption 5.1 insures such a first and last element exists. Now let the grid $\mathcal{G} \vee \mathcal{F}$ consist of boxes of the form

$$Q \equiv \prod_{i=1}^n [\gamma_{j_i}^i, \gamma_{j_i+1}^i].$$

The Riemann integral is only defined for functions, f which are bounded and are equal to zero off some bounded set, D . In what follows f will always be such a function.

Definition E.0.5 *Let f be a bounded function which equals zero off a bounded set, D , and let $\mathcal{G}$ be a grid. For $Q \in \mathcal{G}$, define*

$$M_Q(f) \equiv \sup \{f(\mathbf{x}) : \mathbf{x} \in Q\}, \quad m_Q(f) \equiv \inf \{f(\mathbf{x}) : \mathbf{x} \in Q\}. \quad (5.4)$$

Also define for Q a box, the volume of Q , denoted by $v(Q)$ by

$$v(Q) \equiv \prod_{i=1}^n (b_i - a_i), \quad Q \equiv \prod_{i=1}^n [a_i, b_i].$$

Now define upper sums, $\mathcal{U}_{\mathcal{G}}(f)$ and lower sums, $\mathcal{L}_{\mathcal{G}}(f)$ with respect to the indicated grid, by the formulas

$$\mathcal{U}_{\mathcal{G}}(f) \equiv \sum_{Q \in \mathcal{G}} M_Q(f) v(Q), \quad \mathcal{L}_{\mathcal{G}}(f) \equiv \sum_{Q \in \mathcal{G}} m_Q(f) v(Q).$$

A function of n variables is Riemann integrable when there is a unique number between all the upper and lower sums. This number is the value of the integral.

Note that in this definition, $M_Q(f) = m_Q(f) = 0$ for all but finitely many $Q \in \mathcal{G}$ so there are no convergence questions to be considered here.

Lemma E.0.6 *If $\mathcal{F}$ is a refinement of $\mathcal{G}$ then*

$$\mathcal{U}_{\mathcal{G}}(f) \geq \mathcal{U}_{\mathcal{F}}(f), \quad \mathcal{L}_{\mathcal{G}}(f) \leq \mathcal{L}_{\mathcal{F}}(f).$$

Also if $\mathcal{F}$ and $\mathcal{G}$ are two grids,

$$\mathcal{L}_{\mathcal{G}}(f) \leq \mathcal{U}_{\mathcal{F}}(f).$$

Proof: For $P \in \mathcal{G}$ let $\hat{P}$ denote the set,

$$\{Q \in \mathcal{F} : Q \subseteq P\}.$$

Then $P = \cup \hat{P}$ and

$$\mathcal{L}_{\mathcal{F}}(f) \equiv \sum_{Q \in \mathcal{F}} m_Q(f) v(Q) = \sum_{P \in \mathcal{G}} \sum_{Q \in \hat{P}} m_Q(f) v(Q)$$

$$\geq \sum_{P \in \mathcal{G}} m_P(f) \sum_{Q \in \hat{P}} v(Q) = \sum_{P \in \mathcal{G}} m_P(f) v(P) \equiv \mathcal{L}_{\mathcal{G}}(f).$$

Similarly, the other inequality for the upper sums is valid.

To verify the last assertion of the lemma, use Lemma E.0.4 to write

$$\mathcal{L}_{\mathcal{G}}(f) \leq \mathcal{L}_{\mathcal{G} \vee \mathcal{F}}(f) \leq \mathcal{U}_{\mathcal{G} \vee \mathcal{F}}(f) \leq \mathcal{U}_{\mathcal{F}}(f).$$

This proves the lemma.

This lemma makes it possible to define the Riemann integral.

Definition E.0.7 Define an upper and a lower integral as follows.

$$\begin{aligned} \bar{I}(f) &\equiv \inf \{ \mathcal{U}_{\mathcal{G}}(f) : \mathcal{G} \text{ is a grid} \}, \\ \underline{I}(f) &\equiv \sup \{ \mathcal{L}_{\mathcal{G}}(f) : \mathcal{G} \text{ is a grid} \}. \end{aligned}$$

Lemma E.0.8 $\bar{I}(f) \geq \underline{I}(f)$.

Proof: From Lemma E.0.6 it follows for any two grids $\mathcal{G}$ and $\mathcal{F}$,

$$\mathcal{L}_{\mathcal{G}}(f) \leq \mathcal{U}_{\mathcal{F}}(f).$$

Therefore, taking the supremum for all grids on the left in this inequality,

$$\underline{I}(f) \leq \mathcal{U}_{\mathcal{F}}(f)$$

for all grids $\mathcal{F}$. Taking the infimum in this inequality, yields the conclusion of the lemma.

Definition E.0.9 A bounded function, f which equals zero off a bounded set, D , is said to be Riemann integrable, written as $f \in \mathcal{R}(\mathbb{R}^n)$ exactly when $\underline{I}(f) = \bar{I}(f)$. In this case define

$$\int f dV \equiv \int f dx = \bar{I}(f) = \underline{I}(f).$$

As in the case of integration of functions of one variable, one obtains the Riemann criterion which is stated as the following theorem.

Theorem E.0.10 (Riemann criterion) $f \in \mathcal{R}(\mathbb{R}^n)$ if and only if for all $\varepsilon > 0$ there exists a grid $\mathcal{G}$ such that

$$\mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{G}}(f) < \varepsilon.$$

Proof: If $f \in \mathcal{R}(\mathbb{R}^n)$, then $\bar{I}(f) = \underline{I}(f)$ and so there exist grids $\mathcal{G}$ and $\mathcal{F}$ such that

$$\mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{F}}(f) \leq \bar{I}(f) + \frac{\varepsilon}{2} - \left(\underline{I}(f) - \frac{\varepsilon}{2} \right) = \varepsilon.$$

Then letting $\mathcal{H} = \mathcal{G} \vee \mathcal{F}$, Lemma E.0.6 implies

$$\mathcal{U}_{\mathcal{H}}(f) - \mathcal{L}_{\mathcal{H}}(f) \leq \mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{F}}(f) < \varepsilon.$$

Conversely, if for all $\varepsilon > 0$ there exists $\mathcal{G}$ such that

$$\mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{G}}(f) < \varepsilon,$$

then

$$\bar{I}(f) - \underline{I}(f) \leq \mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{G}}(f) < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this proves the theorem.

E.1 Basic Properties

It is important to know that certain combinations of Riemann integrable functions are Riemann integrable. The following theorem will include all the important cases.

Theorem E.1.1 *Let $f, g \in \mathcal{R}(\mathbb{R}^n)$ and let $\phi : K \rightarrow \mathbb{R}$ be continuous where K is a compact set in $\mathbb{R}^2$ containing $f(\mathbb{R}^n) \times g(\mathbb{R}^n)$. Also suppose that $\phi(0, 0) = 0$. Then defining*

$$h(\mathbf{x}) \equiv \phi(f(\mathbf{x}), g(\mathbf{x})),$$

it follows that h is also in $\mathcal{R}(\mathbb{R}^n)$.

Proof: Let $\varepsilon > 0$ and let $\delta_1 > 0$ be such that if $(y_i, z_i), i = 1, 2$ are points in K , such that $|z_1 - z_2| \leq \delta_1$ and $|y_1 - y_2| \leq \delta_1$, then

$$|\phi(y_1, z_1) - \phi(y_2, z_2)| < \varepsilon.$$

Let $0 < \delta < \min(\delta_1, \varepsilon, 1)$. Let $\mathcal{G}$ be a grid with the property that for $Q \in \mathcal{G}$, the diameter of Q is less than δ and also for $k = f, g$,

$$\mathcal{U}_{\mathcal{G}}(k) - \mathcal{L}_{\mathcal{G}}(k) < \delta^2. \quad (5.5)$$

Then defining for $k = f, g$,

$$\mathcal{P}_k \equiv \{Q \in \mathcal{G} : M_Q(k) - m_Q(k) > \delta\},$$

it follows

$$\begin{aligned} \delta^2 &> \sum_{Q \in \mathcal{G}} (M_Q(k) - m_Q(k)) v(Q) \geq \\ &\sum_{\mathcal{P}_k} (M_Q(k) - m_Q(k)) v(Q) \geq \delta \sum_{\mathcal{P}_k} v(Q) \end{aligned}$$

and so for $k = f, g$,

$$\varepsilon > \delta > \sum_{\mathcal{P}_k} v(Q). \quad (5.6)$$

Suppose for $k = f, g$,

$$M_Q(k) - m_Q(k) \leq \delta.$$

Then if $\mathbf{x}_1, \mathbf{x}_2 \in Q$,

$$|f(\mathbf{x}_1) - f(\mathbf{x}_2)| < \delta, \text{ and } |g(\mathbf{x}_1) - g(\mathbf{x}_2)| < \delta.$$

Therefore,

$$|h(\mathbf{x}_1) - h(\mathbf{x}_2)| \equiv |\phi(f(\mathbf{x}_1), g(\mathbf{x}_1)) - \phi(f(\mathbf{x}_2), g(\mathbf{x}_2))| < \varepsilon$$

and it follows that

$$|M_Q(h) - m_Q(h)| \leq \varepsilon.$$

Now let

$$\mathcal{S} \equiv \{Q \in \mathcal{G} : 0 < M_Q(k) - m_Q(k) \leq \delta, k = f, g\}.$$

Thus the union of the boxes in $\mathcal{S}$ is contained in some large box, R , which depends only on f and g and also, from the assumption that $\phi(0, 0) = 0$, $M_Q(h) - m_Q(h) = 0$ unless $Q \subseteq R$. Then

$$\mathcal{U}_{\mathcal{G}}(h) - \mathcal{L}_{\mathcal{G}}(h) \leq \sum_{Q \in \mathcal{P}_f} (M_Q(h) - m_Q(h)) v(Q) +$$

$$\sum_{Q \in \mathcal{P}_g} (M_Q(h) - m_Q(h)) v(Q) + \sum_{Q \in \mathcal{S}} \delta v(Q).$$

Now since K is compact, it follows $\phi(K)$ is bounded and so there exists a constant, C , depending only on h and ϕ such that $M_Q(h) - m_Q(h) < C$. Therefore, the above inequality implies

$$\mathcal{U}_{\mathcal{G}}(h) - \mathcal{L}_{\mathcal{G}}(h) \leq C \sum_{Q \in \mathcal{P}_f} v(Q) + C \sum_{Q \in \mathcal{P}_g} v(Q) + \sum_{Q \in \mathcal{S}} \delta v(Q),$$

which by 5.6 implies

$$\mathcal{U}_{\mathcal{G}}(h) - \mathcal{L}_{\mathcal{G}}(h) \leq 2C\varepsilon + \delta v(R) \leq 2C\varepsilon + \varepsilon v(R).$$

Since ε is arbitrary, the Riemann criterion is satisfied and so $h \in \mathcal{R}(\mathbb{R}^n)$.

Corollary E.1.2 *Let $f, g \in \mathcal{R}(\mathbb{R}^n)$ and let $a, b \in \mathbb{R}$. Then $af + bg$, fg , and $|f|$ are all in $\mathcal{R}(\mathbb{R}^n)$. Also,*

$$\int_{\mathbb{R}^n} (af + bg) dx = a \int_{\mathbb{R}^n} f dx + b \int_{\mathbb{R}^n} g dx, \quad (5.7)$$

and

$$\int |f| dx \geq \left| \int f dx \right|. \quad (5.8)$$

Proof: Each of the combinations of functions described above is Riemann integrable by Theorem E.1.1. For example, to see $af + bg \in \mathcal{R}(\mathbb{R}^n)$ consider $\phi(y, z) \equiv ay + bz$. This is clearly a continuous function of (y, z) such that $\phi(0, 0) = 0$. To obtain $|f| \in \mathcal{R}(\mathbb{R}^n)$, let $\phi(y, z) \equiv |y|$. It remains to verify the formulas. To do so, let $\mathcal{G}$ be a grid with the property that for $k = f, g, |f|$ and $af + bg$,

$$\mathcal{U}_{\mathcal{G}}(k) - \mathcal{L}_{\mathcal{G}}(k) < \varepsilon. \quad (5.9)$$

Consider 5.7. For each $Q \in \mathcal{G}$ pick a point in Q , $\mathbf{x}_Q$. Then

$$\sum_{Q \in \mathcal{G}} k(\mathbf{x}_Q) v(Q) \in [\mathcal{L}_{\mathcal{G}}(k), \mathcal{U}_{\mathcal{G}}(k)]$$

and so

$$\left| \int k dx - \sum_{Q \in \mathcal{G}} k(\mathbf{x}_Q) v(Q) \right| < \varepsilon.$$

Consequently, since

$$\begin{aligned} & \sum_{Q \in \mathcal{G}} (af + bg)(\mathbf{x}_Q) v(Q) \\ &= a \sum_{Q \in \mathcal{G}} f(\mathbf{x}_Q) v(Q) + b \sum_{Q \in \mathcal{G}} g(\mathbf{x}_Q) v(Q), \end{aligned}$$

it follows

$$\begin{aligned} & \left| \int (af + bg) dx - a \int f dx - b \int g dx \right| \leq \\ & \left| \int (af + bg) dx - \sum_{Q \in \mathcal{G}} (af + bg)(\mathbf{x}_Q) v(Q) \right| + \end{aligned}$$

$$\begin{aligned} & \left| a \sum_{Q \in \mathcal{G}} f(\mathbf{x}_Q) v(Q) - a \int f dx \right| + \left| b \sum_{Q \in \mathcal{G}} g(\mathbf{x}_Q) v(Q) - b \int g dx \right| \\ & \leq \varepsilon + |a| \varepsilon + |b| \varepsilon. \end{aligned}$$

Since ε is arbitrary, this establishes Formula 5.7 and shows the integral is linear.

It remains to establish the inequality 5.8. By 5.9, and the triangle inequality for sums,

$$\begin{aligned} \int |f| dx + \varepsilon & \geq \sum_{Q \in \mathcal{G}} |f(\mathbf{x}_Q)| v(Q) \geq \\ & \geq \left| \sum_{Q \in \mathcal{G}} f(\mathbf{x}_Q) v(Q) \right| \geq \left| \int f dx \right| - \varepsilon. \end{aligned}$$

Then since ε is arbitrary, this establishes the desired inequality. This proves the corollary.

E.2 Which Functions Are Integrable?

Which functions are in $\mathcal{R}(\mathbb{R}^n)$? As in the case of integrals of functions of one variable, this is an important question. It turns out the Riemann integrable functions are characterized by being continuous except on a very small set. This has to do with Jordan content.

Definition E.2.1 A bounded set, E , has Jordan content 0 or content 0 if for every $\varepsilon > 0$ there exists a grid, $\mathcal{G}$ such that

$$\sum_{Q \cap E \neq \emptyset} v(Q) < \varepsilon.$$

This symbol says to sum the volumes of all boxes from $\mathcal{G}$ which have nonempty intersection with E .

Next it is necessary to define the oscillation of a function.

Definition E.2.2 Let f be a function defined on $\mathbb{R}^n$ and let

$$\omega_{f,r}(\mathbf{x}) \equiv \sup \{ |f(\mathbf{z}) - f(\mathbf{y})| : \mathbf{z}, \mathbf{y} \in B(\mathbf{x}, r) \}.$$

This is called the oscillation of f on $B(\mathbf{x}, r)$. Note that this function of r is decreasing in r . Define the oscillation of f as

$$\omega_f(\mathbf{x}) \equiv \lim_{r \rightarrow 0+} \omega_{f,r}(\mathbf{x}).$$

Note that as r decreases, the function, $\omega_{f,r}(\mathbf{x})$ decreases. It is also bounded below by 0 and so the limit must exist and equals $\inf \{ \omega_{f,r}(\mathbf{x}) : r > 0 \}$. (Why?) Then the following simple lemma whose proof follows directly from the definition of continuity gives the reason for this definition.

Lemma E.2.3 A function f is continuous at $\mathbf{x}$ if and only if $\omega_f(\mathbf{x}) = 0$.

This concept of oscillation gives a way to define how discontinuous a function is at a point. The discussion will depend on the following fundamental lemma which gives the existence of something called the Lebesgue number.

Definition E.2.4 Let $\mathfrak{C}$ be a set whose elements are sets of $\mathbb{R}^n$ and let $K \subseteq \mathbb{R}^n$. The set, $\mathfrak{C}$ is called a cover of K if every point of K is contained in some set of $\mathfrak{C}$. If the elements of $\mathfrak{C}$ are open sets, it is called an open cover.

Lemma E.2.5 Let K be sequentially compact and let $\mathfrak{C}$ be an open cover of K . Then there exists $r > 0$ such that whenever $\mathbf{x} \in K$, $B(\mathbf{x}, r)$ is contained in some set of $\mathfrak{C}$.

Proof: Suppose this is not so. Then letting $r_n = 1/n$, there exists $\mathbf{x}_n \in K$ such that $B(\mathbf{x}_n, r_n)$ is not contained in any set of $\mathfrak{C}$. Since K is sequentially compact, there is a subsequence, $\mathbf{x}_{n_k}$ which converges to a point, $\mathbf{x} \in K$. But there exists $\delta > 0$ such that $B(\mathbf{x}, \delta) \subseteq U$ for some $U \in \mathfrak{C}$. Let k be so large that $1/k < \delta/2$ and $|\mathbf{x}_{n_k} - \mathbf{x}| < \delta/2$ also. Then if $\mathbf{z} \in B(\mathbf{x}_{n_k}, r_{n_k})$, it follows

$$|\mathbf{z} - \mathbf{x}| \leq |\mathbf{z} - \mathbf{x}_{n_k}| + |\mathbf{x}_{n_k} - \mathbf{x}| < \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

and so $B(\mathbf{x}_{n_k}, r_{n_k}) \subseteq U$ contrary to supposition. Therefore, the desired number exists after all.

Theorem E.2.6 Let f be a bounded function which equals zero off a bounded set and let W denote the set of points where f fails to be continuous. Then $f \in \mathcal{R}(\mathbb{R}^n)$ if W has content zero. That is, for all $\varepsilon > 0$ there exists a grid, $\mathcal{G}$ such that

$$\sum_{Q \in \mathcal{G}_W} v(Q) < \varepsilon \quad (5.10)$$

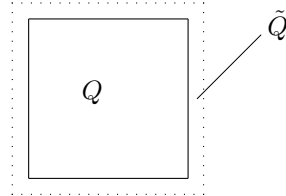
where

$$\mathcal{G}_W \equiv \{Q \in \mathcal{G} : Q \cap W \neq \emptyset\}.$$

Proof: Let W have content zero. Also let $|f(\mathbf{x})| < C/2$ for all $\mathbf{x} \in \mathbb{R}^n$, let $\varepsilon > 0$ be given, and let $\mathcal{G}$ be a grid which satisfies 5.10. Since f equals zero off some bounded set, there exists R such that f equals zero off of $B(\mathbf{0}, \frac{R}{2})$. Thus $W \subseteq B(\mathbf{0}, \frac{R}{2})$. Also note that if $\mathcal{G}$ is a grid for which 5.10 holds, then this inequality continues to hold if $\mathcal{G}$ is replaced with a refined grid. Therefore, you may assume the diameter of every box in $\mathcal{G}$ which intersects $B(\mathbf{0}, R)$ is less than $\frac{R}{3}$ and so all boxes of $\mathcal{G}$ which intersect the set where f is nonzero are contained in $B(\mathbf{0}, R)$. Since W is bounded, $\mathcal{G}_W$ contains only finitely many boxes. Letting

$$Q \equiv \prod_{i=1}^n [a_i, b_i]$$

be one of these boxes, enlarge the box slightly as indicated in the following picture.



The enlarged box is an open set of the form,

$$\tilde{Q} \equiv \prod_{i=1}^n (a_i - \eta_i, b_i + \eta_i)$$

where η_i is chosen small enough that if

$$\prod_{i=1}^n (b_i + \eta_i - (a_i - \eta_i)) \equiv v(\tilde{Q}),$$

and $\widetilde{\mathcal{G}_W}$ denotes those $\tilde{Q}$ for $Q \in \mathcal{G}$ which have nonempty intersection with W , then

$$\sum_{\tilde{Q} \in \widetilde{\mathcal{G}_W}} v(\tilde{Q}) < \varepsilon \quad (5.11)$$

where $\tilde{Q}$ is the box,

$$\prod_{i=1}^n (b_i + 2\eta_i - (a_i - 2\eta_i))$$

For each $\mathbf{x} \in \mathbb{R}^n$, let $r_{\mathbf{x}} < \min(\eta_1/2, \dots, \eta_n/2)$ be such that

$$\omega_{f, r_{\mathbf{x}}}(\mathbf{x}) < \varepsilon + \omega_f(\mathbf{x}). \quad (5.12)$$

Now let $\mathfrak{C}$ denote all intersections of the form $\tilde{Q} \cap B(\mathbf{x}, r_{\mathbf{x}})$ such that $\mathbf{x} \in \overline{B(\mathbf{0}, R)}$ so that $\mathfrak{C}$ is an open cover of the compact set, $\overline{B(\mathbf{0}, R)}$. Let δ be a Lebesgue number for this open cover of $\overline{B(\mathbf{0}, R)}$ and let $\mathcal{F}$ be a refinement of $\mathcal{G}$ such that every box in $\mathcal{F}$ has diameter less than δ . Now let $\mathcal{F}_1$ consist of those boxes of $\mathcal{F}$ which have nonempty intersection with $B(\mathbf{0}, R/2)$. Thus all boxes of $\mathcal{F}_1$ are contained in $B(\mathbf{0}, R)$ and each one is contained in some set of $\mathfrak{C}$. Let $\mathfrak{C}_W$ be those open sets of $\mathfrak{C}$, $\tilde{Q} \cap B(\mathbf{x}, r_{\mathbf{x}})$, for which $\mathbf{x} \in W$. Thus each of these sets is contained in some $\tilde{Q}$ where $Q \in \mathcal{G}_W$. Let $\mathcal{F}_W$ be those sets of $\mathcal{F}_1$ which are subsets of some set of $\mathfrak{C}_W$. Thus

$$\sum_{Q \in \mathcal{F}_W} v(Q) < \varepsilon. \quad (5.13)$$

because each Q in $\mathcal{F}_W$ is contained in a set, $\tilde{Q}$ described above and the sum of the volumes of these is less than ε by 5.11. Then

$$\begin{aligned} \mathcal{U}_{\mathcal{F}}(f) - \mathcal{L}_{\mathcal{F}}(f) &= \sum_{Q \in \mathcal{F}_W} (M_Q(f) - m_Q(f)) v(Q) \\ &+ \sum_{Q \in \mathcal{F}_1 \setminus \mathcal{F}_W} (M_Q(f) - m_Q(f)) v(Q). \end{aligned}$$

If $Q \in \mathcal{F}_1 \setminus \mathcal{F}_W$, then Q must be a subset of some set of $\mathfrak{C} \setminus \mathfrak{C}_W$ since it is not in any set of $\mathfrak{C}_W$. Say $Q \subseteq \tilde{Q}_1 \cap B(\mathbf{x}, r_{\mathbf{x}})$ where $\mathbf{x} \notin W$. Therefore, from 5.12 and the observation that $\mathbf{x} \notin W$, it follows $\omega_f(\mathbf{x}) = 0$ and so

$$M_Q(f) - m_Q(f) \leq \varepsilon.$$

Therefore, from 5.13 and the estimate on f ,

$$\begin{aligned} \mathcal{U}_{\mathcal{F}}(f) - \mathcal{L}_{\mathcal{F}}(f) &\leq \sum_{Q \in \mathcal{F}_W} C v(Q) + \sum_{Q \in \mathcal{F}_1 \setminus \mathcal{F}_W} \varepsilon v(Q) \\ &\leq C\varepsilon + \varepsilon(2R)^n, \end{aligned}$$

the estimate of the second sum coming from the fact

$$B(\mathbf{0}, R) \subseteq \prod_{i=1}^n [-R, R].$$

Since ε is arbitrary, this proves the theorem.¹

Definition E.2.7 A bounded set, E is a Jordan set in $\mathbb{R}^n$ or a contented set in $\mathbb{R}^n$ if $\chi_E \in \mathcal{R}(\mathbb{R}^n)$. The symbol χ_E means

$$\chi_E(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in E \\ 0 & \text{if } \mathbf{x} \notin E \end{cases}$$

It is called the indicator function because it indicates whether $\mathbf{x}$ is in E according to whether it equals 1. For a function $f \in \mathcal{R}(\mathbb{R}^n)$ and E a contented set, $f\chi_E \in \mathcal{R}(\mathbb{R}^n)$ by Corollary E.1.2. Then

$$\int_E f dV \equiv \int f \chi_E dV.$$

So what are examples of contented sets?

Theorem E.2.8 Suppose E is a bounded contented set in $\mathbb{R}^n$ and $f, g : E \rightarrow \mathbb{R}$ are two functions satisfying $f(\mathbf{x}) \geq g(\mathbf{x})$ for all $\mathbf{x} \in E$ and $f\chi_E$ and $g\chi_E$ are both in $\mathcal{R}(\mathbb{R}^n)$. Now define

$$P \equiv \{(\mathbf{x}, x_{n+1}) : \mathbf{x} \in E \text{ and } g(\mathbf{x}) \leq x_{n+1} \leq f(\mathbf{x})\}.$$

Then P is a contented set in $\mathbb{R}^{n+1}$.

Proof: Let $\mathcal{G}$ be a grid such that for $k = f\chi_E, g\chi_E$,

$$\mathcal{U}_{\mathcal{G}}(k) - \mathcal{L}_{\mathcal{G}}(k) < \varepsilon/4. \quad (5.14)$$

Also let $K \geq \sum_{j=1}^m v_n(Q_j)$ where the Q_j are the boxes which intersect E . Let $\{a_i\}_{i=-\infty}^{\infty}$ be a sequence on $\mathbb{R}$, $a_i < a_{i+1}$ for all i , which includes

$$\begin{aligned} &M_{Q_j}(f\chi_E) + \frac{\varepsilon}{4mK}, M_{Q_j}(f\chi_E), M_{Q_j}(g\chi_E), \\ &m_{Q_j}(f\chi_E), m_{Q_j}(g\chi_E), m_{Q_j}(g\chi_E) - \frac{\varepsilon}{4mK} \end{aligned}$$

for all $j = 1, \dots, m$. Now define a grid on $\mathbb{R}^{n+1}$ as follows.

$$\mathcal{G}' \equiv \{Q \times [a_i, a_{i+1}] : Q \in \mathcal{G}, i \in \mathbb{Z}\}$$

In words, this grid consists of all possible boxes of the form $Q \times [a_i, a_{i+1}]$ where $Q \in \mathcal{G}$ and a_i is a term of the sequence just described. It is necessary to verify that for $P \in \mathcal{G}'$, $\chi_P \in \mathcal{R}(\mathbb{R}^{n+1})$. This is done by showing that $\mathcal{U}_{\mathcal{G}'}(\chi_P) - \mathcal{L}_{\mathcal{G}'}(\chi_P) < \varepsilon$ and then noting that $\varepsilon > 0$ was arbitrary. For $\mathcal{G}'$ just described, denote by Q' a box in $\mathcal{G}'$. Thus $Q' = Q \times [a_i, a_{i+1}]$ for some i .

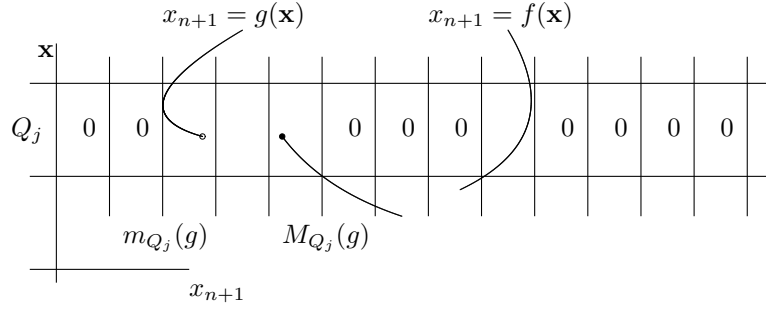
$$\begin{aligned} \mathcal{U}_{\mathcal{G}'}(\chi_P) - \mathcal{L}_{\mathcal{G}'}(\chi_P) &\equiv \sum_{Q' \in \mathcal{G}'} (M_{Q'}(\chi_P) - m_{Q'}(\chi_P)) v_{n+1}(Q') \\ &= \sum_{i=-\infty}^{\infty} \sum_{j=1}^m \left(M_{Q_j'}(\chi_P) - m_{Q_j'}(\chi_P) \right) v_n(Q_j) (a_{i+1} - a_i) \end{aligned}$$

¹In fact one cannot do any better. It can be shown that if a function is Riemann integrable, then it must be the case that for all $\varepsilon > 0$, 5.10 is satisfied for some grid, $\mathcal{G}$. This along with what was just shown is known as Lebesgue's theorem after Lebesgue who discovered it in the early years of the twentieth century. Actually, he also invented a far superior integral which has been the integral of serious mathematicians since that time.

and all sums are bounded because the functions, f and g are given to be bounded. Therefore, there are no limit considerations needed here. Thus

$$\mathcal{U}_{\mathcal{G}'}(\mathcal{X}_P) - \mathcal{L}_{\mathcal{G}'}(\mathcal{X}_P) = \sum_{j=1}^m v_n(Q_j) \sum_{i=-\infty}^{\infty} (M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P)) (a_{i+1} - a_i).$$

Consider the inside sum with the aid of the following picture.



In this picture, the little rectangles represent the boxes $Q_j \times [a_i, a_{i+1}]$ for fixed j . The part of P having $\mathbf{x}$ contained in Q_j is between the two surfaces, $x_{n+1} = g(\mathbf{x})$ and $x_{n+1} = f(\mathbf{x})$ and there is a zero placed in those boxes for which

$$M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) = 0.$$

You see, $\mathcal{X}_P$ has either the value of 1 or the value of 0 depending on whether $(\mathbf{x}, y)$ is contained in P . For the boxes shown with 0 in them, either all of the box is contained in P or none of the box is contained in P . Either way,

$$M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) = 0$$

on these boxes. However, on the boxes intersected by the surfaces, the value of

$$M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P)$$

is 1 because there are points in this box which are not in P as well as points which are in P . Because of the construction of $\mathcal{G}'$ which included all values of

$$M_{Q_j}(f\mathcal{X}_E) + \frac{\varepsilon}{4mK}, M_{Q_j}(f\mathcal{X}_E), \\ M_{Q_j}(g\mathcal{X}_E), m_{Q_j}(f\mathcal{X}_E), m_{Q_j}(g\mathcal{X}_E)$$

for all $j = 1, \dots, m$,

$$\sum_{i=-\infty}^{\infty} (M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P)) (a_{i+1} - a_i) \leq \sum_{\{i: m_{Q_j}(g\mathcal{X}_E) \leq a_i < M_{Q_j}(g\mathcal{X}_E)\}} 1(a_{i+1} - a_i) + \sum_{\{i: m_{Q_j}(f\mathcal{X}_E) \leq a_i < M_{Q_j}(f\mathcal{X}_E)\}} 1(a_{i+1} - a_i) \quad (5.15)$$

The first of the sums in 5.15 contains all possible terms for which

$$M_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P) - m_{Q_j \times [a_i, a_{i+1}]}(\mathcal{X}_P)$$

might be 1 due to the graph of the bottom surface $g\mathcal{X}_E$ while the second sum contains all possible terms for which the expression might be 1 due to the graph of the top surface $f\mathcal{X}_E$.

$$\begin{aligned} &\leq \left(M_{Q_j}(g\mathcal{X}_E) + \frac{\varepsilon}{4mK} - m_{Q_j}(g\mathcal{X}_E) \right) + \left(M_{Q_j}(f\mathcal{X}_E) + \frac{\varepsilon}{4mK} - m_{Q_j}(f\mathcal{X}_E) \right) \\ &= (M_{Q_j}(g\mathcal{X}_E) - m_{Q_j}(g\mathcal{X}_E)) + (M_{Q_j}(f\mathcal{X}_E) - m_{Q_j}(f\mathcal{X}_E)) + \frac{\varepsilon}{2m} \left(\sum_{j=1}^m v(Q_j) \right)^{-1}. \end{aligned}$$

Therefore, by 5.14,

$$\begin{aligned} &\mathcal{U}_{\mathcal{G}'}(\mathcal{X}_P) - \mathcal{L}_{\mathcal{G}'}(\mathcal{X}_P) \leq \\ &\sum_{j=1}^m v_n(Q_j) [(M_{Q_j}(g\mathcal{X}_E) - m_{Q_j}(g\mathcal{X}_E)) + (M_{Q_j}(f\mathcal{X}_E) - m_{Q_j}(f\mathcal{X}_E))] \\ &\quad + \sum_{j=1}^m v(Q_j) \frac{\varepsilon}{2m} \left(\sum_{j=1}^m v(Q_j) \right)^{-1} \\ &= \mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{G}}(f) + \mathcal{U}_{\mathcal{G}}(g) - \mathcal{L}_{\mathcal{G}}(g) + \frac{\varepsilon}{2} \\ &< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this proves the theorem.

Corollary E.2.9 Suppose f and g are continuous functions defined on E , a contented set in $\mathbb{R}^n$ and that $g(\mathbf{x}) \leq f(\mathbf{x})$ for all $\mathbf{x} \in E$. Then

$$P \equiv \{(\mathbf{x}, x_{n+1}) : \mathbf{x} \in E \text{ and } g(\mathbf{x}) \leq x_{n+1} \leq f(\mathbf{x})\}$$

is a contented set in $\mathbb{R}^n$.

Proof: Since E is contented, meaning $\mathcal{X}_E$ is integrable, it follows from Theorem E.2.6 the set of discontinuities of $\mathcal{X}_E$ has Jordan content 0. But the set of discontinuities of $\mathcal{X}_E$ is ∂E defined as those points $\mathbf{x}$ such that $B(\mathbf{x}, r)$ contains points of E and points of E^C for every $r > 0$. Extend f and g to equal 0 off E . Then the set of discontinuities of these extended functions, still denoted as f, g is ∂E which has Jordan content 0. This reduces to the situation of Theorem E.2.8. This proves the corollary.

As an example of how this can be applied, it is obvious a closed interval is a contented set in $\mathbb{R}$. Therefore, if f, g are two continuous functions with $f(x) \geq g(x)$ for $x \in [a, b]$, it follows from the above theorem or its corollary that the set,

$$P_1 \equiv \{(x, y) : g(x) \leq y \leq f(x)\}$$

is a contented set in $\mathbb{R}^2$. Now using the theorem and corollary again, suppose $f_1(x, y) \geq g_1(x, y)$ for $(x, y) \in P_1$ and f, g are continuous. Then the set

$$P_2 \equiv \{(x, y, z) : g_1(x, y) \leq z \leq f_1(x, y)\}$$

is a contented set in $\mathbb{R}^3$. Clearly you can continue this way obtaining examples of contented sets.

Note that as a special case, it follows that every box is a contented set. Therefore, if B_i is a box, functions of the form

$$\sum_{i=1}^m a_i \mathcal{X}_{B_i}$$

are integrable. These functions are called step functions.

The following theorem is analogous to the fact that in one dimension, when you integrate over a point, the answer is 0.

Theorem E.2.10 *If a bounded set, E , has Jordan content 0, then E is a Jordan (contented) set and if f is any bounded function defined on E , then $f\chi_E \in \mathcal{R}(\mathbb{R}^n)$ and*

$$\int_E f dV = 0.$$

Proof: Let $\mathcal{G}$ be a grid with

$$\sum_{Q \cap E \neq \emptyset} v(Q) < \frac{\varepsilon}{1 + (M - m)}.$$

Then

$$\mathcal{U}_{\mathcal{G}}(f\chi_E) \leq \sum_{Q \cap E \neq \emptyset} Mv(Q) \leq \frac{\varepsilon M}{1 + (M - m)}$$

and

$$\mathcal{L}_{\mathcal{G}}(f\chi_E) \geq \sum_{Q \cap E \neq \emptyset} mv(Q) \geq \frac{\varepsilon m}{1 + (M - m)}$$

and so

$$\begin{aligned} \mathcal{U}_{\mathcal{G}}(f\chi_E) - \mathcal{L}_{\mathcal{G}}(f\chi_E) &\leq \sum_{Q \cap E \neq \emptyset} Mv(Q) - \sum_{Q \cap E \neq \emptyset} mv(Q) \\ &= (M - m) \sum_{Q \cap E \neq \emptyset} v(Q) < \frac{\varepsilon(M - m)}{1 + (M - m)} < \varepsilon. \end{aligned}$$

This shows $f\chi_E \in \mathcal{R}(\mathbb{R}^n)$. Now also,

$$m\varepsilon \leq \int f\chi_E dV \leq M\varepsilon$$

and since ε is arbitrary, this shows

$$\int_E f dV \equiv \int f\chi_E dV = 0$$

Why is E contented? Let $\mathcal{G}$ be a grid for which

$$\sum_{Q \cap E \neq \emptyset} v(Q) < \varepsilon$$

Then for this grid,

$$\mathcal{U}_{\mathcal{G}}(\chi_E) - \mathcal{L}_{\mathcal{G}}(\chi_E) \leq \sum_{Q \cap E \neq \emptyset} v(Q) < \varepsilon$$

and this proves the theorem.

Corollary E.2.11 *If $f\chi_{E_i} \in \mathcal{R}(\mathbb{R}^n)$ for $i = 1, 2, \dots, r$ and for all $i \neq j$, $E_i \cap E_j$ is either the empty set or a set of Jordan content 0, then letting $F \equiv \cup_{i=1}^r E_i$, it follows $f\chi_F \in \mathcal{R}(\mathbb{R}^n)$ and*

$$\int f\chi_F dV \equiv \int_F f dV = \sum_{i=1}^r \int_{E_i} f dV.$$

Proof: This is true if $r = 1$. Suppose it is true for r . It will be shown that it is true for $r + 1$. Let $F_r = \cup_{i=1}^r E_i$ and let F_{r+1} be defined similarly. By the induction hypothesis, $f\chi_{F_r} \in \mathcal{R}(\mathbb{R}^n)$. Also, since F_r is a finite union of the E_i , it follows that $F_r \cap E_{r+1}$ is either empty or a set of Jordan content 0.

$$-f\chi_{F_r \cap E_{r+1}} + f\chi_{F_r} + f\chi_{E_{r+1}} = f\chi_{F_{r+1}}$$

and by Theorem E.2.10 each function on the left is in $\mathcal{R}(\mathbb{R}^n)$ and the first one on the left has integral equal to zero. Therefore,

$$\int f\chi_{F_{r+1}} dV = \int f\chi_{F_r} dV + \int f\chi_{E_{r+1}} dV$$

which by induction equals

$$\sum_{i=1}^r \int_{E_i} f dV + \int_{E_{r+1}} f dV = \sum_{i=1}^{r+1} \int_{E_i} f dV$$

and this proves the corollary.

In particular, for

$$Q = \prod_{i=1}^n [a_i, b_i], \quad Q' = \prod_{i=1}^n (a_i, b_i]$$

both are contented sets and

$$\int \chi_Q dV = \int_{Q'} \chi_{Q'} dV = v(Q). \quad (5.16)$$

This is because

$$Q \setminus Q' = \cup_{i=1}^n a_i \times \prod_{j \neq i} (a_j, b_j]$$

a finite union of sets of content 0. It is obvious $\int \chi_Q dV = v(Q)$ because you can use a grid which has Q as one of the boxes and then the upper and lower sums are the same and equal to $v(Q)$. Therefore, the claim about the equality of the two integrals in 5.16 follows right away from Corollary E.2.11. That $\chi_{Q'}$ is integrable follows from

$$\chi_{Q'} = \chi_Q - \chi_{Q \setminus Q'}$$

and each of the two functions on the right is integrable thanks to Theorem E.2.10.

In fact, here is an interesting version of the Riemann criterion which depends on these half open boxes.

Lemma E.2.12 *Suppose f is a bounded function which equals zero off some bounded set. Then $f \in \mathcal{R}(\mathbb{R}^n)$ if and only if for all $\varepsilon > 0$ there exists a grid, $\mathcal{G}$ such that*

$$\sum_{Q \in \mathcal{G}} (M_{Q'}(f) - m_{Q'}(f)) v(Q) < \varepsilon. \quad (5.17)$$

Proof: Since $Q' \subseteq Q$,

$$M_{Q'}(f) - m_{Q'}(f) \leq M_Q(f) - m_Q(f)$$

and therefore, the only if part of the equivalence is obvious.

Conversely, let $\mathcal{G}$ be a grid such that 5.17 holds with ε replaced with $\frac{\varepsilon}{2}$. It is necessary to show that there is a grid such that 5.17 holds with no primes on the Q . Let $\mathcal{F}$ be a

refinement of $\mathcal{G}$ obtained by adding the points $\alpha_k^i + \eta_k$ where $\eta_k \leq \eta$ and is also chosen so small that for each $i = 1, \dots, n$,

$$\alpha_k^i + \eta_k < \alpha_{k+1}^i.$$

You only need to have $\eta_k > 0$ for the finitely many boxes of $\mathcal{G}$ which intersect the bounded set where f is not zero. Then for

$$Q \equiv \prod_{i=1}^n [\alpha_{k_i}^i, \alpha_{k_i+1}^i] \in \mathcal{G},$$

Let

$$\hat{Q} \equiv \prod_{i=1}^n [\alpha_{k_i}^i + \eta_{k_i}, \alpha_{k_i+1}^i]$$

and denote by $\hat{\mathcal{G}}$ the collection of these smaller boxes. For each set, Q in $\mathcal{G}$ there is the smaller set, $\hat{Q}$ along with n boxes, $B_k, k = 1, \dots, n$, one of whose sides is of length η_k and the remainder of whose sides are shorter than the diameter of Q such that the set, Q is the union of $\hat{Q}$ and these sets, B_k . Now suppose f equals zero off the ball $B(\mathbf{0}, \frac{R}{2})$. Then without loss of generality, you may assume the diameter of every box in $\mathcal{G}$ which has nonempty intersection with $B(\mathbf{0}, R)$ is smaller than $\frac{R}{3}$. (If this is not so, simply refine $\mathcal{G}$ to make it so, such a refinement leaving 5.17 valid because refinements do not increase the difference between upper and lower sums in this context either.) Suppose there are P sets of $\mathcal{G}$ contained in $B(\mathbf{0}, R)$ (So these are the only sets of $\mathcal{G}$ which could have nonempty intersection with the set where f is nonzero.) and suppose that for all $\mathbf{x}$, $|f(\mathbf{x})| < C/2$. Then

$$\begin{aligned} \sum_{Q \in \mathcal{F}} (M_Q(f) - m_Q(f)) v(Q) &\leq \sum_{\hat{Q} \in \hat{\mathcal{G}}} (M_{\hat{Q}}(f) - m_{\hat{Q}}(f)) v(Q) \\ &+ \sum_{Q \in \mathcal{F} \setminus \hat{\mathcal{G}}} (M_Q(f) - m_Q(f)) v(Q) \end{aligned}$$

The first term on the right of the inequality in the above is no larger than $\varepsilon/2$ because $M_{\hat{Q}}(f) - m_{\hat{Q}}(f) \leq M_{Q'}(f) - m_{Q'}(f)$ for each Q . Therefore, the above is dominated by

$$\leq \varepsilon/2 + CPnR^{n-1}\eta < \varepsilon$$

whenever η is small enough. Since ε is arbitrary, $f \in \mathcal{R}(\mathbb{R}^n)$ as claimed.

E.3 Iterated Integrals

To evaluate an n dimensional Riemann integral, one uses iterated integrals. Formally, an iterated integral is defined as follows. For f a function defined on $\mathbb{R}^{n+m}$,

$$\mathbf{y} \rightarrow f(\mathbf{x}, \mathbf{y})$$

is a function of $\mathbf{y}$ for each $\mathbf{x} \in \mathbb{R}^n$. Therefore, it might be possible to integrate this function of $\mathbf{y}$ and write

$$\int_{\mathbb{R}^m} f(\mathbf{x}, \mathbf{y}) dV_{\mathbf{y}}.$$

Now the result is clearly a function of $\mathbf{x}$ and so, it might be possible to integrate this and write

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^m} f(\mathbf{x}, \mathbf{y}) dV_{\mathbf{y}} dV_{\mathbf{x}}.$$

This symbol is called an iterated integral, because it involves the iteration of two lower dimensional integrations. Under what conditions are the two iterated integrals equal to the integral

$$\int_{\mathbb{R}^{n+m}} f(\mathbf{z}) \, dV?$$

Definition E.3.1 Let $\mathcal{G}$ be a grid on $\mathbb{R}^{n+m}$ defined by the $n+m$ sequences,

$$\{\alpha_k^i\}_{k=-\infty}^{\infty} \quad i = 1, \dots, n+m.$$

Let $\mathcal{G}_n$ be the grid on $\mathbb{R}^n$ obtained by considering only the first n of these sequences and let $\mathcal{G}_m$ be the grid on $\mathbb{R}^m$ obtained by considering only the last m of the sequences. Thus a typical box in $\mathcal{G}_m$ would be

$$\prod_{i=n+1}^{n+m} [\alpha_{k_i}^i, \alpha_{k_i+1}^i], \quad k_i \geq n+1$$

and a box in $\mathcal{G}_n$ would be of the form

$$\prod_{i=1}^n [\alpha_{k_i}^i, \alpha_{k_i+1}^i], \quad k_i \leq n.$$

Lemma E.3.2 Let $\mathcal{G}$, $\mathcal{G}_n$, and $\mathcal{G}_m$ be the grids defined above. Then

$$\mathcal{G} = \{R \times P : R \in \mathcal{G}_n \text{ and } P \in \mathcal{G}_m\}.$$

Proof: If $Q \in \mathcal{G}$, then Q is clearly of this form. On the other hand, if $R \times P$ is one of the sets described above, then from the above description of R and P , it follows $R \times P$ is one of the sets of $\mathcal{G}$. This proves the lemma.

Now let $\mathcal{G}$ be a grid on $\mathbb{R}^{n+m}$ and suppose

$$\phi(\mathbf{z}) = \sum_{Q \in \mathcal{G}} \phi_Q \mathcal{X}_{Q'}(\mathbf{z}) \quad (5.18)$$

where ϕ_Q equals zero for all but finitely many Q . Thus ϕ is a step function. Recall that for

$$Q = \prod_{i=1}^{n+m} [a_i, b_i], \quad Q' \equiv \prod_{i=1}^{n+m} (a_i, b_i]$$

The function

$$\phi = \sum_{Q \in \mathcal{G}} \phi_Q \mathcal{X}_{Q'}$$

is integrable because it is a finite sum of integrable functions, each function in the sum being integrable because the set of discontinuities has Jordan content 0. (why?) Letting $(\mathbf{x}, \mathbf{y}) = \mathbf{z}$,

$$\begin{aligned} \phi(\mathbf{z}) &= \phi(\mathbf{x}, \mathbf{y}) = \sum_{R \in \mathcal{G}_n} \sum_{P \in \mathcal{G}_m} \phi_{R \times P} \mathcal{X}_{R' \times P'}(\mathbf{x}, \mathbf{y}) \\ &= \sum_{R \in \mathcal{G}_n} \sum_{P \in \mathcal{G}_m} \phi_{R \times P} \mathcal{X}_{R'}(\mathbf{x}) \mathcal{X}_{P'}(\mathbf{y}). \end{aligned} \quad (5.19)$$

For a function of two variables, h , denote by $h(\cdot, \mathbf{y})$ the function, $\mathbf{x} \rightarrow h(\mathbf{x}, \mathbf{y})$ and $h(\mathbf{x}, \cdot)$ the function $\mathbf{y} \rightarrow h(\mathbf{x}, \mathbf{y})$. The following lemma is a preliminary version of Fubini's theorem.

Lemma E.3.3 *Let ϕ be a step function as described in 5.18. Then*

$$\phi(\mathbf{x}, \cdot) \in \mathcal{R}(\mathbb{R}^m), \quad (5.20)$$

$$\int_{\mathbb{R}^m} \phi(\cdot, \mathbf{y}) dV_y \in \mathcal{R}(\mathbb{R}^n), \quad (5.21)$$

and

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \phi(\mathbf{x}, \mathbf{y}) dV_y dV_x = \int_{\mathbb{R}^{n+m}} \phi(\mathbf{z}) dV. \quad (5.22)$$

Proof: To verify 5.20, note that $\phi(\mathbf{x}, \cdot)$ is the step function

$$\phi(\mathbf{x}, \mathbf{y}) = \sum_{P \in \mathcal{G}_m} \phi_{R \times P} \mathcal{X}_{P'}(\mathbf{y}).$$

Where $\mathbf{x} \in R'$ and this is a finite sum of integrable functions because each has set of discontinuities with Jordan content 0. From the description in 5.19,

$$\begin{aligned} \int_{\mathbb{R}^m} \phi(\mathbf{x}, \mathbf{y}) dV_y &= \sum_{R \in \mathcal{G}_n} \sum_{P \in \mathcal{G}_m} \phi_{R \times P} \mathcal{X}_{R'}(\mathbf{x}) v(P) \\ &= \sum_{R \in \mathcal{G}_n} \left(\sum_{P \in \mathcal{G}_m} \phi_{R \times P} v(P) \right) \mathcal{X}_{R'}(\mathbf{x}), \end{aligned} \quad (5.23)$$

another step function. Therefore,

$$\begin{aligned} \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \phi(\mathbf{x}, \mathbf{y}) dV_y dV_x &= \sum_{R \in \mathcal{G}_n} \sum_{P \in \mathcal{G}_m} \phi_{R \times P} v(P) v(R) \\ &= \sum_{Q \in \mathcal{G}} \phi_Q v(Q) = \int_{\mathbb{R}^{n+m}} \phi(\mathbf{z}) dV. \end{aligned}$$

and this proves the lemma.

From 5.23,

$$\begin{aligned} M_{R'_1} \left(\int_{\mathbb{R}^m} \phi(\cdot, \mathbf{y}) dV_y \right) &\equiv \sup \left\{ \sum_{R \in \mathcal{G}_n} \left(\sum_{P \in \mathcal{G}_m} \phi_{R \times P} v(P) \right) \mathcal{X}_{R'}(\mathbf{x}) : \mathbf{x} \in R'_1 \right\} \\ &= \sum_{P \in \mathcal{G}_m} \phi_{R_1 \times P} v(P) \end{aligned} \quad (5.24)$$

because $\int_{\mathbb{R}^m} \phi(\cdot, \mathbf{y}) dV_y$ has the constant value given in 5.24 for $\mathbf{x} \in R'_1$. Similarly,

$$\begin{aligned} m_{R'_1} \left(\int_{\mathbb{R}^m} \phi(\cdot, \mathbf{y}) dV_y \right) &\equiv \inf \left\{ \sum_{R \in \mathcal{G}_n} \left(\sum_{P \in \mathcal{G}_m} \phi_{R \times P} v(P) \right) \mathcal{X}_{R'}(\mathbf{x}) : \mathbf{x} \in R'_1 \right\} \\ &= \sum_{P \in \mathcal{G}_m} \phi_{R_1 \times P} v(P). \end{aligned} \quad (5.25)$$

Theorem E.3.4 (Fubini) *Let $f \in \mathcal{R}(\mathbb{R}^{n+m})$ and suppose also that $f(\mathbf{x}, \cdot) \in \mathcal{R}(\mathbb{R}^m)$ for each $\mathbf{x}$. Then*

$$\int_{\mathbb{R}^m} f(\cdot, \mathbf{y}) dV_y \in \mathcal{R}(\mathbb{R}^n) \quad (5.26)$$

and

$$\int_{\mathbb{R}^{n+m}} f(\mathbf{z}) dV = \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} f(\mathbf{x}, \mathbf{y}) dV_y dV_x. \quad (5.27)$$

Proof: Let $\mathcal{G}$ be a grid such that $\mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{G}}(f) < \varepsilon$ and let $\mathcal{G}_n$ and $\mathcal{G}_m$ be as defined above. Let

$$\phi(\mathbf{z}) \equiv \sum_{Q \in \mathcal{G}} M_{Q'}(f) \mathcal{X}_{Q'}(\mathbf{z}), \quad \psi(\mathbf{z}) \equiv \sum_{Q \in \mathcal{G}} m_{Q'}(f) \mathcal{X}_{Q'}(\mathbf{z}).$$

Observe that $M_{Q'}(f) \leq M_Q(f)$ and $m_{Q'}(f) \geq m_Q(f)$. Then

$$\mathcal{U}_{\mathcal{G}}(f) \geq \int \phi dV, \quad \mathcal{L}_{\mathcal{G}}(f) \leq \int \psi dV.$$

Also $f(\mathbf{z}) \in (\psi(\mathbf{z}), \phi(\mathbf{z}))$ for all $\mathbf{z}$. Thus from 5.24,

$$M_{R'} \left(\int_{\mathbb{R}^m} f(\cdot, \mathbf{y}) dV_y \right) \leq M_{R'} \left(\int_{\mathbb{R}^m} \phi(\cdot, \mathbf{y}) dV_y \right) = \sum_{P \in \mathcal{G}_m} M_{R' \times P'}(f) v(P)$$

and from 5.25,

$$m_{R'} \left(\int_{\mathbb{R}^m} f(\cdot, \mathbf{y}) dV_y \right) \geq m_{R'} \left(\int_{\mathbb{R}^m} \psi(\cdot, \mathbf{y}) dV_y \right) = \sum_{P \in \mathcal{G}_m} m_{R' \times P'}(f) v(P).$$

Therefore,

$$\begin{aligned} \sum_{R \in \mathcal{G}_n} \left[M_{R'} \left(\int_{\mathbb{R}^m} f(\cdot, \mathbf{y}) dV_y \right) - m_{R'} \left(\int_{\mathbb{R}^m} f(\cdot, \mathbf{y}) dV_y \right) \right] v(R) &\leq \\ \sum_{R \in \mathcal{G}_n} \sum_{P \in \mathcal{G}_m} [M_{R' \times P'}(f) - m_{R' \times P'}(f)] v(P) v(R) &\leq \mathcal{U}_{\mathcal{G}}(f) - \mathcal{L}_{\mathcal{G}}(f) < \varepsilon. \end{aligned}$$

This shows, from Lemma E.2.12 and the Riemann criterion, that $\int_{\mathbb{R}^m} f(\cdot, \mathbf{y}) dV_y \in \mathcal{R}(\mathbb{R}^n)$. It remains to verify 5.27. First note

$$\int_{\mathbb{R}^{n+m}} f(\mathbf{z}) dV \in [\mathcal{L}_{\mathcal{G}}(f), \mathcal{U}_{\mathcal{G}}(f)].$$

Next,

$$\begin{aligned} \mathcal{L}_{\mathcal{G}}(f) &\leq \int_{\mathbb{R}^{n+m}} \psi dV = \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \psi dV_y dV_x \leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} f(\mathbf{x}, \mathbf{y}) dV_y dV_x \\ &\leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \phi(\mathbf{x}, \mathbf{y}) dV_y dV_x = \int_{\mathbb{R}^{n+m}} \phi dV \leq \mathcal{U}_{\mathcal{G}}(f). \end{aligned}$$

Therefore,

$$\left| \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} f(\mathbf{x}, \mathbf{y}) dV_y dV_x - \int_{\mathbb{R}^{n+m}} f(\mathbf{z}) dV \right| \leq \varepsilon$$

and since $\varepsilon > 0$ is arbitrary, this proves Fubini's theorem².

Corollary E.3.5 Suppose E is a bounded contented set in $\mathbb{R}^n$ and let ϕ, ψ be continuous functions defined on E such that $\phi(\mathbf{x}) \geq \psi(\mathbf{x})$. Also suppose f is a continuous bounded function defined on the set,

$$P \equiv \{(\mathbf{x}, y) : \psi(\mathbf{x}) \leq y \leq \phi(\mathbf{x})\},$$

It follows $f\mathcal{X}_P \in \mathcal{R}(\mathbb{R}^{n+1})$ and

$$\int_P f dV = \int_E \int_{\psi(\mathbf{x})}^{\phi(\mathbf{x})} f(\mathbf{x}, y) dy dV_x.$$

²Actually, Fubini's theorem usually refers to a much more profound result in the theory of Lebesgue integration.

Proof: Since f is continuous, there is no problem in writing $f(\mathbf{x}, \cdot) \mathcal{X}_{[\psi(\mathbf{x}), \phi(\mathbf{x})]}(\cdot) \in \mathcal{R}(\mathbb{R}^1)$. Also, $f\mathcal{X}_P \in \mathcal{R}(\mathbb{R}^{n+1})$ because P is contented thanks to Corollary E.2.9. Therefore, by Fubini's theorem

$$\begin{aligned} \int_P f dV &= \int_{\mathbb{R}^n} \int_{\mathbb{R}} f \mathcal{X}_P dy dV_x \\ &= \int_E \int_{\psi(\mathbf{x})}^{\phi(\mathbf{x})} f(\mathbf{x}, y) dy dV_x \end{aligned}$$

proving the corollary.

Other versions of this corollary are immediate and should be obvious whenever encountered.

E.4 The Change Of Variables Formula

First recall Theorem D.4.2 on Page 669 which is listed here for convenience.

Theorem E.4.1 *Let $\mathbf{h} : U \rightarrow \mathbb{R}^n$ be a C^1 function with $\mathbf{h}(\mathbf{0}) = \mathbf{0}, D\mathbf{h}(\mathbf{0})^{-1}$ exists. Then there exists an open set, $V \subseteq U$ containing $\mathbf{0}$, flips, $\mathbf{F}_1, \dots, \mathbf{F}_{n-1}$, and primitive functions, $\mathbf{G}_n, \mathbf{G}_{n-1}, \dots, \mathbf{G}_1$ such that for $\mathbf{x} \in V$,*

$$\mathbf{h}(\mathbf{x}) = \mathbf{F}_1 \circ \dots \circ \mathbf{F}_{n-1} \circ \mathbf{G}_n \circ \mathbf{G}_{n-1} \circ \dots \circ \mathbf{G}_1(\mathbf{x}).$$

Also recall Theorem 18.6.12 on Page 451.

Theorem E.4.2 *Let $\phi : [a, b] \rightarrow [c, d]$ be one to one and suppose ϕ' exists and is continuous on $[a, b]$. Then if f is a continuous function defined on $[a, b]$,*

$$\int_c^d f(s) ds = \int_a^b f(\phi(t)) |\phi'(t)| dt$$

The following is a simple corollary to this theorem.

Corollary E.4.3 *Let $\phi : [a, b] \rightarrow [c, d]$ be one to one and suppose ϕ' exists and is continuous on $[a, b]$. Then if f is a continuous function defined on $[a, b]$,*

$$\int_{\mathbb{R}} \mathcal{X}_{[a, b]}(\phi^{-1}(x)) f(x) dx = \int_{\mathbb{R}} \mathcal{X}_{[a, b]}(t) f(\phi(t)) |\phi'(t)| dt$$

Lemma E.4.4 *Let $\mathbf{h} : V \rightarrow \mathbb{R}^n$ be a C^1 function and suppose H is a compact subset of V . Then there exists a constant, C independent of $\mathbf{x} \in H$ such that*

$$|D\mathbf{h}(\mathbf{x}) \mathbf{v}| \leq C |\mathbf{v}|.$$

Proof: Consider the compact set, $H \times \partial B(\mathbf{0}, 1) \subseteq \mathbb{R}^{2n}$. Let $f : H \times \partial B(\mathbf{0}, 1) \rightarrow \mathbb{R}$ be given by $f(\mathbf{x}, \mathbf{v}) = |D\mathbf{h}(\mathbf{x}) \mathbf{v}|$. Then let C denote the maximum value of f . It follows that for $\mathbf{v} \in \mathbb{R}^n$,

$$\left| D\mathbf{h}(\mathbf{x}) \frac{\mathbf{v}}{|\mathbf{v}|} \right| \leq C$$

and so the desired formula follows when you multiply both sides by $|\mathbf{v}|$.

Definition E.4.5 *Let A be an open set. Write $C^k(A; \mathbb{R}^n)$ to denote a C^k function whose domain is A and whose range is in $\mathbb{R}^n$. Let U be an open set in $\mathbb{R}^n$. Then $\mathbf{h} \in C^k(\overline{U}; \mathbb{R}^n)$ if there exists an open set, $V \supseteq \overline{U}$ and a function, $\mathbf{g} \in C^1(V; \mathbb{R}^n)$ such that $\mathbf{g} = \mathbf{h}$ on $\overline{U}$. $f \in C^k(\overline{U})$ means the same thing except that f has values in $\mathbb{R}$.*

Theorem E.4.6 *Let U be a bounded open set such that ∂U has zero content and let $\mathbf{h} \in C(\bar{U}; \mathbb{R}^n)$ be one to one and $D\mathbf{h}(\mathbf{x})^{-1}$ exists for all $\mathbf{x} \in U$. Then $\mathbf{h}(\partial U) = \partial(\mathbf{h}(U))$ and $\partial(\mathbf{h}(U))$ has zero content.*

Proof: Let $\mathbf{x} \in \partial U$ and let $\mathbf{g} = \mathbf{h}$ where $\mathbf{g}$ is a C^1 function defined on an open set containing $\bar{U}$. By the inverse function theorem, $\mathbf{g}$ is locally one to one and an open mapping near $\mathbf{x}$. Thus $\mathbf{g}(\mathbf{x}) = \mathbf{h}(\mathbf{x})$ and is in an open set containing points of $\mathbf{g}(U)$ and points of $\mathbf{g}(U^C)$. These points of $\mathbf{g}(U^C)$ cannot equal any points of $\mathbf{h}(U)$ because $\mathbf{g}$ is one to one locally. Thus $\mathbf{h}(\mathbf{x}) \in \partial(\mathbf{h}(U))$ and so $\mathbf{h}(\partial U) \subseteq \partial(\mathbf{h}(U))$. Now suppose $\mathbf{y} \in \partial(\mathbf{h}(U))$. By the inverse function theorem $\mathbf{y}$ cannot be in the open set $\mathbf{h}(U)$. Since $\mathbf{y} \in \partial(\mathbf{h}(U))$, every ball centered at $\mathbf{y}$ contains points of $\mathbf{h}(U)$ and so $\mathbf{y} \in \overline{\mathbf{h}(U)} \setminus \mathbf{h}(U)$. Thus there exists a sequence, $\{\mathbf{x}_n\} \subseteq U$ such that $\mathbf{h}(\mathbf{x}_n) \rightarrow \mathbf{y}$. But then, by the inverse function theorem, $\mathbf{x}_n \rightarrow \mathbf{h}^{-1}(\mathbf{y})$ and so $\mathbf{h}^{-1}(\mathbf{y}) \in \partial U$. Therefore, $\mathbf{y} \in \mathbf{h}(\partial U)$ and this proves the two sets are equal. It remains to verify the claim about content.

First let H denote a compact set whose interior contains $\bar{U}$ which is also in the interior of the domain of $\mathbf{g}$. Now since ∂U has content zero, it follows that for $\varepsilon > 0$ given, there exists a grid, $\mathcal{G}$ such that if $\mathcal{G}'$ are those boxes of $\mathcal{G}$ which have nonempty intersection with ∂U , then

$$\sum_{Q \in \mathcal{G}'} v(Q) < \varepsilon.$$

and by refining the grid if necessary, no box of $\mathcal{G}$ has nonempty intersection with both $\bar{U}$ and H^C . Refining this grid still more, you can also assume that for all boxes in $\mathcal{G}'$,

$$\frac{l_i}{l_j} < 2$$

where l_i is the length of the i^{th} side. (Thus the boxes are not too far from being cubes.)

Let C be the constant of Lemma E.4.4 applied to $\mathbf{g}$ on H .

Now consider one of these boxes, $Q \in \mathcal{G}'$. If $\mathbf{x}, \mathbf{y} \in Q$, it follows from the chain rule that

$$\mathbf{g}(\mathbf{y}) - \mathbf{g}(\mathbf{x}) = \int_0^1 D\mathbf{g}(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) dt$$

By Lemma E.4.4 applied to H

$$\begin{aligned} |\mathbf{g}(\mathbf{y}) - \mathbf{g}(\mathbf{x})| &\leq \int_0^1 |D\mathbf{g}(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x})| dt \\ &\leq C \int_0^1 |\mathbf{x} - \mathbf{y}| dt \leq C \text{diam}(Q) \\ &= C \left(\sum_{i=1}^n l_i^2 \right)^{1/2} \leq C\sqrt{n}L \end{aligned}$$

where L is the length of the longest side of Q . Thus $\text{diam}(\mathbf{g}(Q)) \leq C\sqrt{n}L$ and so $\mathbf{g}(Q)$ is contained in a cube having sides equal to $C\sqrt{n}L$ and volume equal to

$$C^n n^{n/2} L^n \leq C^n n^{n/2} 2^n l_1 l_2 \cdots l_n = C^n n^{n/2} 2^n v(Q).$$

Denoting by P_Q this cube, it follows

$$\mathbf{h}(\partial U) \subseteq \cup_{Q \in \mathcal{G}'} v(P_Q)$$

and

$$\sum_{Q \in \mathcal{G}'} v(P_Q) \leq C^n n^{n/2} 2^n \sum_{Q \in \mathcal{G}'} v(Q) < \varepsilon C^n n^{n/2} 2^n.$$

Since $\varepsilon > 0$ is arbitrary, this shows $\mathbf{h}(\partial U)$ has content zero as claimed.

Theorem E.4.7 Suppose $f \in C(\overline{U})$ where U is a bounded open set with ∂U having content 0. Then $f\chi_U \in \mathcal{R}(\mathbb{R}^n)$.

Proof: Let H be a compact set whose interior contains $\overline{U}$ which is also contained in the domain of g where g is a continuous functions whose restriction to U equals f . Consider $g\chi_U$, a function whose set of discontinuities has content 0. Then $g\chi_U = f\chi_U \in \mathcal{R}(\mathbb{R}^n)$ as claimed. This is by the big theorem which tells which functions are Riemann integrable.

The following lemma is obvious from the definition of the integral.

Lemma E.4.8 Let U be a bounded open set and let $f\chi_U \in \mathcal{R}(\mathbb{R}^n)$. Then

$$\int f(\mathbf{x} + \mathbf{p})\chi_{U-\mathbf{p}}(\mathbf{x})dx = \int f(\mathbf{x})\chi_U(\mathbf{x})dx$$

A few more lemmas are needed.

Lemma E.4.9 Let S be a nonempty subset of $\mathbb{R}^n$. Define

$$f(\mathbf{x}) \equiv \text{dist}(\mathbf{x}, S) \equiv \inf \{|\mathbf{x} - \mathbf{y}| : \mathbf{y} \in S\}.$$

Then f is continuous.

Proof: Consider $|f(\mathbf{x}) - f(\mathbf{x}_1)|$ and suppose without loss of generality that $f(\mathbf{x}_1) \geq f(\mathbf{x})$. Then choose $\mathbf{y} \in S$ such that $f(\mathbf{x}) + \varepsilon > |\mathbf{x} - \mathbf{y}|$. Then

$$\begin{aligned} |f(\mathbf{x}_1) - f(\mathbf{x})| &= f(\mathbf{x}_1) - f(\mathbf{x}) \leq f(\mathbf{x}_1) - |\mathbf{x} - \mathbf{y}| + \varepsilon \\ &\leq |\mathbf{x}_1 - \mathbf{y}| - |\mathbf{x} - \mathbf{y}| + \varepsilon \\ &\leq |\mathbf{x} - \mathbf{x}_1| + |\mathbf{x} - \mathbf{y}| - |\mathbf{x} - \mathbf{y}| + \varepsilon \\ &= |\mathbf{x} - \mathbf{x}_1| + \varepsilon. \end{aligned}$$

Since ε is arbitrary, it follows that $|f(\mathbf{x}_1) - f(\mathbf{x})| \leq |\mathbf{x} - \mathbf{x}_1|$ and this proves the lemma.

Theorem E.4.10 (Urysohn's lemma for $\mathbb{R}^n$) Let H be a closed subset of an open set, U . Then there exists a continuous function, $g : \mathbb{R}^n \rightarrow [0, 1]$ such that $g(\mathbf{x}) = 1$ for all $\mathbf{x} \in H$ and $g(\mathbf{x}) = 0$ for all $\mathbf{x} \notin U$.

Proof: If $\mathbf{x} \notin C$, a closed set, then $\text{dist}(\mathbf{x}, C) > 0$ because if not, there would exist a sequence of points of C converging to $\mathbf{x}$ and it would follow that $\mathbf{x} \in C$. Therefore, $\text{dist}(\mathbf{x}, H) + \text{dist}(\mathbf{x}, U^C) > 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Now define a continuous function, g as

$$g(\mathbf{x}) \equiv \frac{\text{dist}(\mathbf{x}, U^C)}{\text{dist}(\mathbf{x}, H) + \text{dist}(\mathbf{x}, U^C)}.$$

It is easy to see this verifies the conclusions of the theorem and this proves the theorem.

Definition E.4.11 Define $\text{spt}(f)$ (support of f) to be the closure of the set $\{x : f(x) \neq 0\}$. If V is an open set, $C_c(V)$ will be the set of continuous functions f , defined on $\mathbb{R}^n$ having $\text{spt}(f) \subseteq V$.

Definition E.4.12 If K is a compact subset of an open set, V , then $K \prec \phi \prec V$ if

$$\phi \in C_c(V), \phi(K) = \{1\}, \phi(\mathbb{R}^n) \subseteq [0, 1].$$

Also for $\phi \in C_c(\mathbb{R}^n)$, $K \prec \phi$ if

$$\phi(\mathbb{R}^n) \subseteq [0, 1] \text{ and } \phi(K) = 1.$$

and $\phi \prec V$ if

$$\phi(\mathbb{R}^n) \subseteq [0, 1] \text{ and } \text{spt}(\phi) \subseteq V.$$

Theorem E.4.13 (*Partition of unity*) Let K be a compact subset of $\mathbb{R}^n$ and suppose

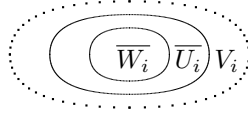
$$K \subseteq V = \cup_{i=1}^n V_i, \quad V_i \text{ open.}$$

Then there exist $\psi_i \prec V_i$ with

$$\sum_{i=1}^n \psi_i(\mathbf{x}) = 1$$

for all $\mathbf{x} \in K$.

Proof: Let $K_1 = K \setminus \cup_{i=2}^n V_i$. Thus K_1 is compact because it is the intersection of a closed set with a compact set and $K_1 \subseteq V_1$. Let $K_1 \subseteq W_1 \subseteq \overline{W}_1 \subseteq V_1$ with $\overline{W}_1$ compact. To obtain W_1 , use Theorem E.4.10 to get f such that $K_1 \prec f \prec V_1$ and let $W_1 \equiv \{\mathbf{x} : f(\mathbf{x}) \neq 0\}$. Thus $W_1, V_2, \dots, V_n$ covers K and $\overline{W}_1 \subseteq V_1$. Let $K_2 = K \setminus (\cup_{i=3}^n V_i \cup W_1)$. Then K_2 is compact and $K_2 \subseteq V_2$. Let $K_2 \subseteq W_2 \subseteq \overline{W}_2 \subseteq V_2$ with $\overline{W}_2$ compact. Continue this way finally obtaining $W_1, \dots, W_n$, $K \subseteq W_1 \cup \dots \cup W_n$, and $\overline{W}_i \subseteq V_i$, $\overline{W}_i$ compact. Now let $\overline{W}_i \subseteq U_i \subseteq \overline{U}_i \subseteq V_i$, $\overline{U}_i$ compact.



By Theorem E.4.10, there exist functions, ϕ_i, γ such that $\overline{U}_i \prec \phi_i \prec V_i$, $\cup_{i=1}^n \overline{W}_i \prec \gamma \prec \cup_{i=1}^n U_i$. Define

$$\psi_i(\mathbf{x}) = \begin{cases} \gamma(\mathbf{x})\phi_i(\mathbf{x}) / \sum_{j=1}^n \phi_j(\mathbf{x}) & \text{if } \sum_{j=1}^n \phi_j(\mathbf{x}) \neq 0, \\ 0 & \text{if } \sum_{j=1}^n \phi_j(\mathbf{x}) = 0. \end{cases}$$

If $\mathbf{x}$ is such that $\sum_{j=1}^n \phi_j(\mathbf{x}) = 0$, then $\mathbf{x} \notin \cup_{i=1}^n \overline{U}_i$. Consequently $\gamma(\mathbf{y}) = 0$ for all $\mathbf{y}$ near $\mathbf{x}$ and so $\psi_i(\mathbf{y}) = 0$ for all $\mathbf{y}$ near $\mathbf{x}$. Hence ψ_i is continuous at such $\mathbf{x}$. If $\sum_{j=1}^n \phi_j(\mathbf{x}) \neq 0$, this situation persists near $\mathbf{x}$ and so ψ_i is continuous at such points. Therefore ψ_i is continuous. If $\mathbf{x} \in K$, then $\gamma(\mathbf{x}) = 1$ and so $\sum_{j=1}^n \psi_j(\mathbf{x}) = 1$. Clearly $0 \leq \psi_i(\mathbf{x}) \leq 1$ and $\text{spt}(\psi_j) \subseteq V_j$. This proves the theorem.

The next lemma contains the main ideas.

Lemma E.4.14 Let U be a bounded open set with ∂U having content 0. Also let $\mathbf{h} \in C^1(\overline{U}; \mathbb{R}^n)$ be one to one on U with $D\mathbf{h}(\mathbf{x})^{-1}$ exists for all $\mathbf{x} \in U$. Let $f \in C(\overline{U})$ be nonnegative. Then

$$\int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f(\mathbf{z}) dV_n = \int \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dV_n$$

Proof: Let $\varepsilon > 0$ be given. Then by Theorem E.4.7,

$$\mathbf{x} \rightarrow \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})|$$

is Riemann integrable. Therefore, there exists a grid, $\mathcal{G}$ such that, letting

$$g(\mathbf{x}) = \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})|,$$

$$\mathcal{L}_{\mathcal{G}}(g) + \varepsilon > \mathcal{U}_{\mathcal{G}}(g).$$

Let K denote the union of the boxes, Q of $\mathcal{G}$ for which $m_Q(g) > 0$. Thus K is a compact subset of U and it is only the terms from these boxes which contribute anything nonzero

to the lower sum. By Theorem D.4.2 on Page 669 which is stated above, it follows that for $\mathbf{p} \in K$, there exists an open set contained in U which contains $\mathbf{p}, O_{\mathbf{p}}$ such that for $\mathbf{x} \in O_{\mathbf{p}} - \mathbf{p}$,

$$\mathbf{h}(\mathbf{x} + \mathbf{p}) - \mathbf{h}(\mathbf{p}) = \mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1} \circ \mathbf{G}_n \circ \cdots \circ \mathbf{G}_1(\mathbf{x})$$

where the $\mathbf{G}_i$ are primitive functions and the $\mathbf{F}_j$ are flips. Finitely many of these open sets, $\{O_j\}_{j=1}^q$ cover K . Let the distinguished point for O_j be denoted by $\mathbf{p}_j$. Now refine $\mathcal{G}$ if necessary such that the diameter of every cell of the new $\mathcal{G}$ which intersects U is smaller than a Lebesgue number for this open cover. Denote by $\mathcal{G}'$ those boxes of $\mathcal{G}$ whose union equals the set, K . Thus every box of $\mathcal{G}'$ is contained in one of these O_j . By Theorem E.4.13 there exists a partition of unity, $\{\psi_j\}$ on $\mathbf{h}(K)$ such that $\psi_j \prec \mathbf{h}(O_j)$. Then

$$\begin{aligned} \mathcal{L}_{\mathcal{G}}(g) &\leq \sum_{Q \in \mathcal{G}'} \int \mathcal{X}_Q(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx \\ &= \sum_{Q \in \mathcal{G}'} \sum_{j=1}^q \int \mathcal{X}_Q(\mathbf{x}) (\psi_j f)(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx. \end{aligned} \quad (5.28)$$

Consider the term $\int \mathcal{X}_Q(\mathbf{x}) (\psi_j f)(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx$. By Lemma E.4.8 and Fubini's theorem this equals

$$\begin{aligned} &\int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \mathcal{X}_{Q-\mathbf{p}_j}(\mathbf{x}) (\psi_j f)(\mathbf{h}(\mathbf{p}_i) + \mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1} \circ \mathbf{G}_n \circ \cdots \circ \mathbf{G}_1(\mathbf{x})) \cdot \\ &|D\mathbf{F}(\mathbf{G}_n \circ \cdots \circ \mathbf{G}_1(\mathbf{x}))| |D\mathbf{G}_n(\mathbf{G}_{n-1} \circ \cdots \circ \mathbf{G}_1(\mathbf{x}))| \cdot \\ &|D\mathbf{G}_{n-1}(\mathbf{G}_{n-2} \circ \cdots \circ \mathbf{G}_1(\mathbf{x}))| \cdot \end{aligned} \quad (5.29)$$

$$\cdots |D\mathbf{G}_2(\mathbf{G}_1(\mathbf{x}))| |D\mathbf{G}_1(\mathbf{x})| dx_1 dV_{n-1}. \quad (5.30)$$

Here dV_{n-1} is with respect to the variables, $x_2, \dots, x_n$. Also $\mathbf{F}$ denotes $\mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1}$. Now

$$\mathbf{G}_1(\mathbf{x}) = (\alpha(\mathbf{x}), x_2, \dots, x_n)^T$$

and is one to one. Therefore, fixing $x_2, \dots, x_n$, $x_1 \rightarrow \alpha(\mathbf{x})$ is one to one. Also, $D\mathbf{G}_1(\mathbf{x}) = \frac{\partial \alpha}{\partial x_1}(\mathbf{x})$. Fixing $x_2, \dots, x_n$, change the variable,

$$y_1 = \alpha(x_1, x_2, \dots, x_n).$$

Thus

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^T = \mathbf{G}_1^{-1}(y_1, x_2, \dots, x_n) \equiv \mathbf{G}_1^{-1}(\mathbf{x}')$$

Then in 5.30 you can use Corollary E.4.3 to write 5.30 as

$$\begin{aligned} &\int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \mathcal{X}_{Q-\mathbf{p}_j}(\mathbf{G}_1^{-1}(\mathbf{x}')) (\psi_j f)(\mathbf{h}(\mathbf{p}_i) + \mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1} \circ \mathbf{G}_n \circ \cdots \circ \mathbf{G}_1(\mathbf{G}_1^{-1}(\mathbf{x}'))) \\ &\cdot |D\mathbf{F}(\mathbf{G}_n \circ \cdots \circ \mathbf{G}_1(\mathbf{G}_1^{-1}(\mathbf{x}')))| |D\mathbf{G}_n(\mathbf{G}_{n-1} \circ \cdots \circ \mathbf{G}_1(\mathbf{G}_1^{-1}(\mathbf{x}')))| \cdot \\ &|D\mathbf{G}_{n-1}(\mathbf{G}_{n-2} \circ \cdots \circ \mathbf{G}_1(\mathbf{G}_1^{-1}(\mathbf{x}')))| \cdots |D\mathbf{G}_2(\mathbf{G}_1(\mathbf{G}_1^{-1}(\mathbf{x}')))| dy_1 dV_{n-1} \end{aligned}$$

which reduces to

$$\begin{aligned} &\int_{\mathbb{R}^n} \mathcal{X}_{Q-\mathbf{p}_j}(\mathbf{G}_1^{-1}(\mathbf{x}')) (\psi_j f)(\mathbf{h}(\mathbf{p}_i) + \mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1} \circ \mathbf{G}_n \circ \cdots \circ \mathbf{G}_2(\mathbf{x}')) \\ &\cdot |D\mathbf{F}(\mathbf{G}_n \circ \cdots \circ \mathbf{G}_2(\mathbf{x}'))| |D\mathbf{G}_n(\mathbf{G}_{n-1} \circ \cdots \circ \mathbf{G}_2(\mathbf{x}'))| \cdot \\ &|D\mathbf{G}_{n-1}(\mathbf{G}_{n-2} \circ \cdots \circ \mathbf{G}_2(\mathbf{x}'))| \cdots |D\mathbf{G}_2(\mathbf{x}')| dV_n. \end{aligned}$$

Now use Fubini's theorem again to make the inside integral taken with respect to x_2 . Exactly the same process yields

$$\begin{aligned} \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} \mathcal{X}_{Q-\mathbf{p}_j} (\mathbf{G}_1^{-1} \circ \mathbf{G}_2^{-1} (\mathbf{x}'')) (\psi_j f) (\mathbf{h}(\mathbf{p}_i) + \mathbf{F}_1 \circ \cdots \circ \mathbf{F}_{n-1} \circ \mathbf{G}_n \circ \cdots \circ \mathbf{G}_3 (\mathbf{x}'')) \\ \cdot |D\mathbf{F}(\mathbf{G}_n \circ \cdots \circ \mathbf{G}_3 (\mathbf{x}''))| |D\mathbf{G}_n(\mathbf{G}_{n-1} \circ \cdots \circ \mathbf{G}_3 (\mathbf{x}''))| \cdot \\ |D\mathbf{G}_{n-1}(\mathbf{G}_{n-2} \circ \cdots \circ \mathbf{G}_3 (\mathbf{x}''))| \cdots dy_2 dV_{n-1}. \end{aligned}$$

Now $\mathbf{F}$ is just a composition of flips and so $|D\mathbf{F}(\mathbf{G}_n \circ \cdots \circ \mathbf{G}_3 (\mathbf{x}''))| = 1$ and so this term can be replaced with 1. Continuing this process, eventually yields an expression of the form

$$\int_{\mathbb{R}^n} \mathcal{X}_{Q-\mathbf{p}_j} (\mathbf{G}_1^{-1} \circ \cdots \circ \mathbf{G}_{n-2}^{-1} \circ \mathbf{G}_{n-1}^{-1} \circ \mathbf{G}_n^{-1} \circ \mathbf{F}^{-1} (\mathbf{y})) (\psi_j f) (\mathbf{h}(\mathbf{p}_i) + \mathbf{y}) dV_n. \quad (5.31)$$

Denoting by $\mathbf{G}^{-1}$ the expression, $\mathbf{G}_1^{-1} \circ \cdots \circ \mathbf{G}_{n-2}^{-1} \circ \mathbf{G}_{n-1}^{-1} \circ \mathbf{G}_n^{-1}$,

$$\mathcal{X}_{Q-\mathbf{p}_j} (\mathbf{G}_1^{-1} \circ \cdots \circ \mathbf{G}_{n-2}^{-1} \circ \mathbf{G}_{n-1}^{-1} \circ \mathbf{G}_n^{-1} \circ \mathbf{F}^{-1} (\mathbf{y})) = 1$$

exactly when $\mathbf{G}^{-1} \circ \mathbf{F}^{-1} (\mathbf{y}) \in Q - \mathbf{p}_j$. Now recall that

$$\mathbf{h}(\mathbf{p}_j + \mathbf{x}) - \mathbf{h}(\mathbf{p}_j) = \mathbf{F} \circ \mathbf{G}(\mathbf{x})$$

and so the above holds exactly when

$$\begin{aligned} \mathbf{y} &= \mathbf{h}(\mathbf{p}_j + \mathbf{G}^{-1} \circ \mathbf{F}^{-1} (\mathbf{y})) - \mathbf{h}(\mathbf{p}_j) \in \mathbf{h}(\mathbf{p}_j + Q - \mathbf{p}_j) - \mathbf{h}(\mathbf{p}_j) \\ &= \mathbf{h}(Q) - \mathbf{h}(\mathbf{p}_j). \end{aligned}$$

Thus 5.31 reduces to

$$\begin{aligned} \int_{\mathbb{R}^n} \mathcal{X}_{\mathbf{h}(Q)-\mathbf{h}(\mathbf{p}_j)} (\mathbf{y}) (\psi_j f) (\mathbf{h}(\mathbf{p}_i) + \mathbf{y}) dV_n \\ = \int_{\mathbb{R}^n} \mathcal{X}_{\mathbf{h}(Q)} (\mathbf{z}) (\psi_j f) (\mathbf{z}) dV_n. \end{aligned}$$

It follows from 5.28

$$\begin{aligned} \mathcal{U}_{\mathcal{G}}(g) - \varepsilon &\leq \mathcal{L}_{\mathcal{G}}(g) \leq \sum_{Q \in \mathcal{G}'} \int \mathcal{X}_Q(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx \\ &= \sum_{Q \in \mathcal{G}'} \sum_{j=1}^q \int \mathcal{X}_Q(\mathbf{x}) (\psi_j f) (\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx \\ &= \sum_{Q \in \mathcal{G}'} \sum_{j=1}^q \int_{\mathbb{R}^n} \mathcal{X}_{\mathbf{h}(Q)} (\mathbf{z}) (\psi_j f) (\mathbf{z}) dV_n \\ &= \sum_{Q \in \mathcal{G}'} \int_{\mathbb{R}^n} \mathcal{X}_{\mathbf{h}(Q)} (\mathbf{z}) f(\mathbf{z}) dV_n \leq \int \mathcal{X}_{\mathbf{h}(U)} (\mathbf{z}) f(\mathbf{z}) dV_n \end{aligned}$$

which implies the inequality,

$$\int \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dV_n \leq \int \mathcal{X}_{\mathbf{h}(U)} (\mathbf{z}) f(\mathbf{z}) dV_n$$

But now you can use the same information just derived to obtain equality.

$$\mathbf{x} = \mathbf{h}^{-1}(\mathbf{z})$$

and so from what was just done,

$$\begin{aligned}
 & \int \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dV_n \\
 &= \int \mathcal{X}_{\mathbf{h}^{-1}(\mathbf{h}(U))}(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dV_n \\
 &\geq \int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f(\mathbf{z}) |\det D\mathbf{h}(\mathbf{h}^{-1}(\mathbf{z}))| |\det D\mathbf{h}^{-1}(\mathbf{z})| dV_n \\
 &= \int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f(\mathbf{z}) dV_n
 \end{aligned}$$

from the chain rule. In fact,

$$I = D\mathbf{h}(\mathbf{h}^{-1}(\mathbf{z})) D\mathbf{h}^{-1}(\mathbf{z})$$

and so

$$1 = |\det D\mathbf{h}(\mathbf{h}^{-1}(\mathbf{z}))| |\det D\mathbf{h}^{-1}(\mathbf{z})|.$$

This proves the lemma.

The change of variables theorem follows.

Theorem E.4.15 *Let U be a bounded open set with ∂U having content 0. Also let $\mathbf{h} \in C^1(\overline{U}; \mathbb{R}^n)$ be one to one on U with $D\mathbf{h}(\mathbf{x})^{-1}$ exists for all $\mathbf{x} \in U$. Let $f \in C(\overline{U})$. Then*

$$\int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f(\mathbf{z}) dz = \int \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx$$

Proof: You note that the formula holds for $f^+ \equiv \frac{|f|+f}{2}$ and $f^- \equiv \frac{|f|-f}{2}$. Now $f = f^+ - f^-$ and so

$$\begin{aligned}
 & \int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f(\mathbf{z}) dz \\
 &= \int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f^+(\mathbf{z}) dz - \int \mathcal{X}_{\mathbf{h}(U)}(\mathbf{z}) f^-(\mathbf{z}) dz \\
 &= \int \mathcal{X}_U(\mathbf{x}) f^+(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx - \int \mathcal{X}_U(\mathbf{x}) f^-(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx \\
 &= \int \mathcal{X}_U(\mathbf{x}) f(\mathbf{h}(\mathbf{x})) |\det D\mathbf{h}(\mathbf{x})| dx.
 \end{aligned}$$

E.5 Some Observations

Some of the above material is very technical. This is because it gives complete answers to the fundamental questions on existence of the integral and related theoretical considerations. However, most of the difficulties are artifacts. They shouldn't even be considered! It was realized early in the twentieth century that these difficulties occur because, from the point of view of mathematics, this is not the right way to define an integral! Better results are obtained much more easily using the Lebesgue integral. Many of the technicalities related to Jordan content disappear almost magically when the right integral is used. However, the Lebesgue integral is more abstract than the Riemann integral and it is not traditional to consider it in a beginning calculus course. If you are interested in the fundamental properties of the integral and the theory behind it, you should abandon the Riemann integral which is an antiquated relic and

begin to study the integral of the last century. An introduction to it is in [23]. Another very good source is [12]. This advanced calculus text does everything in terms of the Lebesgue integral and never bothers to struggle with the inferior Riemann integral. A more general treatment is found in [18], [19], [24], and [20]. There is also a still more general integral called the generalized Riemann integral. A recent book on this subject is [5]. It is far easier to define than the Lebesgue integral but the convergence theorems are much harder to prove. An introduction is also in [19].

Volumes In Higher Dimensions*

F.1 p Dimensional Surfaces*

Everything which was done for 2 dimensional surfaces in $\mathbb{R}^3$ can be done for more general situations. The main change is in the level of generality. You need to first define carefully what is meant by the area of such a surface. The concept of a p dimensional surface in $\mathbb{R}^n$ depends on the following fundamental definition on Page 394. To understand what is going on in this section you must read that material as well as the chapter on spectral theory.

Definition F.1.1 Let $\mathbf{u}_1, \dots, \mathbf{u}_p$ be vectors in $\mathbb{R}^n$. The p dimensional parallelepiped determined by these vectors will be denoted by $P(\mathbf{u}_1, \dots, \mathbf{u}_p)$ and it is defined as

$$P(\mathbf{u}_1, \dots, \mathbf{u}_p) \equiv \left\{ \sum_{j=1}^p s_j \mathbf{u}_j : s_j \in [0, 1] \right\}.$$

Define the volume of this parallelepiped by

$$\text{volume of } P(\mathbf{u}_1, \dots, \mathbf{u}_p) \equiv (\det(\mathbf{u}_i \cdot \mathbf{u}_j))^{1/2}.$$

Suppose then that $\mathbf{x} = \mathbf{f}(\mathbf{u})$ where $\mathbf{u} \in U$, a subset of $\mathbb{R}^p$ and $\mathbf{x}$ is a point in V , a subset of n dimensional space where $n \geq p$. Thus, letting the Cartesian coordinates of $\mathbf{x}$ be given by $\mathbf{x} = (x_1, \dots, x_n)^T$, each x_i being a function of $\mathbf{u}$, an infinitesimal box located at $\mathbf{u}_0$ corresponds to an infinitesimal parallelepiped located at $\mathbf{f}(\mathbf{u}_0)$ which is determined by the p vectors $\left\{ \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_i} du_i \right\}_{i=1}^p$, each of which is tangent to the surface defined by $\mathbf{x} = \mathbf{f}(\mathbf{u})$. (No sum on the repeated index.) From Definition F.1.1, the volume of this infinitesimal parallelepiped located at $\mathbf{f}(\mathbf{u}_0)$ is given by

$$\det \left(\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_i} du_i \cdot \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_j} du_j \right)^{1/2}. \quad (6.1)$$

The matrix in the above formula is a $p \times p$ matrix. Denoting

$$\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_i} = \mathbf{x}_{,i}$$

to save space, this matrix is of the form

$$\overbrace{\begin{pmatrix} du_1 & 0 & \cdots & 0 \\ 0 & du_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & du_p \end{pmatrix}}^{p \times p} \overbrace{\begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_p^T \end{pmatrix}}^{p \times n} \overbrace{\begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_p \end{pmatrix}}^{n \times p} \overbrace{\begin{pmatrix} du_1 & 0 & \cdots & 0 \\ 0 & du_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & du_p \end{pmatrix}}^{p \times p}$$

Therefore, by the theorem which says the determinant of a product equals the product of the determinants, the determinant of the above product equals

$$\det \begin{pmatrix} du_1 & 0 & \cdots & 0 \\ 0 & du_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & du_p \end{pmatrix}^2 \det \left(\begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_p^T \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_p \end{pmatrix} \right) =$$

$$\det \left(\begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_p^T \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_p \end{pmatrix} \right) (du_1 du_2 \cdots du_p)^2$$

and so taking the square root implies the volume of the infinitesimal parallelepiped at $\mathbf{x} = \mathbf{f}(\mathbf{u}_0)$ is

$$\det \left(\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_i} \cdot \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_j} \right)^{1/2} du_1 du_2 \cdots du_p =$$

$$\det \left(D\mathbf{f}(\mathbf{u})^T D\mathbf{f}(\mathbf{u}) \right)^{1/2} du_1 du_2 \cdots du_p$$

Definition F.1.2 Let $\mathbf{x} = \mathbf{f}(\mathbf{u})$ be as described above. Then the symbol, $\frac{\partial(x_1, \dots, x_n)}{\partial(u_1, \dots, u_p)}$, is defined by

$$\det \left(\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_i} \cdot \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_j} \right)^{1/2} \equiv \frac{\partial(x_1, \dots, x_n)}{\partial(u_1, \dots, u_p)}.$$

Also, the symbol, $dV_p \equiv \frac{\partial(x_1, \dots, x_n)}{\partial(u_1, \dots, u_p)} du_1 \cdots du_p$ is called the volume element or area element. Note the use of the subscript, p . This indicates the p dimensional volume element. When $p = 2$ it is customary to write dA . Also, continue referring to $\frac{\partial(x_1, \dots, x_n)}{\partial(u_1, \dots, u_p)}$ as the Jacobian.

This motivates the following fundamental procedure.

Procedure F.1.3 Suppose U is a subset of $\mathbb{R}^p$ and suppose $\mathbf{f} : U \rightarrow \mathbf{f}(U) \subseteq \mathbb{R}^n$ is a one to one and C^1 function. Then if $h : \mathbf{f}(U) \rightarrow \mathbb{R}$, define the p dimensional surface integral, $\int_{\mathbf{f}(U)} h(\mathbf{x}) dV_p$ according to the following formula.

$$\int_{\mathbf{f}(U)} h(\mathbf{x}) dV_p \equiv \int_U h(\mathbf{f}(\mathbf{u})) \frac{\partial(x_1, \dots, x_n)}{\partial(u_1, \dots, u_p)} dV.$$

This new definition of p dimensional volume does not contradict the earlier definition of 2 dimensional area. To review this fact, the two dimensional area of an infinitesimal parallelogram is

$$dA \equiv \left| \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1} du_1 \times \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_2} du_2 \right|.$$

Therefore,

$$dA^2 = \left| \left(\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1} \times \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_2} \right) \right|^2 (du_1 du_2)^2$$

Now recall that for $\mathbf{a}$ and $\mathbf{b}$ vectors in $\mathbb{R}^3$,

$$|\mathbf{a} \times \mathbf{b}|^2 = |\mathbf{a}|^2 |\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2 = \det \begin{pmatrix} \mathbf{a} \cdot \mathbf{a} & \mathbf{a} \cdot \mathbf{b} \\ \mathbf{a} \cdot \mathbf{b} & \mathbf{b} \cdot \mathbf{b} \end{pmatrix}.$$

Letting $\mathbf{a} = \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_1}$ and $\mathbf{b} = \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_2}$ this shows

$$dA = \det \left(\frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_i} \cdot \frac{\partial \mathbf{x}(\mathbf{u}_0)}{\partial u_j} \right)^{1/2} du_1 du_2 \equiv \frac{\partial (x_1, x_2, x_3)}{\partial (u_1, u_2)} du_1 du_2,$$

the same thing. Thus this new definition is just a considerable generalization of the case of two dimensional surfaces. The following example of a two dimensional surface in $\mathbb{R}^3$ was worked earlier.

Example F.1.4 Consider the surface given by $z = x^2$ for $(x, y) \in [0, 1] \times [0, 1] = U$. Find the surface area of this surface.

The first step in using the above is to write this surface in the form $\mathbf{x} = \mathbf{f}(\mathbf{u})$. This is easy to do if you let $\mathbf{u} = (x, y)$. Then $\mathbf{f}(x, y) = (x, y, x^2)$. If you like, let $x = u_1$ and $y = u_2$. What is $\frac{\partial (x_1, x_2, x_3)}{\partial (x, y)}$?

$$D\mathbf{f}(x, y) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2x & 0 \end{pmatrix}$$

and so

$$D\mathbf{f}(x, y)^T = \begin{pmatrix} 1 & 0 & 2x \\ 0 & 1 & 0 \end{pmatrix}$$

Thus in this case,

$$D\mathbf{f}(\mathbf{u})^T D\mathbf{f}(\mathbf{u}) = \begin{pmatrix} 1 + 4x^2 & 0 \\ 0 & 1 \end{pmatrix}$$

and so the area element is $\sqrt{1 + 4x^2} dx dy$ and the surface area is obtained by integrating the function, $h(\mathbf{x}) \equiv 1$. Therefore, this area is

$$\int_U dV = \int_0^1 \int_0^1 \sqrt{1 + 4x^2} dx dy = \frac{1}{2} \sqrt{5} - \frac{1}{4} \ln(-2 + \sqrt{5})$$

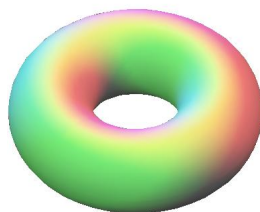
which can be obtained by using the trig. substitution, $2x = \tan \theta$ on the inside integral.

Here is another example of a two dimensional surface in $\mathbb{R}^3$ which was worked earlier.

Example F.1.5 Let $U = [0, 2\pi] \times [0, 2\pi]$ and for $(t, s) \in U$, let

$$\mathbf{f}(t, s) = (2 \cos t + \cos t \cos s, -2 \sin t - \sin t \cos s, \sin s)^T.$$

Find the area of $\mathbf{f}(U)$. This is the surface of a donut shown below.



To find its area,

$$D\mathbf{f}(t, s) = \begin{pmatrix} -2 \sin t - \sin t \cos s & -\cos t \sin s \\ -2 \cos t - \cos t \cos s & \sin t \sin s \\ 0 & \cos s \end{pmatrix}$$

and so

$$D\mathbf{f}(t, s)^T D\mathbf{f}(t, s) = \begin{pmatrix} 4 + 4 \cos s + \cos^2 s & 0 \\ 0 & 1 \end{pmatrix}$$

which implies the area element is

$$\begin{aligned} \det \begin{pmatrix} 4 + 4 \cos s + \cos^2 s & 0 \\ 0 & 1 \end{pmatrix}^{1/2} ds dt &= (4 + 4 \cos s + \cos^2 s)^{1/2} ds dt \\ &= (\cos s + 2) ds dt \end{aligned}$$

and the area is

$$\int_0^{2\pi} \int_0^{2\pi} (\cos s + 2) ds dt = 8\pi^2$$

Example F.1.6 Let $U = [0, 2\pi] \times [0, 2\pi]$ and for $(t, s) \in U$, let

$$\mathbf{f}(t, s) = (2 \cos t + \cos t \cos s, -2 \sin t - \sin t \cos s, \sin s)^T.$$

Find

$$\int_{\mathbf{f}(U)} h dV$$

where $h(x, y, z) = x^2$.

Everything is the same as the preceding example except this time it is an integral of a function. The area element is $(\cos s + 2) ds dt$ and so the integral called for is

$$\int_{\mathbf{f}(U)} h dV = \int_0^{2\pi} \int_0^{2\pi} \left(\overbrace{2 \cos t + \cos t \cos s}^{x \text{ on the surface}} \right)^2 (\cos s + 2) ds dt = 22\pi^2$$

Now here is an example of a three dimensional surface in $\mathbb{R}^4$.

Example F.1.7 Let $U = \{(x, y, z) : x^2 + y^2 + z^2 \leq 4\}$ and for $(x, y, z) \in U$ let $\mathbf{f}(x, y, z) = (x, y, x + y, z)$. Find the three dimensional volume of $\mathbf{f}(U)$.

Note there is no picture here because I am unable to draw one in four dimensions. Nevertheless it is a three dimensional volume which is being computed. Everything is done the same as before.

$$D\mathbf{f}(x, y, z) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so

$$D\mathbf{f}(x, y, z)^T D\mathbf{f}(x, y, z) = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so the volume element is $3 dx dy dz$. Therefore, the volume of $\mathbf{f}(U)$ is

$$\int_U 3 dx dy dz = 3 \left(\frac{4}{3} \pi (8) \right) = 32\pi.$$

One other issue is worth mentioning. Suppose $\mathbf{f}_i : U_i \rightarrow \mathbb{R}^n$ where U_i are sets in $\mathbb{R}^p$ and suppose $\mathbf{f}_1(U_1)$ intersects $\mathbf{f}_2(U_2)$ along C where $C = \mathbf{h}(V)$ for $V \subseteq \mathbb{R}^k$ for $k < p$. Then define integrals and areas over $\mathbf{f}_1(U_1) \cup \mathbf{f}_2(U_2)$ as follows.

$$\int_{\mathbf{f}_1(U_1) \cup \mathbf{f}_2(U_2)} g dV_p \equiv \int_{\mathbf{f}_1(U_1)} g dV_p + \int_{\mathbf{f}_2(U_2)} g dV_p.$$

Admittedly, the set C gets added in twice but this doesn't matter because its p dimensional volume equals zero and therefore, the integrals over this set will also be zero. Why is this? To find the p dimensional volume element on C , it is necessary to find a function, $\mathbf{f}$, mapping $U \subseteq \mathbb{R}^p$ to C . Let $\mathbf{f}(\mathbf{v}, s_1, \dots, s_{p-k}) \equiv \mathbf{h}(\mathbf{v})$. Then $D\mathbf{f}(\mathbf{v}, s_1, \dots, s_{p-k})$ has at least one column of zeros and so $\det(D\mathbf{f}^T D\mathbf{f}) = 0$ showing the p dimensional volume element is zero and so this makes no contribution to the integral as claimed. Clearly something similar holds in the case of many surfaces joined in this way.

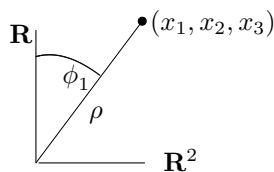
I have been purposely vague about precise mathematical conditions necessary for the above procedures. This is because the precise mathematical conditions which are usually cited are very technical and at the same time far too restrictive. The most general conditions under which these sorts of procedures are valid include things like Lipschitz functions defined on very general sets. These are functions satisfying a Lipschitz condition of the form $|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{y})| \leq K|\mathbf{x} - \mathbf{y}|$. For example, $y = |x|$ is Lipschitz continuous. However, this function does not have a derivative at every point. So it is with Lipschitz functions. However, it turns out these functions have derivatives at enough points to push everything through but this requires considerations involving the Lebesgue integral. Lipschitz functions are also not the most general kind of function for which the above is valid.

F.2 Spherical Coordinates In Many Dimensions*

Sometimes there is a need to deal with spherical coordinates in more than three dimensions. In this section, this concept is defined and formulas are derived for these coordinate systems. Recall polar coordinates are of the form

$$\begin{aligned} x_1 &= \rho \cos \theta \\ x_2 &= \rho \sin \theta \end{aligned}$$

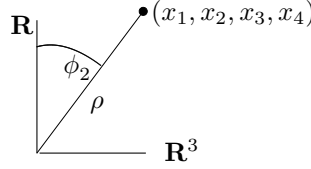
where $\rho > 0$ and $\theta \in [0, 2\pi)$. Here I am writing ρ in place of r to emphasize a pattern which is about to emerge. I will consider polar coordinates as spherical coordinates in two dimensions. This is also the reason I am writing x_1 and x_2 instead of the more usual x and y . Now consider what happens when you go to three dimensions. The situation is depicted in the following picture.



From this picture, you see that $x_3 = \rho \cos \phi_1$. Also the distance between (x_1, x_2) and $(0, 0)$ is $\rho \sin(\phi_1)$. Therefore, using polar coordinates to write (x_1, x_2) in terms of θ and this distance,

$$\begin{aligned} x_1 &= \rho \sin \phi_1 \cos \theta, \\ x_2 &= \rho \sin \phi_1 \sin \theta, \\ x_3 &= \rho \cos \phi_1. \end{aligned}$$

where $\phi_1 \in [0, \pi]$. What was done is to replace ρ with $\rho \sin \phi_1$ and then to add in $x_3 = \rho \cos \phi_1$. Having done this, there is no reason to stop with three dimensions. Consider the following picture:



From this picture, you see that $x_4 = \rho \cos \phi_2$. Also the distance between (x_1, x_2, x_3) and $(0, 0, 0)$ is $\rho \sin(\phi_2)$. Therefore, using polar coordinates to write (x_1, x_2, x_3) in terms of θ, ϕ_1 , and this distance,

$$\begin{aligned} x_1 &= \rho \sin \phi_2 \sin \phi_1 \cos \theta, \\ x_2 &= \rho \sin \phi_2 \sin \phi_1 \sin \theta, \\ x_3 &= \rho \sin \phi_2 \cos \phi_1, \\ x_4 &= \rho \cos \phi_2 \end{aligned}$$

where $\phi_2 \in [0, \pi]$.

Continuing this way, given spherical coordinates in $\mathbb{R}^n$, to get the spherical coordinates in $\mathbb{R}^{n+1}$, you let $x_{n+1} = \rho \cos \phi_{n-1}$ and then replace every occurrence of ρ with $\rho \sin \phi_{n-1}$ to obtain $x_1 \cdots x_n$ in terms of $\phi_1, \phi_2, \dots, \phi_{n-1}, \theta$, and ρ .

For spherical coordinates, it is always the case that ρ measures the distance from the point in $\mathbb{R}^n$ to the origin in $\mathbb{R}^n$, $\mathbf{0}$. Each $\phi_i \in [0, \pi]$, and $\theta \in [0, 2\pi)$. I leave it as an exercise using math induction to prove that these coordinates map $\prod_{i=1}^{n-2} [0, \pi] \times [0, 2\pi) \times (0, \infty)$ one to one onto $\mathbb{R}^n \setminus \{\mathbf{0}\}$.

It is customary to write S^{n-1} for the set $\{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| = 1\}$. Thus a parametrization for this level surface is given by letting $\rho = 1$ in spherical coordinates. I will denote by $S^{n-1}(a)$ the sphere having radius $a > 0$. What would the $n-1$ dimensional volume element on $S^{n-1}(a)$ be? For $S^1(a)$, there is only one parameter, θ . Therefore, the one dimensional volume element is

$$\det \left((-a \sin \theta, a \cos \theta) \begin{pmatrix} -a \sin \theta \\ a \cos \theta \end{pmatrix} \right)^{1/2} d\theta = a d\theta$$

where $\theta \in [0, 2\pi)$.

Next consider S^2 . In this case the two dimensional volume element is

$$\begin{aligned} & \det \left(\begin{pmatrix} -a \sin \phi_1 \sin \theta & a \sin \phi_1 \cos \theta & 0 \\ a \cos \phi_1 \cos \theta & a \cos \phi_1 \sin \theta & -a \sin \phi_1 \end{pmatrix} \right)^{1/2} \\ & \quad \cdot \begin{pmatrix} -a \sin \phi_1 \sin \theta & a \cos \phi_1 \cos \theta \\ a \sin \phi_1 \cos \theta & a \cos \phi_1 \sin \theta \end{pmatrix} d\phi_1 d\theta \\ & = \det \begin{pmatrix} a^2 \sin^2 \phi_1 & 0 \\ 0 & a^2 \end{pmatrix}^{1/2} d\phi_1 d\theta = a^2 (\sin \phi_1) d\phi_1 d\theta \end{aligned}$$

What of $S^3(a) \equiv \{\mathbf{x} \in \mathbb{R}^4 : |\mathbf{x}| = a\}$ and $S^n(a) = \{\mathbf{x} \in \mathbb{R}^{n+1} : |\mathbf{x}| = a\}$. Let $\mathbf{x}_n$ denote the vector $(x_1, \dots, x_n)$. That is, $\mathbf{x}_n$ consists of the first n components of $\mathbf{x}$. Let $D_{\theta \dots \phi_{n-2}}$ denote the derivative with respect to the vector, $(\theta, \phi_1, \dots, \phi_{n-2})$. Then the

volume element on $S^n(a)$ is of the form

$$\det \begin{pmatrix} \begin{pmatrix} \sin(\phi_{n-1}) (D_{\theta \dots \phi_{n-2}} \mathbf{x}_n)_{n \times n-1}^T & (0)_{(n-1) \times 1} \\ (*)_{1 \times n} & -a \sin \phi_{n-1} \end{pmatrix} \\ \cdot \begin{pmatrix} \sin(\phi_{n-1}) (D_{\theta \dots \phi_{n-2}} \mathbf{x}_n)_{n \times (n-1)} & (*)_{n \times 1} \\ (0)_{1 \times n-1} & -a \sin \phi_{n-1} \end{pmatrix} \end{pmatrix}^{1/2} \cdot d\phi_{n-1} \cdots d\phi_1 d\theta$$

Now using block multiplication, this reduces to

$$a \sin(\phi_{n-1}) \det \left(\begin{pmatrix} (D_{\theta \dots \phi_{n-2}} \mathbf{x}_n)_{n \times n-1}^T & (D_{\theta \dots \phi_{n-2}} \mathbf{x}_n)_{n \times (n-1)} \end{pmatrix} \right)^{1/2} \cdot d\phi_{n-1} \cdots d\phi_1 d\theta.$$

That is, to get the volume element in $S^n(a)$, you multiply the volume element on $S^{n-1}(a)$ by $a \sin(\phi_{n-1}) d\phi_{n-1}$. Consequently, beginning with the volume element on $S^1(a)$, you obtain the succession of volume elements for $S^1(a)$, $S^2(a)$, $S^3(a)$, $S^4(a)$.

$$a d\theta, a^2 \sin \phi_1 d\phi_1 d\theta, a^3 \sin \phi_2 \sin \phi_1 d\phi_2 d\phi_1 d\theta, \\ a^4 \sin \phi_3 \sin \phi_2 \sin \phi_1 d\phi_3 d\phi_2 d\phi_1 d\theta, \text{ etc.}$$

In general, the n dimensional volume element on $S^n(a)$ is

$$a^n \left(\prod_{i=1}^{n-1} \sin \phi_i \right) \left(\prod_{i=1}^{n-1} d\phi_i \right) d\theta \quad (6.2)$$

Using similar reasoning, the n dimensional volume element in terms of the spherical coordinates is

$$\rho^{n-1} \left(\prod_{i=1}^{n-2} \sin \phi_i \right) \left(\prod_{i=1}^{n-2} d\phi_i \right) d\theta d\rho. \quad (6.3)$$

Formulas 6.2 and 6.3 are very useful in estimating integrals.

Example F.2.1 For what values of s is the integral $\int_{B(\mathbf{0}, R)} (1 + |\mathbf{x}|^2)^s dV$ bounded independent of R ? Here $B(\mathbf{0}, R)$ is the ball, $\{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| \leq R\}$.

I think you can see immediately that s must be negative but exactly how negative? It turns out it depends on n and using spherical coordinates, you can find just exactly what is needed. It is very hard to overstate the importance of the technique I am about to show you. Write the above integral in n dimensional spherical coordinates.

$$\int_0^R \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} \rho^{n-1} (1 + \rho^2)^s \left(\prod_{i=1}^{n-2} \sin \phi_i \right) \left(\prod_{i=1}^{n-2} d\phi_i \right) d\theta d\rho \\ = \int_0^R \rho^{n-1} (1 + \rho^2)^s \int_{S^{n-1}} dS^{n-1} d\rho = \omega_n \int_0^R \rho^{n-1} (1 + \rho^2)^s d\rho$$

where $dS^{n-1} = \left(\prod_{i=1}^{n-2} \sin \phi_i \right) \left(\prod_{i=1}^{n-2} d\phi_i \right) d\theta$.

$$\omega_n \equiv \int_0^\pi \cdots \int_0^\pi \int_0^{2\pi} \left(\prod_{i=1}^{n-2} \sin \phi_i \right) \left(\prod_{i=1}^{n-2} d\phi_i \right) d\theta$$

and from the above explanation this equals the area of S^{n-1} , the unit sphere,

$$\{\mathbf{x} \in \mathbb{R}^n: |\mathbf{x}| = 1\}.$$

Now the very hard problem has been reduced to considering an easy one variable problem of finding when

$$\int_0^R \rho^{n-1} (1 + \rho^2)^s d\rho$$

is bounded independent of R . I leave it to you to verify using standard one variable calculus that you need $2s + (n - 1) < -1$ so you need $s < -n/2$.

Some Physical Applications

G.1 Product Rule For Matrices

Here is the concept of the product rule extended to matrix multiplication.

Definition G.1.1 Let $A(t)$ be an $m \times n$ matrix. Say $A(t) = (A_{ij}(t))$. Suppose also that $A_{ij}(t)$ is a differentiable function for all i, j . Then define $A'(t) \equiv (A'_{ij}(t))$. That is, $A'(t)$ is the matrix which consists of replacing each entry by its derivative. Such an $m \times n$ matrix in which the entries are differentiable functions is called a differentiable matrix.

The next lemma is just a version of the product rule.

Lemma G.1.2 Let $A(t)$ be an $m \times n$ matrix and let $B(t)$ be an $n \times p$ matrix with the property that all the entries of these matrices are differentiable functions. Then

$$(A(t)B(t))' = A'(t)B(t) + A(t)B'(t).$$

Proof: $(A(t)B(t))' = (C'_{ij}(t))$ where $C_{ij}(t) = A_{ik}(t)B_{kj}(t)$ and the repeated index summation convention is being used. Therefore,

$$\begin{aligned} C'_{ij}(t) &= A'_{ik}(t)B_{kj}(t) + A_{ik}(t)B'_{kj}(t) \\ &= (A'(t)B(t))_{ij} + (A(t)B'(t))_{ij} \\ &= (A'(t)B(t) + A(t)B'(t))_{ij} \end{aligned}$$

Therefore, the ij^{th} entry of $A(t)B(t)$ equals the ij^{th} entry of $A'(t)B(t) + A(t)B'(t)$ and this proves the lemma.

G.2 Moving Coordinate Systems

Let $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ be a right handed¹ orthonormal basis of vectors for each t . It is assumed these vectors are C^1 functions of t . Letting the positive x axis extend in the direction of $\mathbf{i}(t)$, the positive y axis extend in the direction of $\mathbf{j}(t)$, and the positive z axis extend in the direction of $\mathbf{k}(t)$, yields a moving coordinate system. Now let $\mathbf{u} = (u_1, u_2, u_3) \in \mathbb{R}^3$ and let t_0 be some reference time. For example you could let $t_0 = 0$. Then define the components of $\mathbf{u}$ with respect to these vectors, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ at time t_0 as

$$\mathbf{u} \equiv u_1\mathbf{i}(t_0) + u_2\mathbf{j}(t_0) + u_3\mathbf{k}(t_0).$$

¹Recall that right handed implies $\mathbf{i} \times \mathbf{j} = \mathbf{k}$.

Let $\mathbf{u}(t)$ be defined as the vector which has the same components with respect to $\mathbf{i}, \mathbf{j}, \mathbf{k}$ but at time t . Thus

$$\mathbf{u}(t) \equiv u_1 \mathbf{i}(t) + u_2 \mathbf{j}(t) + u_3 \mathbf{k}(t).$$

and the vector has changed although the components have not.

For example, this is exactly the situation in the case of apparently fixed basis vectors on the earth if $\mathbf{u}$ is a position vector from the given spot on the earth's surface to a point regarded as fixed with the earth due to its keeping the same coordinates relative to coordinate axes which are fixed with the earth.

Now define a linear transformation $Q(t)$ mapping $\mathbb{R}^3$ to $\mathbb{R}^3$ by

$$Q(t) \mathbf{u} \equiv u_1 \mathbf{i}(t) + u_2 \mathbf{j}(t) + u_3 \mathbf{k}(t)$$

where

$$\mathbf{u} \equiv u_1 \mathbf{i}(t_0) + u_2 \mathbf{j}(t_0) + u_3 \mathbf{k}(t_0)$$

Thus letting $\mathbf{v}, \mathbf{u} \in \mathbb{R}^3$ be vectors and α, β , scalars,

$$\begin{aligned} Q(t)(\alpha \mathbf{u} + \beta \mathbf{v}) &\equiv (\alpha u_1 + \beta v_1) \mathbf{i}(t) + (\alpha u_2 + \beta v_2) \mathbf{j}(t) + (\alpha u_3 + \beta v_3) \mathbf{k}(t) \\ &= (\alpha u_1 \mathbf{i}(t) + \alpha u_2 \mathbf{j}(t) + \alpha u_3 \mathbf{k}(t)) + (\beta v_1 \mathbf{i}(t) + \beta v_2 \mathbf{j}(t) + \beta v_3 \mathbf{k}(t)) \\ &= \alpha (u_1 \mathbf{i}(t) + u_2 \mathbf{j}(t) + u_3 \mathbf{k}(t)) + \beta (v_1 \mathbf{i}(t) + v_2 \mathbf{j}(t) + v_3 \mathbf{k}(t)) \\ &\equiv \alpha Q(t) \mathbf{u} + \beta Q(t) \mathbf{v} \end{aligned}$$

showing that $Q(t)$ is a linear transformation. Also, $Q(t)$ preserves all distances because, since the vectors, $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ form an orthonormal set,

$$|Q(t) \mathbf{u}| = \left(\sum_{i=1}^3 (u^i)^2 \right)^{1/2} = |\mathbf{u}|.$$

For simplicity, let

$$\mathbf{i}(t) = \mathbf{e}_1(t), \mathbf{j}(t) = \mathbf{e}_2(t), \mathbf{k}(t) = \mathbf{e}_3(t)$$

and

$$\mathbf{i}(t_0) = \mathbf{e}_1(t_0), \mathbf{j}(t_0) = \mathbf{e}_2(t_0), \mathbf{k}(t_0) = \mathbf{e}_3(t_0).$$

Then using the repeated index summation convention,

$$\mathbf{u}(t) = u_j \mathbf{e}_j(t) = u_j \mathbf{e}_j(t) \cdot \mathbf{e}_i(t_0) \mathbf{e}_i(t_0)$$

and so with respect to the basis, $\mathbf{i}(t_0) = \mathbf{e}_1(t_0), \mathbf{j}(t_0) = \mathbf{e}_2(t_0), \mathbf{k}(t_0) = \mathbf{e}_3(t_0)$, the matrix of $Q(t)$ is

$$Q_{ij}(t) = \mathbf{e}_i(t_0) \cdot \mathbf{e}_j(t)$$

Recall this means you take a vector, $\mathbf{u} \in \mathbb{R}^3$ which is a list of the components of $\mathbf{u}$ with respect to $\mathbf{i}(t_0), \mathbf{j}(t_0), \mathbf{k}(t_0)$ and when you multiply by $Q(t)$ you get the components of $\mathbf{u}(t)$ with respect to $\mathbf{i}(t_0), \mathbf{j}(t_0), \mathbf{k}(t_0)$. I will refer to this matrix as $Q(t)$ to save notation.

Lemma G.2.1 *Suppose $Q(t)$ is a real, differentiable $n \times n$ matrix which preserves distances. Then $Q(t)Q(t)^T = Q(t)^T Q(t) = I$. Also, if $\mathbf{u}(t) \equiv Q(t) \mathbf{u}$, then there exists a vector, $\boldsymbol{\Omega}(t)$ such that*

$$\mathbf{u}'(t) = \boldsymbol{\Omega}(t) \times \mathbf{u}(t).$$

Proof: Recall that $(\mathbf{z} \cdot \mathbf{w}) = \frac{1}{4} (|\mathbf{z} + \mathbf{w}|^2 - |\mathbf{z} - \mathbf{w}|^2)$. Therefore,

$$\begin{aligned} (Q(t) \mathbf{u} \cdot Q(t) \mathbf{w}) &= \frac{1}{4} (|Q(t) (\mathbf{u} + \mathbf{w})|^2 - |Q(t) (\mathbf{u} - \mathbf{w})|^2) \\ &= \frac{1}{4} (|\mathbf{u} + \mathbf{w}|^2 - |\mathbf{u} - \mathbf{w}|^2) \\ &= (\mathbf{u} \cdot \mathbf{w}). \end{aligned}$$

This implies

$$(Q(t)^T Q(t) \mathbf{u} \cdot \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})$$

for all $\mathbf{u}, \mathbf{w}$. Therefore, $Q(t)^T Q(t) \mathbf{u} = \mathbf{u}$ and so $Q(t)^T Q(t) = Q(t) Q(t)^T = I$. This proves the first part of the lemma.

It follows from the product rule, Lemma G.1.2 that

$$Q'(t) Q(t)^T + Q(t) Q'(t)^T = 0$$

and so

$$Q'(t) Q(t)^T = - (Q'(t) Q(t)^T)^T. \quad (7.1)$$

From the definition, $Q(t) \mathbf{u} = \mathbf{u}(t)$,

$$\mathbf{u}'(t) = Q'(t) \mathbf{u} = Q'(t) \overbrace{Q(t)^T}^{=\mathbf{u}} \mathbf{u}(t).$$

Then writing the matrix of $Q'(t) Q(t)^T$ with respect to $\mathbf{i}(t_0), \mathbf{j}(t_0), \mathbf{k}(t_0)$, it follows from 7.1 that the matrix of $Q'(t) Q(t)^T$ is of the form

$$\begin{pmatrix} 0 & -\omega_3(t) & \omega_2(t) \\ \omega_3(t) & 0 & -\omega_1(t) \\ -\omega_2(t) & \omega_1(t) & 0 \end{pmatrix}$$

for some time dependent scalars, ω_i . Therefore,

$$\begin{aligned} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}'(t) &= \begin{pmatrix} 0 & -\omega_3(t) & \omega_2(t) \\ \omega_3(t) & 0 & -\omega_1(t) \\ -\omega_2(t) & \omega_1(t) & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}(t) \\ &= \begin{pmatrix} \omega_2(t) u_3(t) - \omega_3(t) u_2(t) \\ \omega_3(t) u_1(t) - \omega_1(t) u_3(t) \\ \omega_1(t) u_2(t) - \omega_2(t) u_1(t) \end{pmatrix} \end{aligned}$$

where the u_i are the components of the vector $\mathbf{u}(t)$ in terms of the fixed vectors

$$\mathbf{i}(t_0), \mathbf{j}(t_0), \mathbf{k}(t_0).$$

Therefore,

$$\mathbf{u}'(t) = \boldsymbol{\Omega}(t) \times \mathbf{u}(t) = Q'(t) Q(t)^T \mathbf{u}(t) \quad (7.2)$$

where

$$\boldsymbol{\Omega}(t) = \omega_1(t) \mathbf{i}(t_0) + \omega_2(t) \mathbf{j}(t_0) + \omega_3(t) \mathbf{k}(t_0).$$

because

$$\boldsymbol{\Omega}(t) \times \mathbf{u}(t) \equiv \begin{vmatrix} \mathbf{i}(t_0) & \mathbf{j}(t_0) & \mathbf{k}(t_0) \\ w_1 & w_2 & w_3 \\ u_1 & u_2 & u_3 \end{vmatrix} \equiv$$

$$\mathbf{i}(t_0) (w_2 u_3 - w_3 u_2) + \mathbf{j}(t_0) (w_3 u_1 - w_1 u_3) + \mathbf{k}(t_0) (w_1 u_2 - w_2 u_1).$$

This proves the lemma and yields the existence part of the following theorem.

Theorem G.2.2 *Let $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ be as described. Then there exists a unique vector $\boldsymbol{\Omega}(t)$ such that if $\mathbf{u}(t)$ is a vector whose components are constant with respect to $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$, then*

$$\mathbf{u}'(t) = \boldsymbol{\Omega}(t) \times \mathbf{u}(t).$$

Proof: It only remains to prove uniqueness. Suppose $\boldsymbol{\Omega}_1$ also works. Then $\mathbf{u}(t) = Q(t)\mathbf{u}$ and so $\mathbf{u}'(t) = Q'(t)\mathbf{u}$ and

$$Q'(t)\mathbf{u} = \boldsymbol{\Omega} \times Q(t)\mathbf{u} = \boldsymbol{\Omega}_1 \times Q(t)\mathbf{u}$$

for all $\mathbf{u}$. Therefore,

$$(\boldsymbol{\Omega} - \boldsymbol{\Omega}_1) \times Q(t)\mathbf{u} = \mathbf{0}$$

for all $\mathbf{u}$ and since $Q(t)$ is one to one and onto, this implies $(\boldsymbol{\Omega} - \boldsymbol{\Omega}_1) \times \mathbf{w} = \mathbf{0}$ for all $\mathbf{w}$ and thus $\boldsymbol{\Omega} - \boldsymbol{\Omega}_1 = \mathbf{0}$. This proves the theorem.

Definition G.2.3 *A **rigid body** in $\mathbb{R}^3$ has a moving coordinate system with the property that for an observer on the rigid body, the vectors, $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ are constant. More generally, a vector $\mathbf{u}(t)$ is said to be fixed with the body if to a person on the body, the vector appears to have the same magnitude and same direction independent of t . Thus $\mathbf{u}(t)$ is fixed with the body if $\mathbf{u}(t) = u_1\mathbf{i}(t) + u_2\mathbf{j}(t) + u_3\mathbf{k}(t)$.*

The following comes from the above discussion.

Theorem G.2.4 *Let $B(t)$ be the set of points in three dimensions occupied by a rigid body. Then there exists a vector $\boldsymbol{\Omega}(t)$ such that whenever $\mathbf{u}(t)$ is fixed with the rigid body,*

$$\mathbf{u}'(t) = \boldsymbol{\Omega}(t) \times \mathbf{u}(t).$$

G.3 Newton's Laws Of Motion

Definition G.3.1 *Let $\mathbf{r}(t)$ denote the position of an object. Then the acceleration of the object is defined to be $\mathbf{r}''(t)$.*

Newton's² first law is: "Every body persists in its state of rest or of uniform motion in a straight line unless it is compelled to change that state by forces impressed on it."

Newton's second law is:

$$\mathbf{F} = m\mathbf{a} \tag{7.3}$$

where $\mathbf{a}$ is the acceleration and m is the mass of the object.

Newton's third law states: "To every action there is always opposed an equal reaction; or, the mutual actions of two bodies upon each other are always equal, and directed to contrary parts."

Of these laws, only the second two are independent of each other, the first law being implied by the second. The third law says roughly that if you apply a force to something, the thing applies the same force back.

²Isaac Newton 1642-1727 is often credited with inventing calculus although this is not correct since most of the ideas were in existence earlier. However, he made major contributions to the subject partly in order to study physics and astronomy. He formulated the laws of gravity, made major contributions to optics, and stated the fundamental laws of mechanics listed here. He invented a version of the binomial theorem when he was only 23 years old and built a reflecting telescope. He showed that Kepler's laws for the motion of the planets came from calculus and his laws of gravitation. In 1686 he published an important book, Principia, in which many of his ideas are found. Newton was also very interested in theology and had strong views on the nature of God which were based on his study of the Bible and early Christian writings. He finished his life as Master of the Mint.

The second law is the one of most interest. Note that the statement of this law depends on the concept of the derivative because the acceleration is defined as a derivative. Newton used calculus and these laws to solve profound problems involving the motion of the planets and other problems in mechanics. The next example involves the concept that if you know the force along with the initial velocity and initial position, then you can determine the position.

Example G.3.2 Let $\mathbf{r}(t)$ denote the position of an object of mass 2 kilogram at time t and suppose the force acting on the object is given by $\mathbf{F}(t) = (t, 1 - t^2, 2e^{-t})$. Suppose $\mathbf{r}(0) = (1, 0, 1)$ meters, and $\mathbf{r}'(0) = (0, 1, 1)$ meters/sec. Find $\mathbf{r}(t)$.

By Newton's second law, $2\mathbf{r}''(t) = \mathbf{F}(t) = (t, 1 - t^2, 2e^{-t})$ and so

$$\mathbf{r}''(t) = (t/2, (1 - t^2)/2, e^{-t}).$$

Therefore the velocity is given by

$$\mathbf{r}'(t) = \left(\frac{t^2}{4}, \frac{t - t^3/3}{2}, -e^{-t} \right) + \mathbf{c}$$

where $\mathbf{c}$ is a constant vector which must be determined from the initial condition given for the velocity. Thus letting $\mathbf{c} = (c_1, c_2, c_3)$,

$$(0, 1, 1) = (0, 0, -1) + (c_1, c_2, c_3)$$

which requires $c_1 = 0$, $c_2 = 1$, and $c_3 = 2$. Therefore, the velocity is found.

$$\mathbf{r}'(t) = \left(\frac{t^2}{4}, \frac{t - t^3/3}{2} + 1, -e^{-t} + 2 \right).$$

Now from this, the displacement must equal

$$\mathbf{r}(t) = \left(\frac{t^3}{12}, \frac{t^2/2 - t^4/12}{2} + t, e^{-t} + 2t \right) + (C_1, C_2, C_3)$$

where the constant vector, (C_1, C_2, C_3) must be determined from the initial condition for the displacement. Thus

$$\mathbf{r}(0) = (1, 0, 1) = (0, 0, 1) + (C_1, C_2, C_3)$$

which means $C_1 = 1$, $C_2 = 0$, and $C_3 = 0$. Therefore, the displacement has also been found.

$$\mathbf{r}(t) = \left(\frac{t^3}{12} + 1, \frac{t^2/2 - t^4/12}{2} + t, e^{-t} + 2t \right) \text{ meters.}$$

Actually, in applications of this sort of thing acceleration does not usually come to you as a nice given function written in terms of simple functions you understand. Rather, it comes as measurements taken by instruments and the position is continuously being updated based on this information. Another situation which often occurs is the case when the forces on the object depend not just on time but also on the position or velocity of the object.

Example G.3.3 An artillery piece is fired at ground level on a level plain. The angle of elevation is $\pi/6$ radians and the speed of the shell is 400 meters per second. How far does the shell fly before hitting the ground?

Neglect air resistance in this problem. Also let the direction of flight be along the positive x axis. Thus the initial velocity is the vector, $400 \cos(\pi/6) \mathbf{i} + 400 \sin(\pi/6) \mathbf{j}$ while the only force experienced by the shell after leaving the artillery piece is the force of gravity, $-mg\mathbf{j}$ where m is the mass of the shell. The acceleration of gravity equals 9.8 meters per sec^2 and so the following needs to be solved.

$$m\mathbf{r}''(t) = -mg\mathbf{j}, \mathbf{r}(0) = (0, 0), \mathbf{r}'(0) = 400 \cos(\pi/6) \mathbf{i} + 400 \sin(\pi/6) \mathbf{j}.$$

Denoting $\mathbf{r}(t)$ as $(x(t), y(t))$,

$$x''(t) = 0, y''(t) = -g.$$

Therefore, $y'(t) = -gt + C$ and from the information on the initial velocity,

$$C = 400 \sin(\pi/6) = 200.$$

Thus

$$y(t) = -4.9t^2 + 200t + D.$$

$D = 0$ because the artillery piece is fired at ground level which requires both x and y to equal zero at this time. Similarly, $x'(t) = 400 \cos(\pi/6)$ so $x(t) = 400 \cos(\pi/6)t = 200\sqrt{3}t$. The shell hits the ground when $y = 0$ and this occurs when $-4.9t^2 + 200t = 0$. Thus $t = 40.8163265306$ seconds and so at this time,

$$x = 200\sqrt{3}(40.8163265306) = 14139.1902659 \text{ meters.}$$

The next example is more complicated because it also takes in to account air resistance. We do not live in a vacume.

Example G.3.4 *A lump of "blue ice" escapes the lavatory of a jet flying at 600 miles per hour at an altitude of 30,000 feet. This blue ice weighs 64 pounds near the earth and experiences a force of air resistance equal to $(-.1)\mathbf{r}'(t)$ pounds. Find the position and velocity of the blue ice as a function of time measured in seconds. Also find the velocity when the lump hits the ground. Such lumps have been known to surprise people on the ground.*

The first thing needed is to obtain information which involves consistent units. The blue ice weighs 32 pounds near the earth. Thus 32 pounds is the force exerted by gravity on the lump and so its mass must be given by Newton's second law as follows.

$$64 = m \times 32.$$

Thus $m = 2$ slugs. The slug is the unit of mass in the system involving feet and pounds. The jet is flying at 600 miles per hour. I want to change this to feet per second. Thus it flies at

$$\frac{600 \times 5280}{60 \times 60} = 880 \text{ feet per second.}$$

The explanation for this is that there are 5280 feet in a mile and so it goes 600×5280 feet in one hour. There are 60×60 seconds in an hour. The position of the lump of blue ice will be computed from a point on the ground directly beneath the airplane at the instant the blue ice escapes and regard the airplane as moving in the direction of the positive x axis. Thus the initial displacement is

$$\mathbf{r}(0) = (0, 30000) \text{ feet}$$

and the initial velocity is

$$\mathbf{r}'(0) = (880, 0) \text{ feet/sec.}$$

The force of gravity is

$$(0, -64) \text{ pounds}$$

and the force due to air resistance is

$$(-.1) \mathbf{r}'(t) \text{ pounds.}$$

Newtons second law yields the following initial value problem for $\mathbf{r}(t) = (r_1(t), r_2(t))$.

$$\begin{aligned} 2(r_1''(t), r_2''(t)) &= (-.1)(r_1'(t), r_2'(t)) + (0, -64), \quad (r_1(0), r_2(0)) = (0, 30000), \\ (r_1'(0), r_2'(0)) &= (880, 0) \end{aligned}$$

Therefore,

$$\begin{aligned} 2r_1''(t) + (.1)r_1'(t) &= 0 \\ 2r_2''(t) + (.1)r_2'(t) &= -64 \\ r_1(0) &= 0 \\ r_2(0) &= 30000 \\ r_1'(0) &= 880 \\ r_2'(0) &= 0 \end{aligned} \tag{7.4}$$

To save on repetition solve

$$mr'' + kr' = c, r(0) = u, r'(0) = v.$$

Divide the differential equation by m and get

$$r'' + (k/m)r' = c/m.$$

Now multiply both sides by $e^{(k/m)t}$. You should check this gives

$$\frac{d}{dt} \left(e^{(k/m)t} r' \right) = (c/m) e^{(k/m)t}$$

Therefore,

$$e^{(k/m)t} r' = \frac{1}{k} e^{\frac{k}{m}t} c + C$$

and using the initial condition, $v = c/k + C$ and so

$$r'(t) = (c/k) + (v - (c/k)) e^{-\frac{k}{m}t}$$

Now this implies

$$r(t) = (c/k)t - \frac{1}{k} m e^{-\frac{k}{m}t} \left(v - \frac{c}{k} \right) + D \tag{7.5}$$

where D is a constant to be determined from the initial conditions. Thus

$$u = -\frac{m}{k} \left(v - \frac{c}{k} \right) + D$$

and so

$$r(t) = (c/k)t - \frac{1}{k} m e^{-\frac{k}{m}t} \left(v - \frac{c}{k} \right) + \left(u + \frac{m}{k} \left(v - \frac{c}{k} \right) \right).$$

Now apply this to the system 7.4 to find

$$\begin{aligned} r_1(t) &= -\frac{1}{(.1)} 2 \left(\exp \left(\frac{-.1}{2} t \right) \right) (880) + \left(\frac{2}{(.1)} (880) \right) \\ &= -17600.0 \exp(-.05t) + 17600.0 \end{aligned}$$

and

$$\begin{aligned} r_2(t) &= (-64/ (.1)) t - \frac{1}{(.1)^2} 2 \left(\exp \left(-\frac{(.1)}{2} t \right) \right) \left(\frac{64}{(.1)} \right) + \left(30000 + \frac{2}{(.1)} \left(\frac{64}{(.1)} \right) \right) \\ &= -640.0t - 12800.0 \exp(-.05t) + 42800.0 \end{aligned}$$

This gives the coordinates of the position. What of the velocity? Using 7.5 in the same way to obtain the velocity,

$$\begin{aligned} r_1'(t) &= 880.0 \exp(-.05t), \\ r_2'(t) &= -640.0 + 640.0 \exp(-.05t). \end{aligned} \tag{7.6}$$

To determine the velocity when the blue ice hits the ground, it is necessary to find the value of t when this event takes place and then to use 7.6 to determine the velocity. It hits ground when $r_2(t) = 0$. Thus it suffices to solve the equation,

$$0 = -640.0t - 12800.0 \exp(-.05t) + 42800.0.$$

This is a fairly hard equation to solve using the methods of algebra. In fact, I do not have a good way to find this value of t using algebra. However if plugging in various values of t using a calculator you eventually find that when $t = 66.14$,

$$-640.0(66.14) - 12800.0 \exp(-.05(66.14)) + 42800.0 = 1.588 \text{ feet.}$$

This is close enough to hitting the ground and so plugging in this value for t yields the approximate velocity,

$$(880.0 \exp(-.05(66.14)), -640.0 + 640.0 \exp(-.05(66.14))) = (32.23, -616.56).$$

Notice how because of air resistance the component of velocity in the horizontal direction is only about 32 feet per second even though this component started out at 880 feet per second while the component in the vertical direction is -616 feet per second even though this component started off at 0 feet per second. You see that air resistance can be very important so it is not enough to pretend, as is often done in beginning physics courses that everything takes place in a vacuum. Actually, this problem used several physical simplifications. It was assumed the force acting on the lump of blue ice by gravity was constant. This is not really true because it actually depends on the distance between the center of mass of the earth and the center of mass of the lump. It was also assumed the air resistance is proportional to the velocity. This is an over simplification when high speeds are involved. However, increasingly correct models can be studied in a systematic way as above.

G.4 Kinetic Energy

Newton's second law is also the basis for the notion of **kinetic energy**. When a force is exerted on an object which causes the object to move, it follows that the force is doing work which manifests itself in a change of velocity of the object. How is the total work done on the object by the force related to the final velocity of the object? By Newton's second law, and letting $\mathbf{v}$ be the velocity,

$$\mathbf{F}(t) = m\mathbf{v}'(t).$$

Now in a small increment of time, $(t, t + dt)$, the work done on the object would be approximately equal to

$$dW = \mathbf{F}(t) \cdot \mathbf{v}(t) dt. \tag{7.7}$$

If no work has been done at time $t = 0$, then 7.7 implies

$$\frac{dW}{dt} = \mathbf{F} \cdot \mathbf{v}, \quad W(0) = 0.$$

Hence,

$$\frac{dW}{dt} = m\mathbf{v}'(t) \cdot \mathbf{v}(t) = \frac{m}{2} \frac{d}{dt} |\mathbf{v}(t)|^2.$$

Therefore, the total work done up to time t would be $W(t) = \frac{m}{2} |\mathbf{v}(t)|^2 - \frac{m}{2} |\mathbf{v}_0|^2$ where $|\mathbf{v}_0|$ denotes the initial speed of the object. This difference represents the change in the kinetic energy.

G.5 Kepler's First And Second Laws

Kepler's first law states the planets move in elliptical orbits about the sun. There are actually three laws which will be discussed later from a different point of view. Here I will only consider the first law. It comes from Newton's law of gravitation which says the force between two balls or point masses having mass M and m is

$$GMm \frac{1}{r^2}$$

where r is the distance between their centers of mass. Letting M denote the mass of the sun whose center of mass is taken to be the origin and m the mass of a planet, it follows from Newton's second law that for $\mathbf{r}(t)$ the position vector of the planet

$$\mathbf{r}'' = -GMm \frac{1}{r^2} \mathbf{u} \quad (7.8)$$

where $\mathbf{u}$ is the unit vector $\mathbf{r}/|\mathbf{r}| = \mathbf{r}/r$ where also $r = |\mathbf{r}|$. Before beginning this derivation, recall the following vector identity involving the cross product.

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c}) \mathbf{b} - (\mathbf{b} \cdot \mathbf{c}) \mathbf{a}$$

Also

$$\begin{aligned} \mathbf{r}' \times \mathbf{r} &= (r\mathbf{u})' \times r\mathbf{u} = (r'\mathbf{u} + r\mathbf{u}') \times r\mathbf{u} \\ &= r'r\mathbf{u} \times \mathbf{u} + r^2\mathbf{u}' \times \mathbf{u} = r^2\mathbf{u}' \times \mathbf{u}. \end{aligned}$$

Then from 7.8, properties of the cross product, and Newton's second law,

$$(\mathbf{r}' \times \mathbf{r})' = \mathbf{r}'' \times \mathbf{r} = -GMm \frac{1}{r^2} \mathbf{u} \times r\mathbf{u} = \mathbf{0}$$

and so there exists a constant vector $\mathbf{k}$ such that

$$\mathbf{r}' \times \mathbf{r} = \mathbf{k}. \quad (7.9)$$

Thus $\mathbf{r} \cdot \mathbf{k} = 0$ and so the motion of the planet is in a plane.

Next,

$$\begin{aligned} (\mathbf{k} \times \mathbf{r}')' &= \mathbf{k} \times \mathbf{r}'' = -\left(\mathbf{k} \times \frac{GM}{r^2} \mathbf{u}\right) \\ &= -\frac{GM}{r^2} (\mathbf{k} \times \mathbf{u}) = -\frac{GM}{r^2} ((\mathbf{r}' \times \mathbf{r}) \times \mathbf{u}) \\ &= -GM ((\mathbf{u}' \times \mathbf{u}) \times \mathbf{u}) = -GM \left[(\mathbf{u}' \cdot \mathbf{u}) \mathbf{u} - |\mathbf{u}|^2 \mathbf{u}' \right]. \end{aligned}$$

Now $(\mathbf{u}' \cdot \mathbf{u}) = 0$ because $1 = \mathbf{u} \cdot \mathbf{u}$ and taking the derivative of both sides yields the result. Therefore,

$$(\mathbf{k} \times \mathbf{r}')' = GM\mathbf{u}'$$

and so

$$\mathbf{k} \times \mathbf{r}' = GM\mathbf{u} + \mathbf{w}$$

for $\mathbf{w}$ a fixed vector independent of t . Without loss of generality, it can be assumed $\mathbf{w} = w\mathbf{i}$. Just rotate the coordinate axes in the plane of motion of the planet to make this happen. Then

$$\begin{aligned} \mathbf{r} \cdot (\mathbf{k} \times \mathbf{r}') &= GMr + w\mathbf{r} \cdot \mathbf{i} \\ &= GMr + wr \cos \theta \end{aligned} \quad (7.10)$$

where θ is the angle between $\mathbf{r}$ and $\mathbf{i}$. Also from Lemma 16.4.19 on Page 389 and 7.9,

$$\begin{aligned} \mathbf{r} \cdot (\mathbf{k} \times \mathbf{r}') &= -\mathbf{r} \cdot (\mathbf{r}' \times \mathbf{k}) \\ &= -\mathbf{r} \times \mathbf{r}' \cdot \mathbf{k} \\ &= \mathbf{r}' \times \mathbf{r} \cdot \mathbf{k} \\ &= \mathbf{k} \cdot \mathbf{k} = |\mathbf{k}|^2 \equiv k^2, \end{aligned}$$

a constant. Therefore, letting x and y be the usual rectangular coordinates,

$$k^2 = GM\sqrt{x^2 + y^2} + wx.$$

Then

$$\begin{aligned} G^2M^2(x^2 + y^2) &= (k^2 - wx)^2 \\ &= k^4 - 2k^2wx + w^2x^2 \end{aligned}$$

and this is the equation for either an ellipse, a hyperbola, or a parabola. Since planet's orbits are not unbounded, this implies it must be the equation for an ellipse. Solving 7.10 for r gives

$$r = \frac{k^2}{GM + w \cos(\theta)} = \frac{(k^2/GM)}{1 + e \cos(\theta)}$$

where $e = w/GM$. This is a more elegant way to describe the curve. If $e < 1$, this must be an ellipse while if $e = 0$ it must be a circle. If $e = 1$, the curve is a parabola and if $e > 1$ the curve is a hyperbola. You should verify this is the case.

Here I will give a rough argument for the equal area law which is also called Kepler's second law in the context of planetary motion. This will be done more carefully later. If the force experienced by an object of mass m at position $\mathbf{r}$ is $k(r)\mathbf{u}$ where $\mathbf{u} = \mathbf{r}/r$, this is called a central force field. Gravity is such an example because in this case,

$$k(r) = -\frac{GMm}{r^2}.$$

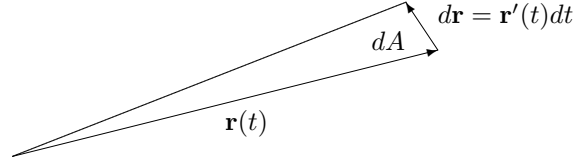
Any motion caused by a central force field moves in a plane and satisfies something called the equal area rule which states the position vector sweeps out area at a constant rate. To see this is the case,

$$(\mathbf{r} \times \mathbf{r}')' = \mathbf{r} \times \mathbf{r}'' = \mathbf{r} \times k(r)\mathbf{u} = \mathbf{0}$$

and so there exists a constant vector $\mathbf{k}$ such that

$$\mathbf{r} \times \mathbf{r}' = \mathbf{k}.$$

Thus $\mathbf{r} \cdot \mathbf{k} = 0$ so the motion is in a plane. Also letting dA denote an increment of area corresponding to dt an increment of time, the following picture is an illustration of this increment of area.



Remember now how to take the area of a triangle using the cross product. This yields

$$dA = \frac{1}{2} \left| \mathbf{r} \times \overbrace{\mathbf{r}' dt}^{d\mathbf{r}} \right| = \frac{1}{2} |\mathbf{k}| dt$$

and so

$$\frac{dA}{dt} = \frac{1}{2} |\mathbf{k}|$$

which says the radius vector sweeps out area at a constant rate.

G.6 Impulse And Momentum

Work and energy involve a force acting on an object for some distance. Impulse involves a force which acts on an object for an interval of time.

Definition G.6.1 Let $\mathbf{F}$ be a force which acts on an object during the time interval, $[a, b]$. The **impulse** of this force is

$$\int_a^b \mathbf{F}(t) dt.$$

This is defined as

$$\left(\int_a^b F_1(t) dt, \int_a^b F_2(t) dt, \int_a^b F_3(t) dt \right).$$

The **linear momentum** of an object of mass m and velocity $\mathbf{v}$ is defined as

$$\text{Linear momentum} = m\mathbf{v}.$$

The notion of impulse and momentum are related in the following theorem.

Theorem G.6.2 Let $\mathbf{F}$ be a force acting on an object of mass m . Then the impulse equals the change in momentum. More precisely,

$$\int_a^b \mathbf{F}(t) dt = m\mathbf{v}(b) - m\mathbf{v}(a).$$

Proof: This is really just the fundamental theorem of calculus and Newton's second law applied to the components of $\mathbf{F}$.

$$\int_a^b \mathbf{F}(t) dt = \int_a^b m \frac{d\mathbf{v}}{dt} dt = m\mathbf{v}(b) - m\mathbf{v}(a) \quad (7.11)$$

Now suppose two point masses, A and B collide. Newton's third law says the force exerted by mass A on mass B is equal in magnitude but opposite in direction to the force exerted by mass B on mass A . Letting the collision take place in the time interval, $[a, b]$ and denoting the two masses by m_A and m_B and their velocities by $\mathbf{v}_A$ and $\mathbf{v}_B$ it follows that

$$m_A \mathbf{v}_A(b) - m_A \mathbf{v}_A(a) = \int_a^b (\text{Force of } B \text{ on } A) dt$$

and

$$\begin{aligned} m_B \mathbf{v}_B(b) - m_B \mathbf{v}_B(a) &= \int_a^b (\text{Force of } A \text{ on } B) dt \\ &= - \int_a^b (\text{Force of } B \text{ on } A) dt \\ &= - (m_A \mathbf{v}_A(b) - m_A \mathbf{v}_A(a)) \end{aligned}$$

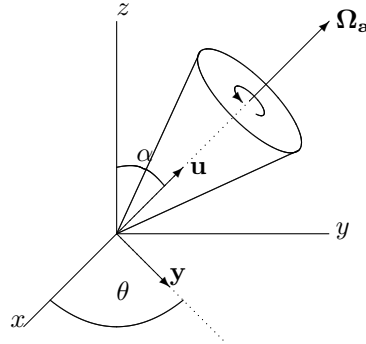
and this shows

$$m_B \mathbf{v}_B(b) + m_A \mathbf{v}_A(b) = m_B \mathbf{v}_B(a) + m_A \mathbf{v}_A(a).$$

In other words, in a collision between two masses the total linear momentum before the collision equals the total linear momentum after the collision. This is known as the conservation of linear momentum.

G.6.1 The Spinning Top

To begin with consider a spinning top as illustrated in the following picture.



For the purpose of this discussion, consider the top as a large number of point masses, m_i , located at the positions, $\mathbf{r}_i(t)$ for $i = 1, 2, \dots, N$ and these masses are symmetrically arranged relative to the axis of the top. As the top spins, the axis of symmetry is observed to move around the z axis. This is called precession and you will see it occur whenever you spin a top. What is the speed of this precession? In other words, what is θ' ? The following discussion follows one given in Sears and Zemansky [26].

Imagine a coordinate system which is fixed relative to the moving top. Thus in this coordinate system the points of the top are fixed. Let the standard unit vectors of the coordinate system moving with the top be denoted by $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$. From Theorem G.2.2 on Page 710, there exists an angular velocity vector $\mathbf{\Omega}(t)$ such that if $\mathbf{u}(t)$ is the position vector of a point fixed in the top, ($\mathbf{u}(t) = u_1 \mathbf{i}(t) + u_2 \mathbf{j}(t) + u_3 \mathbf{k}(t)$),

$$\mathbf{u}'(t) = \mathbf{\Omega}(t) \times \mathbf{u}(t).$$

The vector $\mathbf{\Omega}_a$ shown in the picture is the vector for which

$$\mathbf{r}'_i(t) \equiv \mathbf{\Omega}_a \times \mathbf{r}_i(t)$$

is the velocity of the i^{th} point mass due to rotation about the axis of the top. Thus $\mathbf{\Omega}(t) = \mathbf{\Omega}_a(t) + \mathbf{\Omega}_p(t)$ and it is assumed $\mathbf{\Omega}_p(t)$ is very small relative to $\mathbf{\Omega}_a$. In other words, it is assumed the axis of the top moves very slowly relative to the speed of the points in the top which are spinning very fast around the axis of the top. The angular momentum, $\mathbf{L}$ is defined by

$$\mathbf{L} \equiv \sum_{i=1}^N \mathbf{r}_i \times m_i \mathbf{v}_i \quad (7.12)$$

where $\mathbf{v}_i$ equals the velocity of the i^{th} point mass. Thus $\mathbf{v}_i = \mathbf{\Omega}(t) \times \mathbf{r}_i$ and from the above assumption, $\mathbf{v}_i$ may be taken equal to $\mathbf{\Omega}_a \times \mathbf{r}_i$. Therefore, $\mathbf{L}$ is essentially given by

$$\begin{aligned} \mathbf{L} &\equiv \sum_{i=1}^N m_i \mathbf{r}_i \times (\mathbf{\Omega}_a \times \mathbf{r}_i) \\ &= \sum_{i=1}^N m_i \left(|\mathbf{r}_i|^2 \mathbf{\Omega}_a - (\mathbf{r}_i \cdot \mathbf{\Omega}_a) \mathbf{r}_i \right). \end{aligned}$$

By symmetry of the top, this last expression equals a multiple of $\mathbf{\Omega}_a$. Thus $\mathbf{L}$ is parallel to $\mathbf{\Omega}_a$. Also,

$$\begin{aligned} \mathbf{L} \cdot \mathbf{\Omega}_a &= \sum_{i=1}^N m_i \mathbf{\Omega}_a \cdot \mathbf{r}_i \times (\mathbf{\Omega}_a \times \mathbf{r}_i) \\ &= \sum_{i=1}^N m_i (\mathbf{\Omega}_a \times \mathbf{r}_i) \cdot (\mathbf{\Omega}_a \times \mathbf{r}_i) \\ &= \sum_{i=1}^N m_i |\mathbf{\Omega}_a \times \mathbf{r}_i|^2 = \sum_{i=1}^N m_i |\mathbf{\Omega}_a|^2 |\mathbf{r}_i|^2 \sin^2(\beta_i) \end{aligned}$$

where β_i denotes the angle between the position vector of the i^{th} point mass and the axis of the top. Since this expression is positive, this also shows $\mathbf{L}$ has the same direction as $\mathbf{\Omega}_a$. Let $\omega \equiv |\mathbf{\Omega}_a|$. Then the above expression is of the form

$$\mathbf{L} \cdot \mathbf{\Omega}_a = I\omega^2,$$

where

$$I \equiv \sum_{i=1}^N m_i |\mathbf{r}_i|^2 \sin^2(\beta_i).$$

Thus, to get I you take the mass of the i^{th} point mass, multiply it by the square of its distance to the axis of the top and add all these up. This is defined as the moment of inertia of the top about the axis of the top. Letting $\mathbf{u}$ denote a unit vector in the direction of the axis of the top, this implies

$$\mathbf{L} = I\omega\mathbf{u}. \quad (7.13)$$

Note the simple description of the angular momentum in terms of the moment of inertia. Referring to the above picture, define the vector, $\mathbf{y}$ to be the projection of the vector, $\mathbf{u}$ on the xy plane. Thus

$$\mathbf{y} = \mathbf{u} - (\mathbf{u} \cdot \mathbf{k}) \mathbf{k}$$

and

$$(\mathbf{u} \cdot \mathbf{i}) = (\mathbf{y} \cdot \mathbf{i}) = \sin \alpha \cos \theta. \quad (7.14)$$

Now also from 7.12,

$$\begin{aligned} \frac{d\mathbf{L}}{dt} &= \sum_{i=1}^N m_i \overbrace{\mathbf{r}'_i \times \mathbf{v}_i}^{=0} + \mathbf{r}_i \times m_i \mathbf{v}'_i \\ &= \sum_{i=1}^N \mathbf{r}_i \times m_i \mathbf{v}'_i = - \sum_{i=1}^N \mathbf{r}_i \times m_i g \mathbf{k} \end{aligned}$$

where g is the acceleration of gravity. From 7.13, 7.14, and the above,

$$\begin{aligned} \frac{d\mathbf{L}}{dt} \cdot \mathbf{i} &= I\omega \left(\frac{d\mathbf{u}}{dt} \cdot \mathbf{i} \right) = I\omega \left(\frac{d\mathbf{y}}{dt} \cdot \mathbf{i} \right) \\ &= (-I\omega \sin \alpha \sin \theta) \theta' = - \sum_{i=1}^N \mathbf{r}_i \times m_i g \mathbf{k} \cdot \mathbf{i} \\ &= - \sum_{i=1}^N m_i g \mathbf{r}_i \cdot \mathbf{k} \times \mathbf{i} = - \sum_{i=1}^N m_i g \mathbf{r}_i \cdot \mathbf{j}. \end{aligned} \quad (7.15)$$

To simplify this further, recall the following definition of the center of mass.

Definition G.6.3 Define the total mass, M by

$$M = \sum_{i=1}^N m_i$$

and the center of mass, $\mathbf{r}_0$ by

$$\mathbf{r}_0 \equiv \frac{\sum_{i=1}^N \mathbf{r}_i m_i}{M}. \quad (7.16)$$

In terms of the center of mass, the last expression equals

$$\begin{aligned} -Mg \mathbf{r}_0 \cdot \mathbf{j} &= -Mg (\mathbf{r}_0 - (\mathbf{r}_0 \cdot \mathbf{k}) \mathbf{k} + (\mathbf{r}_0 \cdot \mathbf{k}) \mathbf{k}) \cdot \mathbf{j} \\ &= -Mg (\mathbf{r}_0 - (\mathbf{r}_0 \cdot \mathbf{k}) \mathbf{k}) \cdot \mathbf{j} \\ &= -Mg |\mathbf{r}_0 - (\mathbf{r}_0 \cdot \mathbf{k}) \mathbf{k}| \cos \theta \\ &= -Mg |\mathbf{r}_0| \sin \alpha \cos \left(\frac{\pi}{2} - \theta \right). \end{aligned}$$

Note that by symmetry, $\mathbf{r}_0(t)$ is on the axis of the top, is in the same direction as $\mathbf{L}$, $\mathbf{u}$, and $\boldsymbol{\Omega}_a$, and also $|\mathbf{r}_0|$ is independent of t . Therefore, from the second line of 7.15,

$$(-I\omega \sin \alpha \sin \theta) \theta' = -Mg |\mathbf{r}_0| \sin \alpha \sin \theta.$$

which shows

$$\theta' = \frac{Mg |\mathbf{r}_0|}{I\omega}. \quad (7.17)$$

From 7.17, the angular velocity of precession does not depend on α in the picture. It also is slower when ω is large and I is large.

The above discussion is a considerable simplification of the problem of a spinning top obtained from an assumption that $\boldsymbol{\Omega}_a$ is approximately equal to $\boldsymbol{\Omega}$. It also leaves out all considerations of friction and the observation that the axis of symmetry wobbles. This wobbling is called **nutation**. The full mathematical treatment of this problem involves the Euler angles and some fairly complicated differential equations obtained using techniques discussed in advanced physics classes. Lagrange studied these types of problems back in the 1700's.

G.6.2 Kinetic Energy

The next problem is that of understanding the total kinetic energy of a collection of moving point masses. Consider a possibly large number of point masses, m_i located at the positions $\mathbf{r}_i$ for $i = 1, 2, \dots, N$. Thus the velocity of the i^{th} point mass is $\mathbf{r}'_i = \mathbf{v}_i$. The kinetic energy of the mass m_i is defined by

$$\frac{1}{2} m_i |\mathbf{r}'_i|^2.$$

(This is a very good time to review the presentation on kinetic energy given on Page 714.) The total kinetic energy of the collection of masses is then

$$E = \sum_{i=1}^N \frac{1}{2} m_i |\mathbf{r}'_i|^2. \quad (7.18)$$

As these masses move about, so does the center of mass, $\mathbf{r}_0$. Thus $\mathbf{r}_0$ is a function of t just as the other $\mathbf{r}_i$. From 7.18 the total kinetic energy is

$$\begin{aligned} E &= \sum_{i=1}^N \frac{1}{2} m_i |\mathbf{r}'_i - \mathbf{r}'_0 + \mathbf{r}'_0|^2 \\ &= \sum_{i=1}^N \frac{1}{2} m_i \left[|\mathbf{r}'_i - \mathbf{r}'_0|^2 + |\mathbf{r}'_0|^2 + 2(\mathbf{r}'_i - \mathbf{r}'_0) \cdot \mathbf{r}'_0 \right]. \end{aligned} \quad (7.19)$$

Now

$$\begin{aligned} \sum_{i=1}^N m_i (\mathbf{r}'_i - \mathbf{r}'_0) \cdot \mathbf{r}'_0 &= \left(\sum_{i=1}^N m_i (\mathbf{r}_i - \mathbf{r}_0) \right)' \cdot \mathbf{r}'_0 \\ &= 0 \end{aligned}$$

because from 7.16

$$\begin{aligned} \sum_{i=1}^N m_i (\mathbf{r}_i - \mathbf{r}_0) &= \sum_{i=1}^N m_i \mathbf{r}_i - \sum_{i=1}^N m_i \mathbf{r}_0 \\ &= \sum_{i=1}^N m_i \mathbf{r}_i - \sum_{i=1}^N m_i \left(\frac{\sum_{i=1}^N \mathbf{r}_i m_i}{\sum_{i=1}^N m_i} \right) = \mathbf{0}. \end{aligned}$$

Let $M \equiv \sum_{i=1}^N m_i$ be the total mass. Then 7.19 reduces to

$$\begin{aligned} E &= \sum_{i=1}^N \frac{1}{2} m_i \left[|\mathbf{r}'_i - \mathbf{r}'_0|^2 + |\mathbf{r}'_0|^2 \right] \\ &= \frac{1}{2} M |\mathbf{r}'_0|^2 + \sum_{i=1}^N \frac{1}{2} m_i |\mathbf{r}'_i - \mathbf{r}'_0|^2. \end{aligned} \quad (7.20)$$

The first term is just the kinetic energy of a point mass equal to the sum of all the masses involved, located at the center of mass of the system of masses while the second term represents kinetic energy which comes from the relative velocities of the masses taken with respect to the center of mass. It is this term which is considered more carefully in the case where the system of masses maintain distance between each other.

To illustrate the contrast between the case where the masses maintain a constant distance and one in which they don't, take a hard boiled egg and spin it and then take

a raw egg and give it a spin. You will certainly feel a big difference in the way the two eggs respond. Incidentally, this is a good way to tell whether the egg has been hard boiled or is raw and can be used to prevent messiness which could occur if you think it is hard boiled and it really isn't.

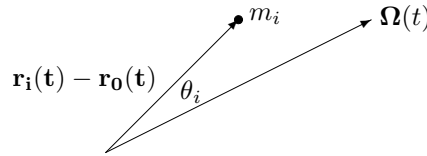
Now let $\mathbf{e}_1(t)$, $\mathbf{e}_2(t)$, and $\mathbf{e}_3(t)$ be an orthonormal set of vectors which is fixed in the body undergoing rigid body motion. This means that $\mathbf{r}_i(t) - \mathbf{r}_0(t)$ has components which are constant in t with respect to the vectors, $\mathbf{e}_i(t)$. By Theorem G.2.2 on Page 710 there exists a vector, $\boldsymbol{\Omega}(t)$ which does not depend on i such that

$$\mathbf{r}'_i(t) - \mathbf{r}'_0(t) = \boldsymbol{\Omega}(t) \times (\mathbf{r}_i(t) - \mathbf{r}_0(t)).$$

Now using this in 7.20,

$$\begin{aligned} E &= \frac{1}{2}M |\mathbf{r}'_0|^2 + \sum_{i=1}^N \frac{1}{2}m_i |\boldsymbol{\Omega}(t) \times (\mathbf{r}_i(t) - \mathbf{r}_0(t))|^2 \\ &= \frac{1}{2}M |\mathbf{r}'_0|^2 + \frac{1}{2} \left(\sum_{i=1}^N m_i |\mathbf{r}_i(t) - \mathbf{r}_0(t)|^2 \sin^2 \theta_i \right) |\boldsymbol{\Omega}(t)|^2 \\ &= \frac{1}{2}M |\mathbf{r}'_0|^2 + \frac{1}{2} \left(\sum_{i=1}^N m_i |\mathbf{r}_i(0) - \mathbf{r}_0(0)|^2 \sin^2 \theta_i \right) |\boldsymbol{\Omega}(t)|^2 \end{aligned}$$

where θ_i is the angle between $\boldsymbol{\Omega}(t)$ and the vector, $\mathbf{r}_i(t) - \mathbf{r}_0(t)$. Therefore, $|\mathbf{r}_i(t) - \mathbf{r}_0(t)| \sin \theta_i$ is the distance between the point mass, m_i located at $\mathbf{r}_i$ and a line through the center of mass, $\mathbf{r}_0$ with direction, $\boldsymbol{\Omega}$ as indicated in the following picture.



Thus the expression, $\sum_{i=1}^N m_i |\mathbf{r}_i(0) - \mathbf{r}_0(0)|^2 \sin^2 \theta_i$ plays the role of a mass in the definition of kinetic energy except instead of the speed, substitute the angular speed, $|\boldsymbol{\Omega}(t)|$. It is this expression which is called the moment of inertia about the line whose direction is $\boldsymbol{\Omega}(t)$.

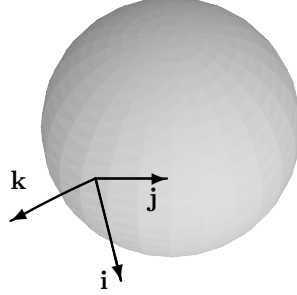
In both of these examples, the center of mass and the moment of inertia occurred in a natural way.

G.7 Acceleration With Respect To Moving Coordinate Systems*

The idea is you have a coordinate system which is moving and this results in strange forces experienced relative to these moving coordinates systems. A good example is what we experience every day living on a rotating ball. Relative to our supposedly fixed coordinate system, we experience forces which account for many phenomena which are observed.

G.7.1 The Coriolis Acceleration

Imagine a point on the surface of the earth. Now consider unit vectors, one pointing South, one pointing East and one pointing directly away from the center of the earth.



Denote the first as $\mathbf{i}$, the second as $\mathbf{j}$ and the third as $\mathbf{k}$. If you are standing on the earth, you will consider these vectors as fixed, but of course they are not. As the earth turns, they change direction and so each is in reality a function of t . Nevertheless, it is with respect to these apparently fixed vectors that you wish to understand acceleration, velocities, and displacements.

In general, let $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ be an orthonormal basis of vectors for each t , like the vectors described in the first paragraph. It is assumed these vectors are C^1 functions of t . Letting the positive x axis extend in the direction of $\mathbf{i}(t)$, the positive y axis extend in the direction of $\mathbf{j}(t)$, and the positive z axis extend in the direction of $\mathbf{k}(t)$, yields a moving coordinate system. By Theorem G.2.2 on Page 710, there exists an angular velocity vector, $\boldsymbol{\Omega}(t)$ such that if $\mathbf{u}(t)$ is any vector which has constant components with respect to $\mathbf{i}(t), \mathbf{j}(t)$, and $\mathbf{k}(t)$, then

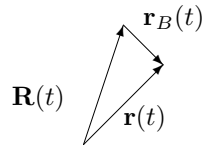
$$\boldsymbol{\Omega} \times \mathbf{u} = \mathbf{u}' \quad (7.21)$$

Now let $\mathbf{R}(t)$ be a position vector of the moving coordinate system and let

$$\mathbf{r}(t) = \mathbf{R}(t) + \mathbf{r}_B(t)$$

where

$$\mathbf{r}_B(t) \equiv x(t)\mathbf{i}(t) + y(t)\mathbf{j}(t) + z(t)\mathbf{k}(t).$$



In the example of the earth, $\mathbf{R}(t)$ is the position vector of a point $\mathbf{p}(t)$ on the earth's surface and $\mathbf{r}_B(t)$ is the position vector of another point from $\mathbf{p}(t)$, thus regarding $\mathbf{p}(t)$ as the origin. $\mathbf{r}_B(t)$ is the position vector of a point as perceived by the observer on the earth with respect to the vectors he thinks of as fixed. Similarly, $\mathbf{v}_B(t)$ and $\mathbf{a}_B(t)$ will be the velocity and acceleration relative to $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$, and so $\mathbf{v}_B = x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}$ and $\mathbf{a}_B = x''\mathbf{i} + y''\mathbf{j} + z''\mathbf{k}$. Then

$$\mathbf{v} \equiv \mathbf{r}' = \mathbf{R}' + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} + x\mathbf{i}' + y\mathbf{j}' + z\mathbf{k}'.$$

By 7.21, if $\mathbf{e} \in \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, $\mathbf{e}' = \boldsymbol{\Omega} \times \mathbf{e}$ because the components of these vectors with respect to $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are constant. Therefore,

$$\begin{aligned} x\mathbf{i}' + y\mathbf{j}' + z\mathbf{k}' &= x\boldsymbol{\Omega} \times \mathbf{i} + y\boldsymbol{\Omega} \times \mathbf{j} + z\boldsymbol{\Omega} \times \mathbf{k} \\ &= \boldsymbol{\Omega} \times (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) \end{aligned}$$

and consequently,

$$\mathbf{v} = \mathbf{R}' + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} + \boldsymbol{\Omega} \times \mathbf{r}_B = \mathbf{R}' + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} + \boldsymbol{\Omega} \times (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}).$$

Now consider the acceleration. Quantities which are relative to the moving coordinate system are distinguished by using the subscript, B .

$$\begin{aligned} \mathbf{a} = \mathbf{v}' &= \mathbf{R}'' + x''\mathbf{i} + y''\mathbf{j} + z''\mathbf{k} + \overbrace{x'\mathbf{i}' + y'\mathbf{j}' + z'\mathbf{k}'}^{\boldsymbol{\Omega} \times \mathbf{v}_B} + \boldsymbol{\Omega}' \times \mathbf{r}_B \\ &+ \boldsymbol{\Omega} \times \left(\overbrace{x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}}^{\mathbf{v}_B} + \overbrace{x\mathbf{i}' + y\mathbf{j}' + z\mathbf{k}'}^{\boldsymbol{\Omega} \times \mathbf{r}_B(t)} \right) \\ &= \mathbf{R}'' + \mathbf{a}_B + \boldsymbol{\Omega}' \times \mathbf{r}_B + 2\boldsymbol{\Omega} \times \mathbf{v}_B + \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}_B). \end{aligned}$$

The acceleration $\mathbf{a}_B$ is that perceived by an observer for whom the moving coordinate system is fixed. The term $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}_B)$ is called the centripetal acceleration. Solving for $\mathbf{a}_B$,

$$\mathbf{a}_B = \mathbf{a} - \mathbf{R}'' - \boldsymbol{\Omega}' \times \mathbf{r}_B - 2\boldsymbol{\Omega} \times \mathbf{v}_B - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}_B). \quad (7.22)$$

Here the term $-(\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}_B))$ is called the centrifugal acceleration, it being an acceleration felt by the observer relative to the moving coordinate system which he regards as fixed, and the term $-2\boldsymbol{\Omega} \times \mathbf{v}_B$ is called the Coriolis acceleration, an acceleration experienced by the observer as he moves relative to the moving coordinate system. The mass multiplied by the Coriolis acceleration defines the Coriolis force.

There is a ride found in some amusement parks in which the victims stand next to a circular wall covered with a carpet or some rough material. Then the whole circular room begins to revolve faster and faster. At some point, the bottom drops out and the victims are held in place by friction. The force they feel which keeps them stuck to the wall is called centrifugal force and it causes centrifugal acceleration. It is not necessary to move relative to coordinates fixed with the revolving wall in order to feel this force and it is pretty predictable. However, if the nauseated victim moves relative to the rotating wall, he will feel the effects of the Coriolis force and this force is really strange. The difference between these forces is that the Coriolis force is caused by movement relative to the moving coordinate system and the centrifugal force is not.

G.7.2 The Coriolis Acceleration On The Rotating Earth

Now consider the earth. Let $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$ be the usual basis vectors attached to the rotating earth. Thus $\mathbf{k}^*$ is fixed in space with $\mathbf{k}^*$ pointing in the direction of the north pole from the center of the earth while $\mathbf{i}^*$ and $\mathbf{j}^*$ point to fixed points on the surface of the earth. Thus $\mathbf{i}^*$ and $\mathbf{j}^*$ depend on t while $\mathbf{k}^*$ does not. Let $\mathbf{i}, \mathbf{j}, \mathbf{k}$ be the unit vectors described earlier with $\mathbf{i}$ pointing South, $\mathbf{j}$ pointing East, and $\mathbf{k}$ pointing away from the center of the earth at some point of the rotating earth's surface, $\mathbf{p}$. Letting $\mathbf{R}(t)$ be the position vector of the point $\mathbf{p}$, from the center of the earth, observe the coordinates of $\mathbf{R}(t)$ are constant with respect to $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$. Also, since the earth rotates from West to East and the speed of a point on the surface of the earth relative to an observer fixed in space is $\omega|\mathbf{R}|\sin\phi$ where ω is the angular speed of the earth about an axis through the poles, it follows from the geometric definition of the cross product that

$$\mathbf{R}' = \omega \mathbf{k}^* \times \mathbf{R}$$

Therefore, $\boldsymbol{\Omega} = \omega \mathbf{k}^*$ and so

$$\mathbf{R}'' = \overbrace{\boldsymbol{\Omega}' \times \mathbf{R}}^{=0} + \boldsymbol{\Omega} \times \mathbf{R}' = \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R})$$

since $\boldsymbol{\Omega}$ does not depend on t . Formula 7.22 implies

$$\mathbf{a}_B = \mathbf{a} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R}) - 2\boldsymbol{\Omega} \times \mathbf{v}_B - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}_B). \quad (7.23)$$

In this formula, you can totally ignore the term $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}_B)$ because it is so small whenever you are considering motion near some point on the earth's surface. To see

this, note $\omega \overbrace{(24)(3600)}^{\text{seconds in a day}} = 2\pi$, and so $\omega = 7.2722 \times 10^{-5}$ in radians per second. If you are using seconds to measure time and feet to measure distance, this term is therefore, no larger than

$$(7.2722 \times 10^{-5})^2 |\mathbf{r}_B|.$$

Clearly this is not worth considering in the presence of the acceleration due to gravity which is approximately 32 feet per second squared near the surface of the earth.

If the acceleration $\mathbf{a}$, is due to gravity, then

$$\begin{aligned} \mathbf{a}_B &= \mathbf{a} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R}) - 2\boldsymbol{\Omega} \times \mathbf{v}_B = \\ &= \underbrace{-\frac{GM(\mathbf{R} + \mathbf{r}_B)}{|\mathbf{R} + \mathbf{r}_B|^3}}_{\equiv \mathbf{g}} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R}) - 2\boldsymbol{\Omega} \times \mathbf{v}_B \equiv \mathbf{g} - 2\boldsymbol{\Omega} \times \mathbf{v}_B. \end{aligned}$$

Note that

$$\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R}) = (\boldsymbol{\Omega} \cdot \mathbf{R}) \boldsymbol{\Omega} - |\boldsymbol{\Omega}|^2 \mathbf{R}$$

and so $\mathbf{g}$, the acceleration relative to the moving coordinate system on the earth is not directed exactly toward the center of the earth except at the poles and at the equator, although the components of acceleration which are in other directions are very small when compared with the acceleration due to the force of gravity and are often neglected. Therefore, if the only force acting on an object is due to gravity, the following formula describes the acceleration relative to a coordinate system moving with the earth's surface.

$$\mathbf{a}_B = \mathbf{g} - 2(\boldsymbol{\Omega} \times \mathbf{v}_B)$$

While the vector, $\boldsymbol{\Omega}$ is quite small, if the relative velocity, $\mathbf{v}_B$ is large, the Coriolis acceleration could be significant. This is described in terms of the vectors $\mathbf{i}(t), \mathbf{j}(t), \mathbf{k}(t)$ next.

Letting (ρ, θ, ϕ) be the usual spherical coordinates of the point $\mathbf{p}(t)$ on the surface taken with respect to $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$ the usual way with ϕ the polar angle, it follows the $\mathbf{i}^*, \mathbf{j}^*, \mathbf{k}^*$ coordinates of this point are

$$\begin{pmatrix} \rho \sin(\phi) \cos(\theta) \\ \rho \sin(\phi) \sin(\theta) \\ \rho \cos(\phi) \end{pmatrix}.$$

It follows,

$$\begin{aligned} \mathbf{i} &= \cos(\phi) \cos(\theta) \mathbf{i}^* + \cos(\phi) \sin(\theta) \mathbf{j}^* - \sin(\phi) \mathbf{k}^* \\ \mathbf{j} &= -\sin(\theta) \mathbf{i}^* + \cos(\theta) \mathbf{j}^* + 0 \mathbf{k}^* \end{aligned}$$

and

$$\mathbf{k} = \sin(\phi) \cos(\theta) \mathbf{i}^* + \sin(\phi) \sin(\theta) \mathbf{j}^* + \cos(\phi) \mathbf{k}^*.$$

It is necessary to obtain $\mathbf{k}^*$ in terms of the vectors, $\mathbf{i}, \mathbf{j}, \mathbf{k}$. Thus the following equation needs to be solved for a, b, c to find $\mathbf{k}^* = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$

$$\overbrace{\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}^{\mathbf{k}^*} = \begin{pmatrix} \cos(\phi) \cos(\theta) & -\sin(\theta) & \sin(\phi) \cos(\theta) \\ \cos(\phi) \sin(\theta) & \cos(\theta) & \sin(\phi) \sin(\theta) \\ -\sin(\phi) & 0 & \cos(\phi) \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad (7.24)$$

The first column is $\mathbf{i}$, the second is $\mathbf{j}$ and the third is $\mathbf{k}$ in the above matrix. The solution is $a = -\sin(\phi)$, $b = 0$, and $c = \cos(\phi)$.

Now the Coriolis acceleration on the earth equals

$$2(\boldsymbol{\Omega} \times \mathbf{v}_B) = 2\omega \left(\overbrace{-\sin(\phi)\mathbf{i} + 0\mathbf{j} + \cos(\phi)\mathbf{k}}^{\mathbf{k}^*} \right) \times (x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}).$$

This equals

$$2\omega [(-y' \cos \phi)\mathbf{i} + (x' \cos \phi + z' \sin \phi)\mathbf{j} - (y' \sin \phi)\mathbf{k}]. \quad (7.25)$$

Remember ϕ is fixed and pertains to the fixed point, $\mathbf{p}(t)$ on the earth's surface. Therefore, if the acceleration, $\mathbf{a}$ is due to gravity,

$$\mathbf{a}_B = \mathbf{g} - 2\omega [(-y' \cos \phi)\mathbf{i} + (x' \cos \phi + z' \sin \phi)\mathbf{j} - (y' \sin \phi)\mathbf{k}]$$

where $\mathbf{g} = -\frac{GM(\mathbf{R} + \mathbf{r}_B)}{|\mathbf{R} + \mathbf{r}_B|^3} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R})$ as explained above. The term $\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R})$ is pretty small and so it will be neglected. However, the Coriolis force will not be neglected.

Example G.7.1 Suppose a rock is dropped from a tall building. Where will it strike?

Assume $\mathbf{a} = -g\mathbf{k}$ and the $\mathbf{j}$ component of $\mathbf{a}_B$ is approximately

$$-2\omega (x' \cos \phi + z' \sin \phi).$$

The dominant term in this expression is clearly the second one because x' will be small. Also, the $\mathbf{i}$ and $\mathbf{k}$ contributions will be very small. Therefore, the following equation is descriptive of the situation.

$$\mathbf{a}_B = -g\mathbf{k} - 2z'\omega \sin \phi \mathbf{j}.$$

$z' = -gt$ approximately. Therefore, considering the $\mathbf{j}$ component, this is

$$2gt\omega \sin \phi.$$

Two integrations give $(\omega gt^3/3) \sin \phi$ for the $\mathbf{j}$ component of the relative displacement at time t .

This shows the rock does not fall directly towards the center of the earth as expected but slightly to the east.

Example G.7.2 In 1851 Foucault set a pendulum vibrating and observed the earth rotate out from under it. It was a very long pendulum with a heavy weight at the end so that it would vibrate for a long time without stopping³. This is what allowed him to observe the earth rotate out from under it. Clearly such a pendulum will take 24 hours for the plane of vibration to appear to make one complete revolution at the north pole. It is also reasonable to expect that no such observed rotation would take place on the equator. Is it possible to predict what will take place at various latitudes?

Using 7.25, in 7.23,

$$\mathbf{a}_B = \mathbf{a} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{R}) - 2\omega [(-y' \cos \phi)\mathbf{i} + (x' \cos \phi + z' \sin \phi)\mathbf{j} - (y' \sin \phi)\mathbf{k}].$$

³There is such a pendulum in the Eyring building at BYU and to keep people from touching it, there is a little sign which says Warning! 1000 ohms.

Neglecting the small term, $\mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{R})$, this becomes

$$= -g\mathbf{k} + \mathbf{T}/m - 2\omega [(-y' \cos \phi)\mathbf{i} + (x' \cos \phi + z' \sin \phi)\mathbf{j} - (y' \sin \phi)\mathbf{k}]$$

where $\mathbf{T}$, the tension in the string of the pendulum, is directed towards the point at which the pendulum is supported, and m is the mass of the pendulum bob. The pendulum can be thought of as the position vector from $(0, 0, l)$ to the surface of the sphere $x^2 + y^2 + (z - l)^2 = l^2$. Therefore,

$$\mathbf{T} = -T \frac{x}{l} \mathbf{i} - T \frac{y}{l} \mathbf{j} + T \frac{l - z}{l} \mathbf{k}$$

and consequently, the differential equations of relative motion are

$$\begin{aligned} x'' &= -T \frac{x}{ml} + 2\omega y' \cos \phi \\ y'' &= -T \frac{y}{ml} - 2\omega (x' \cos \phi + z' \sin \phi) \end{aligned}$$

and

$$z'' = T \frac{l - z}{ml} - g + 2\omega y' \sin \phi.$$

If the vibrations of the pendulum are small so that for practical purposes, $z'' = z = 0$, the last equation may be solved for T to get

$$gm - 2\omega y' \sin(\phi) m = T.$$

Therefore, the first two equations become

$$x'' = -(gm - 2\omega m y' \sin \phi) \frac{x}{ml} + 2\omega y' \cos \phi$$

and

$$y'' = -(gm - 2\omega m y' \sin \phi) \frac{y}{ml} - 2\omega (x' \cos \phi + z' \sin \phi).$$

All terms of the form xy' or $y'y$ can be neglected because it is assumed x and y remain small. Also, the pendulum is assumed to be long with a heavy weight so that x' and y' are also small. With these simplifying assumptions, the equations of motion become

$$x'' + g \frac{x}{l} = 2\omega y' \cos \phi$$

and

$$y'' + g \frac{y}{l} = -2\omega x' \cos \phi.$$

These equations are of the form

$$x'' + a^2 x = by', \quad y'' + a^2 y = -bx' \quad (7.26)$$

where $a^2 = \frac{g}{l}$ and $b = 2\omega \cos \phi$. Then it is fairly tedious but routine to verify that for each constant, c ,

$$x = c \sin\left(\frac{bt}{2}\right) \sin\left(\frac{\sqrt{b^2 + 4a^2}}{2} t\right), \quad y = c \cos\left(\frac{bt}{2}\right) \sin\left(\frac{\sqrt{b^2 + 4a^2}}{2} t\right) \quad (7.27)$$

yields a solution to 7.26 along with the initial conditions,

$$x(0) = 0, y(0) = 0, x'(0) = 0, y'(0) = \frac{c\sqrt{b^2 + 4a^2}}{2}. \quad (7.28)$$

It is clear from experiments with the pendulum that the earth does indeed rotate out from under it causing the plane of vibration of the pendulum to appear to rotate. The purpose of this discussion is not to establish these self evident facts but to predict how long it takes for the plane of vibration to make one revolution. Therefore, there will be some instant in time at which the pendulum will be vibrating in a plane determined by $\mathbf{k}$ and $\mathbf{j}$. (Recall $\mathbf{k}$ points away from the center of the earth and $\mathbf{j}$ points East.) At this instant in time, defined as $t = 0$, the conditions of 7.28 will hold for some value of c and so the solution to 7.26 having these initial conditions will be those of 7.27 by uniqueness of the initial value problem. Writing these solutions differently,

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = c \begin{pmatrix} \sin\left(\frac{bt}{2}\right) \\ \cos\left(\frac{bt}{2}\right) \end{pmatrix} \sin\left(\frac{\sqrt{b^2 + 4a^2}}{2}t\right)$$

This is very interesting! The vector, $c \begin{pmatrix} \sin\left(\frac{bt}{2}\right) \\ \cos\left(\frac{bt}{2}\right) \end{pmatrix}$ always has magnitude equal to $|c|$ but its direction changes very slowly because b is very small. The plane of vibration is determined by this vector and the vector $\mathbf{k}$. The term $\sin\left(\frac{\sqrt{b^2 + 4a^2}}{2}t\right)$ changes relatively fast and takes values between -1 and 1 . This is what describes the actual observed vibrations of the pendulum. Thus the plane of vibration will have made one complete revolution when $t = P$ for

$$\frac{bP}{2} \equiv 2\pi.$$

Therefore, the time it takes for the earth to turn out from under the pendulum is

$$P = \frac{4\pi}{2\omega \cos \phi} = \frac{2\pi}{\omega} \sec \phi.$$

Since ω is the angular speed of the rotating earth, it follows $\omega = \frac{2\pi}{24} = \frac{\pi}{12}$ in radians per hour. Therefore, the above formula implies

$$P = 24 \sec \phi.$$

I think this is really amazing. You could actually determine latitude, not by taking readings with instruments using the North Star but by doing an experiment with a big pendulum. You would set it vibrating, observe P in hours, and then solve the above equation for ϕ . Also note the pendulum would not appear to change its plane of vibration at the equator because $\lim_{\phi \rightarrow \pi/2} \sec \phi = \infty$.

The Coriolis acceleration is also responsible for the phenomenon of the next example.

Example G.7.3 *It is known that low pressure areas rotate counterclockwise as seen from above in the Northern hemisphere but clockwise in the Southern hemisphere. Why?*

Neglect accelerations other than the Coriolis acceleration and the following acceleration which comes from an assumption that the point $\mathbf{p}(t)$ is the location of the lowest pressure.

$$\mathbf{a} = -a(r_B) \mathbf{r}_B$$

where $r_B = r$ will denote the distance from the fixed point $\mathbf{p}(t)$ on the earth's surface which is also the lowest pressure point. Of course the situation could be more complicated but this will suffice to explain the above question. Then the acceleration observed by a person on the earth relative to the apparently fixed vectors, $\mathbf{i}, \mathbf{k}, \mathbf{j}$, is

$$\mathbf{a}_B = -a(r_B)(x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) - 2\omega[-y'\cos(\phi)\mathbf{i} + (x'\cos(\phi) + z'\sin(\phi))\mathbf{j} - (y'\sin(\phi)\mathbf{k})]$$

Therefore, one obtains some differential equations from $\mathbf{a}_B = x''\mathbf{i} + y''\mathbf{j} + z''\mathbf{k}$ by matching the components. These are

$$\begin{aligned}x'' + a(r_B)x &= 2\omega y' \cos \phi \\y'' + a(r_B)y &= -2\omega x' \cos \phi - 2\omega z' \sin(\phi) \\z'' + a(r_B)z &= 2\omega y' \sin \phi\end{aligned}$$

Now remember, the vectors, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are fixed relative to the earth and so are constant vectors. Therefore, from the properties of the determinant and the above differential equations,

$$\begin{aligned}(\mathbf{r}'_B \times \mathbf{r}_B)' &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x' & y' & z' \\ x & y & z \end{vmatrix}' = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x'' & y'' & z'' \\ x & y & z \end{vmatrix} \\ &= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ A & B & C \\ x & y & z \end{vmatrix}\end{aligned}$$

where

$$\begin{aligned}A &= -a(r_B)x + 2\omega y' \cos \phi, \\B &= -a(r_B)y - 2\omega x' \cos \phi - 2\omega z' \sin(\phi), \\C &= -a(r_B)z + 2\omega y' \sin \phi\end{aligned}$$

Then the $\mathbf{k}^{th}$ component of this cross product equals

$$\omega \cos(\phi) (y^2 + x^2)' + 2\omega x z' \sin(\phi).$$

The first term will be negative because it is assumed $\mathbf{p}(t)$ is the location of low pressure causing $y^2 + x^2$ to be a decreasing function. If it is assumed there is not a substantial motion in the $\mathbf{k}$ direction, so that z is fairly constant and the last term can be neglected, then the $\mathbf{k}^{th}$ component of $(\mathbf{r}'_B \times \mathbf{r}_B)'$ is negative provided $\phi \in (0, \frac{\pi}{2})$ and positive if $\phi \in (\frac{\pi}{2}, \pi)$. Beginning with a point at rest, this implies $\mathbf{r}'_B \times \mathbf{r}_B = \mathbf{0}$ initially and then the above implies its $\mathbf{k}^{th}$ component is negative in the upper hemisphere when $\phi < \pi/2$ and positive in the lower hemisphere when $\phi > \pi/2$. Using the right hand and the geometric definition of the cross product, this shows clockwise rotation in the lower hemisphere and counter clockwise rotation in the upper hemisphere.

Note also that as ϕ gets close to $\pi/2$ near the equator, the above reasoning tends to break down because $\cos(\phi)$ becomes close to zero. Therefore, the motion towards the low pressure has to be more pronounced in comparison with the motion in the $\mathbf{k}$ direction in order to draw this conclusion.

G.8 Exercises

1. Show that the solution to $\mathbf{v}' + r\mathbf{v} = \mathbf{c}$ with the initial condition, $\mathbf{v}(0) = \mathbf{v}_0$ is $\mathbf{v}(t) = (\mathbf{v}_0 - \frac{\mathbf{c}}{r})e^{-rt} + (\mathbf{c}/r)$. If $\mathbf{v}$ is velocity and $r = k/m$ where k is a constant for air resistance and m is the mass, and $\mathbf{c} = \mathbf{f}/m$, argue from Newton's second law that this is the equation for finding the velocity, $\mathbf{v}$ of an object acted on by air resistance proportional to the velocity and a constant force, $\mathbf{f}$, possibly from gravity. Does there exist a terminal velocity? What is it? **Hint:** To find the solution to the equation, multiply both sides by e^{rt} . Verify that then $\frac{d}{dt}(e^{rt}\mathbf{v}) = \mathbf{c}e^{rt}$. Then integrating both sides, $e^{rt}\mathbf{v}(t) = \frac{1}{r}\mathbf{c}e^{rt} + \mathbf{C}$. Now you need to find $\mathbf{C}$ from using the initial condition which states $\mathbf{v}(0) = \mathbf{v}_0$.

2. Verify Formula 7.11 carefully by considering the components.
3. Suppose that the air resistance is proportional to the velocity but it is desired to find the constant of proportionality. Describe how you could find this constant.
4. Suppose an object having mass equal to 5 kilograms experiences a time dependent force, $\mathbf{F}(t) = e^{-t}\mathbf{i} + \cos(t)\mathbf{j} + t^2\mathbf{k}$ meters per sec^2 . Suppose also that the object is at the point $(0, 1, 1)$ meters at time $t = 0$ and that its initial velocity at this time is $\mathbf{v} = \mathbf{i} + \mathbf{j} - \mathbf{k}$ meters per sec. Find the position of the object as a function of t .
5. Fill in the details for the derivation of kinetic energy. In particular verify that $m\mathbf{v}'(t) \cdot \mathbf{v}(t) = \frac{m}{2} \frac{d}{dt} |\mathbf{v}(t)|^2$. Also, why would $dW = \mathbf{F}(t) \cdot \mathbf{v}(t) dt$?
6. Suppose the force acting on an object, $\mathbf{F}$ is always perpendicular to the velocity of the object. Thus $\mathbf{F} \cdot \mathbf{v} = 0$. Show that the Kinetic energy of the object is constant. Such forces are sometimes called forces of constraint because they do not contribute to the speed of the object, only its direction.
7. A cannon is fired at an angle, θ from ground level on a vast plain. The speed of the ball as it leaves the mouth of the cannon is known to be s meters per second. Neglecting air resistance, find a formula for how far the cannon ball goes before hitting the ground. Show that the maximum range for the cannon ball is achieved when $\theta = \pi/4$.
8. Suppose in the context of Problem 7 that the cannon ball has mass 10 kilograms and it experiences a force of air resistance which is $.01\mathbf{v}$ Newtons where $\mathbf{v}$ is the velocity in meters per second. The acceleration of gravity is 9.8 meters per sec^2 . Also suppose that the initial speed is 100 meters per second. Find a formula for the displacement, $\mathbf{r}(t)$ of the cannon ball. If the angle of elevation equals $\pi/4$, use a calculator or other means to estimate the time before the cannon ball hits the ground.
9. Show that Newton's first law can be obtained from the second law.
10. Show that if $\mathbf{v}'(t) = \mathbf{0}$, for all $t \in (a, b)$, then there exists a constant vector, $\mathbf{z}$ independent of t such that $\mathbf{v}(t) = \mathbf{z}$ for all t .
11. Suppose an object moves in three dimensional space in such a way that the only force acting on the object is directed toward a single fixed point in three dimensional space. Verify that the motion of the object takes place in a plane. **Hint:** Let $\mathbf{r}(t)$ denote the position vector of the object from the fixed point. Then the force acting on the object must be of the form $g(\mathbf{r}(t))\mathbf{r}(t)$ and by Newton's second law, this equals $m\mathbf{r}''(t)$. Therefore,

$$m\mathbf{r}'' \times \mathbf{r} = g(\mathbf{r})\mathbf{r} \times \mathbf{r} = \mathbf{0}.$$

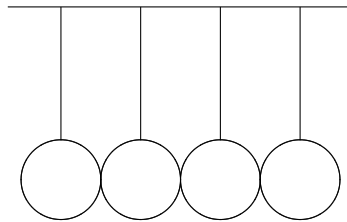
Now argue that $\mathbf{r}'' \times \mathbf{r} = (\mathbf{r}' \times \mathbf{r})'$, showing that $(\mathbf{r}' \times \mathbf{r})$ must equal a constant vector, $\mathbf{z}$. Therefore, what can be said about $\mathbf{z}$ and $\mathbf{r}$?

12. Suppose the only forces acting on an object are the force of gravity, $-mg\mathbf{k}$ and a force, $\mathbf{F}$ which is perpendicular to the motion of the object. Thus $\mathbf{F} \cdot \mathbf{v} = 0$. Show that the total energy of the object,

$$E \equiv \frac{1}{2}m|\mathbf{v}|^2 + mgz$$

is constant. Here $\mathbf{v}$ is the velocity and the first term is the kinetic energy while the second is the potential energy. **Hint:** Use Newton's second law to show that the time derivative of the above expression equals zero.

13. Using Problem 12, suppose an object slides down a frictionless inclined plane from a height of 100 feet. When it reaches the bottom, how fast will it be going? Assume it starts from rest.
14. The ballistic pendulum is an interesting device which is used to determine the speed of a bullet. It is a large massive block of wood hanging from a long string. A rifle is fired into the block of wood which then moves. The speed of the bullet can be determined from measuring how high the block of wood rises. Explain how this can be done and why. **Hint:** Let v be the speed of the bullet which has mass m and let the block of wood have mass M . By conservation of momentum $mv = (m + M)V$ where V is the speed of the block of wood immediately after the collision. Thus the energy is $\frac{1}{2}(m + M)V^2$ and this block of wood rises to a height of h . Now use Problem 12.
15. In the experiment of Problem 14, show that the kinetic energy before the collision is greater than the kinetic energy after the collision. Thus linear momentum is conserved but energy is not. Such a collision is called inelastic.
16. There is a popular toy consisting of identical steel balls suspended from strings of equal length as illustrated in the following picture.



The ball at the right is lifted and allowed to swing. When it collides with the other balls, the ball on the left is observed to swing away from the others with the same speed the ball on the right had when it collided. Why does this happen? Why don't two or more of the stationary balls start to move, perhaps at a slower speed? This is an example of an elastic collision because energy is conserved. Of course this could change if you fixed things so the balls would stick to each other.

17. An illustration used in many beginning physics books is that of firing a rifle horizontally and dropping an identical bullet from the same height above the perfectly flat ground followed by an assertion that the two bullets will hit the ground at exactly the same time. Is this true on the rotating earth assuming the experiment takes place over a large perfectly flat field so the curvature of the earth is not an issue? Explain. What other irregularities will occur? Recall the Coriolis force is $2\omega [(-y' \cos \phi) \mathbf{i} + (x' \cos \phi + z' \sin \phi) \mathbf{j} - (y' \sin \phi) \mathbf{k}]$ where $\mathbf{k}$ points away from the center of the earth, $\mathbf{j}$ points East, and $\mathbf{i}$ points South.
18. Suppose you have n masses, $m_1, \dots, m_n$. Let the position vector of the i^{th} mass be $\mathbf{r}_i(t)$. The center of mass of these is defined to be

$$\mathbf{R}(t) \equiv \frac{\sum_{i=1}^n \mathbf{r}_i m_i}{\sum_{i=1}^n m_i} \equiv \frac{\sum_{i=1}^n \mathbf{r}_i(t) m_i}{M}.$$

Let $\mathbf{r}_{Bi}(t) = \mathbf{r}_i(t) - \mathbf{R}(t)$. Show that $\sum_{i=1}^n m_i \mathbf{r}_i(t) - \sum_i m_i \mathbf{R}(t) = \mathbf{0}$.

19. Suppose you have n masses, $m_1, \dots, m_n$ which make up a moving rigid body. Let $\mathbf{R}(t)$ denote the position vector of the center of mass of these n masses. Find a formula for the total kinetic energy in terms of this position vector, the angular velocity vector, and the position vector of each mass from the center of mass.
Hint: Use Problem 18.

G.9 Exercises With Answers

1. Show that the solution to $\mathbf{v}' + r\mathbf{v} = \mathbf{c}$ with the initial condition, $\mathbf{v}(0) = \mathbf{v}_0$ is $\mathbf{v}(t) = (\mathbf{v}_0 - \frac{\mathbf{c}}{r})e^{-rt} + (\mathbf{c}/r)$. If $\mathbf{v}$ is velocity and $r = k/m$ where k is a constant for air resistance and m is the mass, and $\mathbf{c} = \mathbf{f}/m$, argue from Newton's second law that this is the equation for finding the velocity, $\mathbf{v}$ of an object acted on by air resistance proportional to the velocity and a constant force, $\mathbf{f}$, possibly from gravity. Does there exist a terminal velocity? What is it?

Multiply both sides of the differential equation by e^{rt} . Then the left side becomes $\frac{d}{dt}(e^{rt}\mathbf{v}) = e^{rt}\mathbf{c}$. Now integrate both sides. This gives $e^{rt}\mathbf{v}(t) = \mathbf{C} + \frac{e^{rt}}{r}\mathbf{c}$. You finish the rest.

2. Suppose an object having mass equal to 5 kilograms experiences a time dependent force, $\mathbf{F}(t) = e^{-t}\mathbf{i} + \cos(t)\mathbf{j} + t^2\mathbf{k}$ meters per sec². Suppose also that the object is at the point $(0, 1, 1)$ meters at time $t = 0$ and that its initial velocity at this time is $\mathbf{v} = \mathbf{i} + \mathbf{j} - \mathbf{k}$ meters per sec. Find the position of the object as a function of t .

This is done by using Newton's law. Thus $5\frac{d^2\mathbf{r}}{dt^2} = e^{-t}\mathbf{i} + \cos(t)\mathbf{j} + t^2\mathbf{k}$ and so $5\frac{d\mathbf{r}}{dt} = -e^{-t}\mathbf{i} + \sin(t)\mathbf{j} + (t^3/3)\mathbf{k} + \mathbf{C}$. Find the constant, $\mathbf{C}$ by using the given initial velocity. Next do another integration obtaining another constant vector which will be determined by using the given initial position of the object.

3. Fill in the details for the derivation of kinetic energy. In particular verify that $m\mathbf{v}'(t) \cdot \mathbf{v}(t) = \frac{m}{2} \frac{d}{dt} |\mathbf{v}(t)|^2$. Also, why would $dW = \mathbf{F}(t) \cdot \mathbf{v}(t) dt$?

Remember $|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v}$. Now use the product rule.

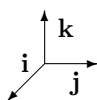
4. Suppose the force acting on an object, $\mathbf{F}$ is always perpendicular to the velocity of the object. Thus $\mathbf{F} \cdot \mathbf{v} = 0$. Show that the Kinetic energy of the object is constant. Such forces are sometimes called forces of constraint because they do not contribute to the speed of the object, only its direction.

$0 = \mathbf{F} \cdot \mathbf{v} = m\mathbf{v}' \cdot \mathbf{v}$. Explain why this is $\frac{d}{dt} \left(m \frac{1}{2} |\mathbf{v}|^2 \right)$, the derivative of the kinetic energy.

Curvilinear Coordinates

H.1 Basis vectors

The usual basis vectors are denoted by $\mathbf{i}, \mathbf{j}, \mathbf{k}$ and are as the following picture describes.

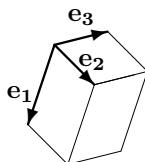


The vectors, $\mathbf{i}, \mathbf{j}, \mathbf{k}$, are fixed. If $\mathbf{v}$ is a vector, there are unique scalars called components such that $\mathbf{v} = v^1 \mathbf{i} + v^2 \mathbf{j} + v^3 \mathbf{k}$. This is what it means to say $\mathbf{i}, \mathbf{j}, \mathbf{k}$ is a basis.

Now suppose $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are three vectors which satisfy

$$\mathbf{e}_1 \times \mathbf{e}_2 \cdot \mathbf{e}_3 \neq 0.$$

Recall this means the volume of the box spanned by the three vectors is not zero.



Suppose $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are as just described. Does it follow that they form a basis? That is, do there exist unique scalars, v^1, v^2 , and v^3 such that

$$\mathbf{v} = v^i \mathbf{e}_i.$$

This is the content of the following theorem.

Theorem H.1.1 *If $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are three vectors, then they form a basis if and only if*

$$\mathbf{e}_1 \times \mathbf{e}_2 \cdot \mathbf{e}_3 \neq 0.$$

Proof: Suppose first the above condition holds. Let $\mathbf{i}_1 \equiv \mathbf{i}, \mathbf{i}_2 \equiv \mathbf{j}, \mathbf{i}_3 \equiv \mathbf{k}$ and suppose $\mathbf{v} = u^j \mathbf{i}_j$. Therefore, u^1, u^2, u^3 are the components of the vector $\mathbf{v}$ with respect to the usual basis vectors. Also let

$$\mathbf{e}_i = a_i^j \mathbf{i}_j$$

thus writing each $\mathbf{e}_i$ in terms of the vectors $\mathbf{i}_j$. Then from the definition of the box product in terms of the usual basis vectors,

$$0 \neq \mathbf{e}_1 \times \mathbf{e}_2 \cdot \mathbf{e}_3 = \det \left(a_i^j \right)$$

$$\equiv \det \begin{pmatrix} a_1^1 & a_1^2 & a_1^3 \\ a_2^1 & a_2^2 & a_2^3 \\ a_3^1 & a_3^2 & a_3^3 \end{pmatrix} = \det \begin{pmatrix} a_1^1 & a_2^1 & a_3^1 \\ a_1^2 & a_2^2 & a_3^2 \\ a_1^3 & a_2^3 & a_3^3 \end{pmatrix} \quad (8.1)$$

Does there exist a unique solution to

$$\mathbf{v} \equiv u^j \mathbf{i}_j = v^i \mathbf{e}_i = a_i^j v^i \mathbf{i}_j? \quad (8.2)$$

In other words, is there a unique solution to the system of equations

$$u^j = a_i^j v^i, \quad j = 1, 2, 3.$$

This is in the form

$$\mathbf{u} = A\mathbf{v} \quad (8.3)$$

where A is a matrix which, by 8.1, has non zero determinant. Therefore, there exists a unique solution.

If the box product is equal to zero, then the system which needs to be solved is of the form 8.3 where $\det(A) = 0$. Therefore, A has an eigenvector, $\mathbf{v} = (v_1, v_2, v_3)^T$ which corresponds to the eigen value $\lambda = 0$. Therefore, 8.2 shows $\mathbf{0}$ has more than one set of components with respect to the vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and this set of vectors is therefore not a basis.

This gives a simple geometric condition which determines whether a list of three vectors forms a basis in $\mathbb{R}^3$. One simply takes the box product. If the box product is not equal to zero, then the vectors form a basis. If not, the list of three vectors does not form a basis. This condition generalizes to $\mathbb{R}^n$ as follows. If $\mathbf{e}_i = a_i^j \mathbf{i}_j$, then $\{\mathbf{e}_i\}_{i=1}^n$ forms a basis if and only if $\det(a_i^j) \neq 0$.

These vectors $\{\mathbf{e}_i\}$ may or may not be orthonormal. In any case, it is convenient to define something called the dual basis.

Definition H.1.2 Let $\{\mathbf{e}_i\}_{i=1}^n$ form a basis for $\mathbb{R}^n$. Then $\{\mathbf{e}^i\}_{i=1}^n$ is called the dual basis if

$$\mathbf{e}^i \cdot \mathbf{e}_j = \delta_j^i \equiv \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (8.4)$$

Theorem H.1.3 If $\{\mathbf{e}_i\}_{i=1}^n$ is a basis then $\{\mathbf{e}^i\}_{i=1}^n$ is also a basis provided 8.4 holds.

Proof: Suppose

$$\mathbf{v} = v_i \mathbf{e}^i. \quad (8.5)$$

Then taking the dot product of both sides of 8.5 with $\mathbf{e}_j$, yields

$$v_j = \mathbf{v} \cdot \mathbf{e}_j. \quad (8.6)$$

Thus there is at most one choice of scalars v_j such that $\mathbf{v} = v_j \mathbf{e}^j$ and it is given by 8.6.

$$(\mathbf{v} - \mathbf{v} \cdot \mathbf{e}_j \mathbf{e}^j) \cdot \mathbf{e}_k = 0$$

and so, since $\{\mathbf{e}_i\}_{i=1}^n$ is a basis,

$$(\mathbf{v} - \mathbf{v} \cdot \mathbf{e}_j \mathbf{e}^j) \cdot \mathbf{w} = 0$$

for all vectors, $\mathbf{w}$. It follows $\mathbf{v} - \mathbf{v} \cdot \mathbf{e}_j \mathbf{e}^j = \mathbf{0}$ and this shows $\{\mathbf{e}^i\}_{i=1}^n$ is a basis.

The above gives formulas for the components of a vector $\mathbf{v}$, v_i , with respect to the dual basis, found to be $v_j = \mathbf{v} \cdot \mathbf{e}_j$. In the same way, the components of a vector with respect to the basis $\{\mathbf{e}_i\}_{i=1}^n$ can be obtained. Let $\mathbf{v}$ be any vector and let

$$\mathbf{v} = v^j \mathbf{e}_j. \quad (8.7)$$

Then using 8.4 and taking the dot product of both sides of 8.7 with $\mathbf{e}^i$, $v^i = \mathbf{e}^i \cdot \mathbf{v}$.

Does there exist a dual basis and is it uniquely determined?

Theorem H.1.4 *If $\{\mathbf{e}_i\}_{i=1}^n$ is a basis for $\mathbb{R}^n$, then there exists a unique dual basis, $\{\mathbf{e}^j\}_{j=1}^n$ satisfying*

$$\mathbf{e}^j \cdot \mathbf{e}_i = \delta_i^j.$$

Proof: First the dual basis is unique. Suppose $\{\mathbf{f}^j\}_{j=1}^n$ is another set of vectors which satisfies $\mathbf{f}^j \cdot \mathbf{e}_i = \delta_i^j$. Then

$$\mathbf{f}^j = \mathbf{f}^j \cdot \mathbf{e}_i \mathbf{e}^i = \delta_i^j \mathbf{e}^i = \mathbf{e}^j.$$

This shows uniqueness.

What about existence? Note that from the definition, the dual basis to $\{\mathbf{i}_j\}_{j=1}^n$ is just $\mathbf{i}^j = \mathbf{i}_j$. Letting

$$\mathbf{e}_i = a_i^j \mathbf{i}_j$$

where the vectors, $\{\mathbf{i}_j\}_{j=1}^n$ are the standard basis vectors, it follows that since the $\mathbf{e}_i$ form a basis that the matrix whose ij^{th} entry is a_i^j is an invertible matrix. Letting

$$\mathbf{e}^k = b_r^k \mathbf{i}^r,$$

b_r^k must be chosen such that

$$\overbrace{b_r^k \mathbf{i}^r}^{\mathbf{e}^k} \cdot a_i^j \mathbf{i}_j = b_r^k a_i^j \delta_j^r = b_j^k a_i^j = \delta_i^k.$$

But this is nothing more than the matrix equation for $B \equiv (b_r^k)$ which is of the form

$$AB = I$$

where $A = (a_i^j)$ and has an inverse. There exists a unique solution to this equation given by $B = A^{-1}$ and this proves the existence of the dual basis.

Summarizing what has been shown so far, $\{\mathbf{e}_i\}_{i=1}^n$ is a basis for $\mathbb{R}^n$ if and only if when $\mathbf{e}_i = a_i^j \mathbf{i}_j$,

$$\det(a_i^j) \neq 0. \quad (8.8)$$

If $\{\mathbf{e}_i\}_{i=1}^n$ is a basis, then there exists a unique dual basis, $\{\mathbf{e}^j\}_{j=1}^n$ satisfying

$$\mathbf{e}^j \cdot \mathbf{e}_i = \delta_i^j, \quad (8.9)$$

and that if $\mathbf{v}$ is any vector,

$$\mathbf{v} = v_j \mathbf{e}^j, \quad \mathbf{v} = v^j \mathbf{e}_j. \quad (8.10)$$

The components of $\mathbf{v}$ which have the index on the top are called the contravariant components of the vector while the components which have the index on the bottom are called the covariant components. In general $v_i \neq v^i$! The formulas for these components in terms of the dot product are

$$v_j = \mathbf{v} \cdot \mathbf{e}_j, \quad v^j = \mathbf{v} \cdot \mathbf{e}^j. \quad (8.11)$$

Define $g_{ij} \equiv \mathbf{e}_i \cdot \mathbf{e}_j$ and $g^{ij} \equiv \mathbf{e}^i \cdot \mathbf{e}^j$. The next theorem describes the process of raising or lowering an index.

Theorem H.1.5 *The following hold.*

$$g^{ij} \mathbf{e}_j = \mathbf{e}^i, \quad g_{ij} \mathbf{e}^j = \mathbf{e}_i, \quad (8.12)$$

$$g^{ij} v_j = v^i, \quad g_{ij} v^j = v_i, \quad (8.13)$$

$$g^{ij} g_{jk} = \delta_k^i, \quad (8.14)$$

$$\det(g_{ij}) > 0, \quad \det(g^{ij}) > 0. \quad (8.15)$$

Proof: First,

$$\mathbf{e}^i = \mathbf{e}^i \cdot \mathbf{e}^j \mathbf{e}_j = g^{ij} \mathbf{e}_j$$

by 8.10 and 8.11. Similarly, by 8.10 and 8.11,

$$\mathbf{e}_i = \mathbf{e}_i \cdot \mathbf{e}_j \mathbf{e}^j = g_{ij} \mathbf{e}^j.$$

This verifies 8.12. To verify 8.13,

$$v^i = \mathbf{e}^i \cdot \mathbf{v} = g^{ij} \mathbf{e}_j \cdot \mathbf{v} = g^{ij} v_j.$$

The proof of the remaining formula in 8.13 is similar.

To verify 8.14,

$$g^{ij} g_{jk} = \mathbf{e}^i \cdot \mathbf{e}^j \mathbf{e}_j \cdot \mathbf{e}_k = ((\mathbf{e}^i \cdot \mathbf{e}^j) \mathbf{e}_j) \cdot \mathbf{e}_k = \mathbf{e}^i \cdot \mathbf{e}_k = \delta_k^i.$$

This shows the two determinants in 8.15 are non zero because the two matrices are inverses of each other. It only remains to verify that one of these is greater than zero. Letting $\mathbf{e}_i = a_i^j \mathbf{i}_j = b_i^j \mathbf{j}^j$, it follows since $\mathbf{i}_j = \mathbf{j}^j$, $a_i^j = b_i^j$. Therefore,

$$\mathbf{e}_i \cdot \mathbf{e}_j = a_i^r \mathbf{i}_r \cdot b_j^k \mathbf{i}^k = a_i^r b_j^k \delta_r^k = a_i^k b_j^k = a_i^k a_j^k.$$

It follows that for G the matrix whose ij^{th} entry is $\mathbf{e}_i \cdot \mathbf{e}_j$, $G = AA^T$ where the ik^{th} entry of A is a_i^k . Therefore, $\det(G) = \det(A) \det(A^T) = \det(A)^2 > 0$. It follows from 8.14 that if H is the matrix whose ij^{th} entry is g^{ij} , then $GH = I$ and so $H = G^{-1}$ and

$$\det(G) \det(G^{-1}) = \det(g^{ij}) \det(G) = 1.$$

Therefore, $\det(G^{-1}) > 0$ also. This proves the theorem.

Definition H.1.6 The matrix $(g_{ij}) = G$ is called the metric tensor.

H.2 Exercises

1. Let $\mathbf{e}_1 = \mathbf{i} + \mathbf{j}$, $\mathbf{e}_2 = \mathbf{i} - \mathbf{j}$, $\mathbf{e}_3 = \mathbf{j} + \mathbf{k}$. Find $\mathbf{e}^1, \mathbf{e}^2, \mathbf{e}^3$, (g_{ij}) , (g^{ij}) . If $\mathbf{v} = \mathbf{i} + 2\mathbf{j} + \mathbf{k}$, find v^i and v_j , the contravariant and covariant components of the vector.
2. Let $\mathbf{e}^1 = 2\mathbf{i} + \mathbf{j}$, $\mathbf{e}^2 = \mathbf{i} - 2\mathbf{j}$, $\mathbf{e}^3 = \mathbf{k}$. Find $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, (g_{ij}) , (g^{ij}) . If $\mathbf{v} = 2\mathbf{i} - 2\mathbf{j} + \mathbf{k}$, find v^i and v_j , the contravariant and covariant components of the vector.
3. Suppose $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ have the property that $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ whenever $i \neq j$. Show that the same is true of the dual basis and that in fact, $\mathbf{e}^i$ is a multiple of $\mathbf{e}_i$.
4. Let $\mathbf{e}_1, \dots, \mathbf{e}_3$ be a basis for $\mathbb{R}^3$ and let $\mathbf{v} = v^i \mathbf{e}_i = v_i \mathbf{e}^i$, $\mathbf{w} = w^j \mathbf{e}_j = w_j \mathbf{e}^j$ be two vectors. Show that

$$\mathbf{v} \cdot \mathbf{w} = g_{ij} v^i w^j = g^{ij} v_i w_j.$$

5. Show that if $\{\mathbf{e}_i\}_{i=1}^3$ is a basis in $\mathbb{R}^3$

$$\mathbf{e}^1 = \frac{\mathbf{e}_2 \times \mathbf{e}_3}{\mathbf{e}_2 \times \mathbf{e}_3 \cdot \mathbf{e}_1}, \quad \mathbf{e}^2 = \frac{\mathbf{e}_1 \times \mathbf{e}_2}{\mathbf{e}_1 \times \mathbf{e}_3 \cdot \mathbf{e}_2}, \quad \mathbf{e}^3 = \frac{\mathbf{e}_1 \times \mathbf{e}_2}{\mathbf{e}_1 \times \mathbf{e}_2 \cdot \mathbf{e}_3}.$$

6. Let $\{\mathbf{e}_i\}_{i=1}^n$ be a basis and define

$$\mathbf{e}_i^* \equiv \frac{\mathbf{e}_i}{|\mathbf{e}_i|}, \quad \mathbf{e}^{*i} \equiv \mathbf{e}^i |\mathbf{e}_i|.$$

Show that $\mathbf{e}^{*i} \cdot \mathbf{e}_j^* = \delta_j^i$.

7. If $\mathbf{v}$ is a vector, v_i^* and v^{*i} , are defined by

$$\mathbf{v} \equiv v_i^* \mathbf{e}^{*i} \equiv v^{*i} \mathbf{e}_i^*.$$

These are called the physical components of $\mathbf{v}$. Show that

$$v_i^* = \frac{v_i}{|\mathbf{e}_i|}, \quad v^{*i} = v^i |\mathbf{e}_i| \quad (\text{No summation on } i).$$

H.3 Curvilinear Coordinates

With the algebraic preparation of the last section, consider curvilinear coordinates. Let $D \subseteq \mathbb{R}^n$ be an open set and let $\mathbf{M}: D \rightarrow \mathbb{R}^n$ satisfy

$$\mathbf{M} \text{ is } C^2, \quad (8.16)$$

$$\mathbf{M} \text{ is one to one.} \quad (8.17)$$

Letting $\mathbf{x} \in D$,

$$\mathbf{M}(\mathbf{x}) = M^k(\mathbf{x}) \mathbf{i}_k$$

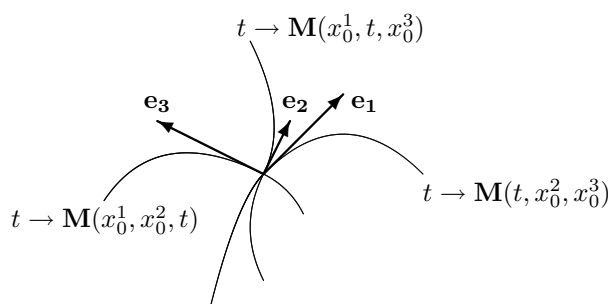
where, as usual, $\mathbf{i}_k$ are the standard basis vectors for $\mathbb{R}^n$, $\mathbf{i}_k$ being the vector in $\mathbb{R}^n$ which has a one in the k^{th} coordinate and a 0 in every other spot. For a fixed $\mathbf{x} \in D$, consider the curves,

$$t \rightarrow \mathbf{M}(\mathbf{x} + t\mathbf{i}_k)$$

for $t \in I$, some open interval containing 0. Then for the point $\mathbf{x}$, let

$$\mathbf{e}_k \equiv \frac{\partial \mathbf{M}}{\partial x^k}(\mathbf{x}) \equiv \frac{d}{dt}(\mathbf{M}(\mathbf{x} + t\mathbf{i}_k))|_{t=0}.$$

Denote this vector as $\mathbf{e}_k(\mathbf{x})$ to emphasize its dependence on $\mathbf{x}$. The following picture illustrates the situation in $\mathbb{R}^3$.



Is $\{\mathbf{e}_k\}_{k=1}^n$ a basis? This will be so if

$$\det \left(\frac{\partial M^i}{\partial x^k} \right) \neq 0. \quad (8.18)$$

Let

$$y^i = M^i(\mathbf{x}) \quad i = 1, \dots, n \quad (8.19)$$

so that the y^i are the usual coordinates with respect to the usual basis vectors $\{\mathbf{i}_k\}_{k=1}^n$ of the point $\mathbf{M}(\mathbf{x})$. Letting $\mathbf{x} \equiv (x^1, \dots, x^n)$, it follows from the inverse function theorem of advanced calculus that $\mathbf{M}(D)$ is open, and that 8.18, 8.16, and 8.17 imply

the equations 8.19 define each x^i as a C^2 function of $\mathbf{y} \equiv (y^1, \dots, y^n)^T$. Thus, abusing notation slightly, the equations 8.19 are equivalent to

$$x^i = x^i(\mathbf{y}), \quad i = 1, \dots, n$$

where x^i is a C^2 function. Thus

$$\nabla x^k(\mathbf{y}) = \frac{\partial x^k(\mathbf{y})}{\partial y^j} \mathbf{i}^j.$$

Then

$$\nabla x^k(\mathbf{y}) \cdot \mathbf{e}_j = \frac{\partial x^k}{\partial y^s} \mathbf{i}^s \cdot \frac{\partial y^r}{\partial x^j} \mathbf{i}_r = \frac{\partial x^k}{\partial y^s} \frac{\partial y^s}{\partial x^j} = \delta_j^k$$

by the chain rule. Therefore, the dual basis is given by

$$\mathbf{e}^k(\mathbf{x}) = \nabla x^k(\mathbf{y}). \quad (8.20)$$

Notice that it might be hard or even impossible to solve algebraically for x^i in terms of the y^j . Thus the straight forward approach to finding $\mathbf{e}^k$ by 8.20 might be impossible. Also, this approach leads to an expression in terms of the $\mathbf{y}$ coordinates rather than the desired $\mathbf{x}$ coordinates. Therefore, it is expedient to use another method to obtain these vectors. The vectors, $\mathbf{e}^k(\mathbf{x})$ may always be found by using formula 8.12 and the result is in terms of the curvilinear coordinates, $\mathbf{x}$. Consider the following example.

Example H.3.1 $D \equiv (0, \infty) \times (0, \pi) \times (0, 2\pi)$ and

$$\begin{pmatrix} y^1 \\ y^2 \\ y^3 \end{pmatrix} = \begin{pmatrix} x^1 \sin(x^2) \cos(x^3) \\ x^1 \sin(x^2) \sin(x^3) \\ x^1 \cos(x^2) \end{pmatrix}$$

(These are spherical coordinates, usually written as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \rho \sin(\phi) \cos(\theta) \\ \rho \sin(\phi) \sin(\theta) \\ \rho \cos(\phi) \end{pmatrix}$$

where (ρ, ϕ, θ) are the spherical coordinates.) Thus

$$\mathbf{e}_1(\mathbf{x}) = \sin(x^2) \cos(x^3) \mathbf{i}_1 + \sin(x^2) \sin(x^3) \mathbf{i}_2 + \cos(x^2) \mathbf{i}_3,$$

$$\begin{aligned} \mathbf{e}_2(\mathbf{x}) &= x^1 \cos(x^2) \cos(x^3) \mathbf{i}_1 \\ &+ x^1 \cos(x^2) \sin(x^3) \mathbf{i}_2 - x^1 \sin(x^2) \mathbf{i}_3, \end{aligned}$$

$$\mathbf{e}_3(\mathbf{x}) = -x^1 \sin(x^2) \sin(x^3) \mathbf{i}_1 + x^1 \sin(x^2) \cos(x^3) \mathbf{i}_2 + 0 \mathbf{i}_3.$$

It follows the metric tensor is

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (x^1)^2 & 0 \\ 0 & 0 & (x^1)^2 \sin^2(x^2) \end{pmatrix} = (g_{ij}) = (\mathbf{e}_i \cdot \mathbf{e}_j). \quad (8.21)$$

Therefore, by Theorem H.1.5

$$\begin{aligned} G^{-1} &= (g^{ij}) \\ &= (\mathbf{e}^i, \mathbf{e}^j) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & (x^1)^{-2} & 0 \\ 0 & 0 & (x^1)^{-2} \sin^{-2}(x^2) \end{pmatrix}. \end{aligned}$$

To obtain the dual basis, use Theorem H.1.5 to write

$$\begin{aligned}\mathbf{e}^1 &= g^{1j} \mathbf{e}_j = \mathbf{e}_1 \\ \mathbf{e}^2 &= g^{2j} \mathbf{e}_j = (x^1)^{-2} \mathbf{e}_2 \\ \mathbf{e}^3 &= g^{3j} \mathbf{e}_j = (x^1)^{-2} \sin^{-2}(x^2) \mathbf{e}_3.\end{aligned}$$

It is natural to ask if there exists a transformation $\mathbf{M}$ such that

$$\frac{\partial \mathbf{M}}{\partial x^1} = \mathbf{i} = \mathbf{i}_1, \frac{\partial \mathbf{M}}{\partial x^2} = \mathbf{j} = \mathbf{i}_2, \frac{\partial \mathbf{M}}{\partial x^3} = \mathbf{k} = \mathbf{i}_3. \quad (8.22)$$

Let

$$\mathbf{M}(x^1, x^2, x^3) \equiv x^1 \mathbf{i} + x^2 \mathbf{j} + x^3 \mathbf{k}.$$

Then 8.22 holds for this transformation.

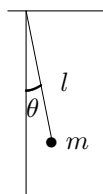
H.4 Exercises

1. Let

$$\begin{pmatrix} y^1 \\ y^2 \\ y^3 \end{pmatrix} = \begin{pmatrix} x^1 + 2x^2 \\ x^2 + x^3 \\ x^1 - 2x^2 \end{pmatrix}$$

where the y^i are the rectangular coordinates of the point. Find $\mathbf{e}^i, \mathbf{e}_i, i = 1, 2, 3$, and find $(g_{ij})(\mathbf{x})$ and $(g^{ij})(\mathbf{x})$.

2. Let $\mathbf{y} = \mathbf{y}(\mathbf{x}, t)$ where t signifies time and $\mathbf{x} \in U \subseteq \mathbb{R}^m$ for U an open set, while $\mathbf{y} \in \mathbb{R}^n$ and suppose $\mathbf{x}$ is a function of t . Physically, this corresponds to an object moving over a surface in $\mathbb{R}^n$ which may be changing as a function of t . The point $\mathbf{y} = \mathbf{y}(\mathbf{x}(t), t)$ is the point in $\mathbb{R}^n$ corresponding to t . For example, consider the pendulum



in which $n = 2, l$ is fixed and $y^1 = l \sin \theta, y^2 = l - l \cos \theta$. Thus, in this simple example, $m = 1$. If l were changing in a known way with respect to t , then this would be of the form $\mathbf{y} = \mathbf{y}(\mathbf{x}, t)$. The kinetic energy is defined as

$$T \equiv \frac{1}{2} m \dot{\mathbf{y}} \cdot \dot{\mathbf{y}} \quad (*)$$

where the dot on the top signifies differentiation with respect to t . Show that

$$\frac{\partial T}{\partial \dot{x}^k} = m \dot{\mathbf{y}} \cdot \frac{\partial \mathbf{y}}{\partial x^k}.$$

Hint: First show that

$$\dot{\mathbf{y}} = \frac{\partial \mathbf{y}}{\partial x^j} \dot{x}^j + \frac{\partial \mathbf{y}}{\partial t} \quad (**)$$

and so

$$\frac{\partial \dot{\mathbf{y}}}{\partial \dot{x}^j} = \frac{\partial \mathbf{y}}{\partial x^j}.$$

3. ↑ Show that

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{x}^k} \right) = m \dot{\mathbf{y}} \cdot \frac{\partial \mathbf{y}}{\partial x^k} + m \dot{\mathbf{y}} \cdot \frac{\partial^2 \mathbf{y}}{\partial x^k \partial x^r} \dot{x}^r + m \dot{\mathbf{y}} \cdot \frac{\partial^2 \mathbf{y}}{\partial t \partial x^k}.$$

4. ↑ Show that

$$\frac{\partial T}{\partial x^k} = m \dot{\mathbf{y}} \cdot \left(\frac{\partial^2 \mathbf{y}}{\partial x^r \partial x^k} \dot{x}^r + \frac{\partial^2 \mathbf{y}}{\partial t \partial x^k} \right).$$

Hint: Use * and **.

5. ↑ Now show from Newton's second law (mass times acceleration equals force) that for $\mathbf{F}$ the force,

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{x}^k} \right) - \frac{\partial T}{\partial x^k} = m \ddot{\mathbf{y}} \cdot \frac{\partial \mathbf{y}}{\partial x^k} = \mathbf{F} \cdot \frac{\partial \mathbf{y}}{\partial x^k}. \quad (***)$$

6. ↑ In the example of the simple pendulum above,

$$\mathbf{y} = \begin{pmatrix} l \sin \theta \\ l - l \cos \theta \end{pmatrix} = l \sin \theta \mathbf{i} + (l - l \cos \theta) \mathbf{j}.$$

Use *** to find a differential equation which describes the vibrations of the pendulum in terms of θ . First write the kinetic energy and then consider the force acting on the mass which is

$$-mg\mathbf{j}.$$

7. The above problem is fairly easy to do without the formalism developed. Now consider the case where $\mathbf{x} = (\rho, \theta, \phi)$, spherical coordinates, and write differential equations for ρ , θ , and ϕ to describe the motion of an object in terms of these coordinates given a force, $\mathbf{F}$.
8. Suppose the pendulum is not assumed to vibrate in a plane. Let it be suspended at the origin and consider spherical coordinates. Find differential equations for θ and ϕ .
9. If there are many masses, $m_\alpha, \alpha = 1, \dots, R$, the kinetic energy is the sum of the kinetic energies of the individual masses. Thus,

$$T \equiv \frac{1}{2} \sum_{\alpha=1}^R m_\alpha |\dot{\mathbf{y}}_\alpha|^2.$$

Generalize the above problems to show that, assuming

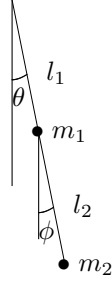
$$\mathbf{y}_\alpha = \mathbf{y}_\alpha(\mathbf{x}, t),$$

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{x}^k} \right) - \frac{\partial T}{\partial x^k} = \sum_{\alpha=1}^R \mathbf{F}_\alpha \cdot \frac{\partial \mathbf{y}_\alpha}{\partial x^k}$$

where $\mathbf{F}_\alpha$ is the force acting on m_α .

10. Discuss the equivalence of these formulae with Newton's second law, force equals mass times acceleration. What is gained from the above so called Lagrangian formalism?

11. The double pendulum has two masses instead of only one.

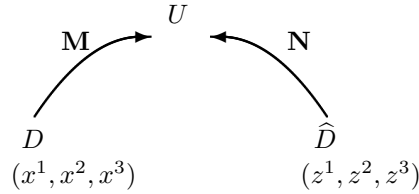


Write differential equations for θ and ϕ to describe the motion of the double pendulum.

H.5 Transformation Of Coordinates.

The scalars $\{x^i\}$ are called curvilinear coordinates. Note they can be used to identify a point in $\mathbb{R}^n$ and $\mathbf{x} = (x^1, \dots, x^n)$ is a point in $\mathbb{R}^n$. The basis vectors associated with this particular set of curvilinear coordinates at a point identified by $\mathbf{x}$ are denoted by $\mathbf{e}_i(\mathbf{x})$ and the dual basis vectors at this point are denoted by $\mathbf{e}^j(\mathbf{x})$. What if other curvilinear coordinates are used? How do you write $\mathbf{e}^k(\mathbf{x})$ in terms of the vectors, $\mathbf{e}^j(\mathbf{z})$ where $\mathbf{z}$ is some other type of curvilinear coordinates? This is next.

Consider the following picture in which U is an open set in $\mathbb{R}^n$, D , and $\hat{D}$ are open sets in $\mathbb{R}^n$, and $\mathbf{M}, \mathbf{N}$ are C^2 mappings which are one to one from D and $\hat{D}$ respectively. Suppose that a point in U is identified by the curvilinear coordinates $\mathbf{x}$ in D and $\mathbf{z}$ in $\hat{D}$.



Thus $\mathbf{M}(\mathbf{x}) = \mathbf{N}(\mathbf{z})$. Now by the chain rule,

$$\mathbf{e}_i(\mathbf{z}) \equiv \frac{\partial \mathbf{N}}{\partial z^i} = \frac{\partial \mathbf{M}}{\partial x^j} \frac{\partial x^j}{\partial z^i} = \frac{\partial x^j}{\partial z^i} \mathbf{e}_j(\mathbf{x}). \quad (8.23)$$

Define the covariant and contravariant coordinates for the various curvilinear coordinates in the obvious way. Thus,

$$\mathbf{v} = v_i(\mathbf{x}) \mathbf{e}^i(\mathbf{x}) = v^i(\mathbf{x}) \mathbf{e}_i(\mathbf{x}) = v_j(\mathbf{z}) \mathbf{e}^j(\mathbf{z}) = v^j(\mathbf{z}) \mathbf{e}_j(\mathbf{z}).$$

Then here is a fundamental theorem about transforming the vectors and coordinates.

Theorem H.5.1 *The following transformation rules hold for pairs of curvilinear coordinates.*

$$v_i(\mathbf{z}) = \frac{\partial x^j}{\partial z^i} v_j(\mathbf{x}), \quad v^i(\mathbf{z}) = \frac{\partial z^i}{\partial x^j} v^j(\mathbf{x}), \quad (8.24)$$

$$\mathbf{e}_i(\mathbf{z}) = \frac{\partial x^j}{\partial z^i} \mathbf{e}_j(\mathbf{x}), \quad \mathbf{e}^i(\mathbf{z}) = \frac{\partial z^i}{\partial x^j} \mathbf{e}^j(\mathbf{x}), \quad (8.25)$$

$$g_{ij}(\mathbf{z}) = \frac{\partial x^r}{\partial z^i} \frac{\partial x^s}{\partial z^j} g_{rs}(\mathbf{x}), \quad g^{ij}(\mathbf{z}) = \frac{\partial z^i}{\partial x^r} \frac{\partial z^j}{\partial x^s} g^{rs}(\mathbf{x}). \quad (8.26)$$

Proof: The first part of 8.25 is shown in 8.23. Then, from 8.23,

$$\begin{aligned}\mathbf{e}^i(\mathbf{z}) &= \mathbf{e}^i(\mathbf{z}) \cdot \mathbf{e}_j(\mathbf{x}) \mathbf{e}^j(\mathbf{x}) = \mathbf{e}^i(\mathbf{z}) \cdot \frac{\partial z^k}{\partial x^j} \mathbf{e}_k(\mathbf{z}) \mathbf{e}^j(\mathbf{x}) \\ &= \delta_k^i \frac{\partial z^k}{\partial x^j} \mathbf{e}^j(\mathbf{x}) = \frac{\partial z^i}{\partial x^j} \mathbf{e}^j(\mathbf{x})\end{aligned}$$

and this proves the second part of 8.25. Now to show that 8.24,

$$v_i(\mathbf{z}) = \mathbf{v} \cdot \mathbf{e}_i(\mathbf{z}) = \mathbf{v} \cdot \frac{\partial x^j}{\partial z_i} \mathbf{e}_j(\mathbf{x}) = \frac{\partial x^j}{\partial z_i} v_j(\mathbf{x})$$

and

$$v^i(\mathbf{z}) = \mathbf{v} \cdot \mathbf{e}^i(\mathbf{z}) = \mathbf{v} \cdot \frac{\partial z^i}{\partial x^j} \mathbf{e}^j(\mathbf{x}) = \frac{\partial z^i}{\partial x^j} v^j(\mathbf{x}).$$

To verify 8.26,

$$g_{ij}(\mathbf{z}) = \mathbf{e}_i(\mathbf{z}) \cdot \mathbf{e}_j(\mathbf{z}) = \mathbf{e}_r(\mathbf{x}) \frac{\partial x^r}{\partial z^i} \cdot \mathbf{e}_s(\mathbf{x}) \frac{\partial x^s}{\partial z^j} = g_{rs}(\mathbf{x}) \frac{\partial x^r}{\partial z^i} \frac{\partial x^s}{\partial z^j}.$$

This proves the theorem.

Denote by $\mathbf{y}$ the curvilinear coordinates with the property that

$$\mathbf{e}^k(\mathbf{y}) = \mathbf{i}_k = \mathbf{e}_k(\mathbf{y}).$$

H.6 Differentiation and Christoffel Symbols

Let $\mathbf{F}: U \rightarrow \mathbb{R}^n$ be differentiable. Thus $\mathbf{F}$ is a vector field and it is used to model force, velocity, acceleration, or any other vector quantity which may change from point to point in U . Then

$$\frac{\partial \mathbf{F}(\mathbf{x})}{\partial x^j}$$

is a vector and so there exist scalars, $F_{,j}^i(\mathbf{x})$ and $F_{i,j}(\mathbf{x})$ such that

$$\frac{\partial \mathbf{F}(\mathbf{x})}{\partial x^j} = F_{,j}^i(\mathbf{x}) \mathbf{e}_i(\mathbf{x}) = F_{i,j}(\mathbf{x}) \mathbf{e}^i(\mathbf{x}). \quad (8.27)$$

Consider how these scalars, $F_{,j}^i(\mathbf{x})$ and $F_{i,j}(\mathbf{x})$ transform when the coordinates are changed.

Theorem H.6.1 *If $\mathbf{x}$ and $\mathbf{z}$ are curvilinear coordinates,*

$$F_{,s}^r(\mathbf{x}) = F_{,j}^i(\mathbf{z}) \frac{\partial x^r}{\partial z^i} \frac{\partial z^j}{\partial x^s}, \quad F_{r,s}(\mathbf{x}) \frac{\partial x^r}{\partial z^i} \frac{\partial x^s}{\partial z^j} = F_{i,j}(\mathbf{z}). \quad (8.28)$$

Proof:

$$\begin{aligned}F_{,s}^r(\mathbf{x}) \mathbf{e}_r(\mathbf{x}) &\equiv \frac{\partial \mathbf{F}(\mathbf{x})}{\partial x^s} = \frac{\partial \mathbf{F}(\mathbf{z})}{\partial z^j} \frac{\partial z^j}{\partial x^s} \equiv \\ &F_{,j}^i(\mathbf{z}) \mathbf{e}_i(\mathbf{z}) \frac{\partial z^j}{\partial x^s} = F_{,j}^i(\mathbf{z}) \frac{\partial x^r}{\partial z^i} \frac{\partial z^j}{\partial x^s} \mathbf{e}_r(\mathbf{x})\end{aligned}$$

which shows the first formula of 8.27. To show the other formula,

$$\begin{aligned}F_{i,j}(\mathbf{z}) \mathbf{e}^i(\mathbf{z}) &\equiv \frac{\partial \mathbf{F}(\mathbf{z})}{\partial z^j} = \frac{\partial \mathbf{F}(\mathbf{x})}{\partial x^s} \frac{\partial x^s}{\partial z^j} \equiv \\ &F_{r,s}(\mathbf{x}) \mathbf{e}^r(\mathbf{x}) \frac{\partial x^s}{\partial z^j} = F_{r,s}(\mathbf{x}) \frac{\partial x^r}{\partial z^i} \frac{\partial x^s}{\partial z^j} \mathbf{e}^i(\mathbf{z}),\end{aligned}$$

and this shows the second formula for transforming these scalars. This proves the theorem.

Now $\mathbf{F}(\mathbf{x}) = F^i(\mathbf{x}) \mathbf{e}_i(\mathbf{x})$ and so by the product rule,

$$\frac{\partial \mathbf{F}}{\partial x^j} = \frac{\partial F^i}{\partial x^j} \mathbf{e}_i(\mathbf{x}) + F^i(\mathbf{x}) \frac{\partial \mathbf{e}_i(\mathbf{x})}{\partial x^j}.$$

Now $\frac{\partial \mathbf{e}_i(\mathbf{x})}{\partial x^j}$ is a vector and so there exist scalars, $\left\{ \begin{smallmatrix} k \\ ij \end{smallmatrix} \right\}$ such that

$$\frac{\partial \mathbf{e}_i(\mathbf{x})}{\partial x^j} = \left\{ \begin{smallmatrix} k \\ ij \end{smallmatrix} \right\} \mathbf{e}_k(\mathbf{x}).$$

Therefore,

$$\frac{\partial \mathbf{F}}{\partial x^j} = \frac{\partial F^k}{\partial x^j} \mathbf{e}_k(\mathbf{x}) + F^i(\mathbf{x}) \left\{ \begin{smallmatrix} k \\ ij \end{smallmatrix} \right\} \mathbf{e}_k(\mathbf{x})$$

which shows

$$F^k_{,j}(\mathbf{x}) = \frac{\partial F^k}{\partial x^j} + F^i(\mathbf{x}) \left\{ \begin{smallmatrix} k \\ ij \end{smallmatrix} \right\}.$$

This is sometimes called the covariant derivative.

These scalars are called the Christoffel symbols of the second kind. The next theorem is devoted to properties of these Christoffel symbols. Before stating the theorem, recall that the mapping, $\mathbf{M}$ which defines the curvilinear coordinates is C^2 . The reason for this is to be able to assert the mixed partial derivatives are equal.

Theorem H.6.2 *The Christoffel symbols of the second kind satisfy the following*

$$\frac{\partial \mathbf{e}_i(\mathbf{x})}{\partial x^j} = \left\{ \begin{smallmatrix} k \\ ij \end{smallmatrix} \right\} \mathbf{e}_k(\mathbf{x}), \quad (8.29)$$

$$\frac{\partial \mathbf{e}^i(\mathbf{x})}{\partial x^j} = -\left\{ \begin{smallmatrix} i \\ kj \end{smallmatrix} \right\} \mathbf{e}^k(\mathbf{x}), \quad (8.30)$$

$$\left\{ \begin{smallmatrix} k \\ ij \end{smallmatrix} \right\} = \left\{ \begin{smallmatrix} k \\ ji \end{smallmatrix} \right\}, \quad (8.31)$$

$$\left\{ \begin{smallmatrix} m \\ ik \end{smallmatrix} \right\} = \frac{g^{jm}}{2} \left[\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{kj}}{\partial x^i} - \frac{\partial g_{ik}}{\partial x^j} \right]. \quad (8.32)$$

Proof: Formula 8.29 is the definition of the Christoffel symbols. Consider 8.30 next. To do so, note

$$\mathbf{e}^i(\mathbf{x}) \cdot \mathbf{e}_k(\mathbf{x}) = \delta_k^i.$$

Then from the product rule,

$$\frac{\partial \mathbf{e}^i(\mathbf{x})}{\partial x^j} \cdot \mathbf{e}_k(\mathbf{x}) + \mathbf{e}^i(\mathbf{x}) \cdot \frac{\partial \mathbf{e}_k(\mathbf{x})}{\partial x^j} = 0.$$

Now from the definition,

$$\frac{\partial \mathbf{e}^i(\mathbf{x})}{\partial x^j} \cdot \mathbf{e}_k(\mathbf{x}) = -\mathbf{e}^i(\mathbf{x}) \cdot \left\{ \begin{smallmatrix} r \\ kj \end{smallmatrix} \right\} \mathbf{e}_r(\mathbf{x}) = -\left\{ \begin{smallmatrix} i \\ kj \end{smallmatrix} \right\}.$$

But also,

$$\frac{\partial \mathbf{e}^i(\mathbf{x})}{\partial x^j} = \frac{\partial \mathbf{e}^i(\mathbf{x})}{\partial x^j} \cdot \mathbf{e}_k(\mathbf{x}) \mathbf{e}^k(\mathbf{x}) = -\left\{ \begin{smallmatrix} i \\ kj \end{smallmatrix} \right\} \mathbf{e}^k(\mathbf{x}).$$

This verifies 8.30.

Letting $\frac{\partial \mathbf{M}(\mathbf{x})}{\partial x^j} = \mathbf{e}_j(\mathbf{x})$, it follows from equality of mixed partial derivatives,

$$\left\{ \begin{matrix} k \\ ij \end{matrix} \right\} \mathbf{e}_k(\mathbf{x}) = \frac{\partial \mathbf{e}_i}{\partial x^j} = \frac{\partial^2 \mathbf{M}}{\partial x^j \partial x^i} = \frac{\partial^2 \mathbf{M}}{\partial x^i \partial x^j} = \frac{\partial \mathbf{e}_j}{\partial x^i} = \left\{ \begin{matrix} k \\ ji \end{matrix} \right\} \mathbf{e}_k(\mathbf{x}),$$

which shows 8.31. It remains to show 8.32.

$$\frac{\partial g_{ij}}{\partial x^k} = \frac{\partial \mathbf{e}_i}{\partial x^k} \cdot \mathbf{e}_j + \mathbf{e}_i \cdot \frac{\partial \mathbf{e}_j}{\partial x^k} = \left\{ \begin{matrix} r \\ ik \end{matrix} \right\} \mathbf{e}_r \cdot \mathbf{e}_j + \mathbf{e}_i \cdot \mathbf{e}_r \left\{ \begin{matrix} r \\ jk \end{matrix} \right\}.$$

Therefore,

$$\frac{\partial g_{ij}}{\partial x^k} = \left\{ \begin{matrix} r \\ ik \end{matrix} \right\} g_{rj} + \left\{ \begin{matrix} r \\ jk \end{matrix} \right\} g_{ri}. \quad (8.33)$$

Switching i and k while remembering 8.31 yields

$$\frac{\partial g_{kj}}{\partial x^i} = \left\{ \begin{matrix} r \\ ik \end{matrix} \right\} g_{rj} + \left\{ \begin{matrix} r \\ ji \end{matrix} \right\} g_{rk}. \quad (8.34)$$

Now switching j and k in 8.33,

$$\frac{\partial g_{ik}}{\partial x^j} = \left\{ \begin{matrix} r \\ ij \end{matrix} \right\} g_{rk} + \left\{ \begin{matrix} r \\ jk \end{matrix} \right\} g_{ri}. \quad (8.35)$$

Adding 8.33 to 8.34 and subtracting 8.35 yields

$$\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{kj}}{\partial x^i} - \frac{\partial g_{ik}}{\partial x^j} = 2 \left\{ \begin{matrix} r \\ ik \end{matrix} \right\} g_{rj}.$$

Now multiplying both sides by g^{jm} and using the fact shown earlier in Theorem H.1.5 that

$$g_{rj} g^{jm} = \delta_r^m,$$

$$2 \left\{ \begin{matrix} m \\ ik \end{matrix} \right\} = g^{jm} \left(\frac{\partial g_{ij}}{\partial x^k} + \frac{\partial g_{kj}}{\partial x^i} - \frac{\partial g_{ik}}{\partial x^j} \right)$$

which proves 8.32.

This is a very interesting formula because it shows the Christoffel symbols are completely determined by the metric tensor and its derivatives.

H.7 Gradients And Divergence

What is the gradient and divergence of a vector field in general curvilinear coordinates? As before, $\mathbf{y}$ will denote the standard coordinates with respect to the usual basis vectors. Thus

$$\mathbf{N}(\mathbf{y}) \equiv y^k \mathbf{i}_k, \quad \mathbf{e}_k(\mathbf{y}) = \mathbf{i}_k = \mathbf{e}^k(\mathbf{y}).$$

Let $\phi : U \rightarrow \mathbb{R}$ be a differentiable scalar function, sometimes called a “scalar field” in this subject. Write $\phi(\mathbf{x})$ to denote the value of ϕ at the point whose coordinates are $\mathbf{x}$. In general, this convention is followed for any field, vector or scalar. Thus $\mathbf{F}(\mathbf{x})$ is the value of a vector field at the point of U determined by the coordinates $\mathbf{x}$. In the standard coordinates, the formula for the gradient of ϕ is known. It is given by the following formula.

$$\nabla \phi(\mathbf{y}) = \frac{\partial \phi(\mathbf{y})}{\partial y^k} \mathbf{e}^k(\mathbf{y}).$$

(Recall $\mathbf{e}^k(\mathbf{y}) = \mathbf{e}_k(\mathbf{y}) = \mathbf{i}_k$.) Therefore, using the chain rule, if the coordinates of the point of U are given in terms of the curvilinear coordinates $\mathbf{x}$,

$$\begin{aligned}\nabla\phi(\mathbf{x}) &= \nabla\phi(\mathbf{y}) \\ &= \frac{\partial\phi(\mathbf{x})}{\partial x^r} \frac{\partial x^r}{\partial y^k} \frac{\partial y^k}{\partial x^s} \mathbf{e}^s(\mathbf{x}) = \frac{\partial\phi(\mathbf{x})}{\partial x^r} \delta_s^r \mathbf{e}^s(\mathbf{x}) = \frac{\partial\phi(\mathbf{x})}{\partial x^r} \mathbf{e}^r(\mathbf{x}).\end{aligned}$$

This shows the covariant components of $\nabla\phi(\mathbf{x})$ are

$$(\nabla\phi(\mathbf{x}))_r = \frac{\partial\phi(\mathbf{x})}{\partial x^r}. \quad (8.36)$$

To find the contravariant components, “raise the index” in the usual way. Thus

$$(\nabla\phi(\mathbf{x}))^r = g^{rk}(\mathbf{x}) (\nabla\phi(\mathbf{x}))_k = g^{rk}(\mathbf{x}) \frac{\partial\phi(\mathbf{x})}{\partial x^k}. \quad (8.37)$$

What about the divergence of a vector field? The divergence of a vector field, $\mathbf{F}$ defined on U is a scalar field, $\text{div}(\mathbf{F})$ which in terms of the usual coordinates $\mathbf{y}$ is

$$\frac{\partial F^k}{\partial y^k}(\mathbf{y}) = F^k_{,k}(\mathbf{y})$$

in terms of the usual coordinates $\mathbf{y}$. The reason the above equation holds in this case is that $\mathbf{e}_k(\mathbf{y})$ is a constant and so the Christoffel symbols are zero. What is the correct expression for the divergence in an arbitrary coordinate system? From Theorem H.6.1,

$$\begin{aligned}F^i_{,j}(\mathbf{y}) &= F^r_{,s}(\mathbf{x}) \frac{\partial x^s}{\partial y^j} \frac{\partial y^i}{\partial x^r} \\ &= \left(\frac{\partial F^r}{\partial x^s}(\mathbf{x}) + F^k(\mathbf{x}) \left\{ \begin{matrix} r \\ ks \end{matrix} \right\}(\mathbf{x}) \right) \frac{\partial x^s}{\partial y^j} \frac{\partial y^i}{\partial x^r}.\end{aligned}$$

Letting $j = i$ yields

$$\begin{aligned}\text{div}(\mathbf{F}) &= \left(\frac{\partial F^r}{\partial x^s}(\mathbf{x}) + F^k(\mathbf{x}) \left\{ \begin{matrix} r \\ ks \end{matrix} \right\}(\mathbf{x}) \right) \frac{\partial x^s}{\partial y^i} \frac{\partial y^i}{\partial x^r} \\ &= \left(\frac{\partial F^r}{\partial x^s}(\mathbf{x}) + F^k(\mathbf{x}) \left\{ \begin{matrix} r \\ ks \end{matrix} \right\}(\mathbf{x}) \right) \delta_r^s \\ &= \left(\frac{\partial F^r}{\partial x^r}(\mathbf{x}) + F^k(\mathbf{x}) \left\{ \begin{matrix} r \\ kr \end{matrix} \right\}(\mathbf{x}) \right).\end{aligned} \quad (8.38)$$

The Christoffel symbols $\left\{ \begin{matrix} r \\ kr \end{matrix} \right\}$ are simplified using the description of it in Theorem H.6.2. Thus, from this theorem,

$$\left\{ \begin{matrix} r \\ kr \end{matrix} \right\} = \frac{g^{jr}}{2} \left[\frac{\partial g_{rj}}{\partial x^k} + \frac{\partial g_{kj}}{\partial x^r} - \frac{\partial g_{rk}}{\partial x^j} \right]$$

Now consider $\frac{g^{jr}}{2}$ times the last two terms in $[\cdot]$. Relabeling the indices r and j in the second term implies

$$\frac{g^{jr}}{2} \frac{\partial g_{kj}}{\partial x^r} - \frac{g^{jr}}{2} \frac{\partial g_{rk}}{\partial x^j} = \frac{g^{jr}}{2} \frac{\partial g_{kj}}{\partial x^r} - \frac{g^{rj}}{2} \frac{\partial g_{jk}}{\partial x^r} = 0.$$

Therefore,

$$\left\{ \begin{matrix} r \\ kr \end{matrix} \right\} = \frac{g^{jr}}{2} \frac{\partial g_{rj}}{\partial x^k}. \quad (8.39)$$

Now recall $g \equiv \det(g_{ij}) = \det(G) > 0$ from Theorem H.1.5. Also from the formula for the inverse of a matrix and this theorem,

$$g^{jr} = A^{rj} (\det G)^{-1} = A^{jr} (\det G)^{-1}$$

where A^{rj} is the rj^{th} cofactor of the matrix (g_{ij}) . Also recall that

$$g = \sum_{r=1}^n g_{rj} A^{rj} \text{ no sum on } j.$$

Therefore, g is a function of the variables $\{g_{rj}\}$ and

$$\frac{\partial g}{\partial g_{rj}} = A^{rj}.$$

From 8.39,

$$\left\{ \begin{matrix} r \\ rk \end{matrix} \right\} = \frac{g^{jr}}{2} \frac{\partial g_{rj}}{\partial x^k} = \frac{1}{2g} \frac{\partial g_{rj}}{\partial x^k} A^{jr} = \frac{1}{2g} \frac{\partial g}{\partial g_{rj}} \frac{\partial g_{rj}}{\partial x^k} = \frac{1}{2g} \frac{\partial g}{\partial x^k}$$

and so from 8.38,

$$\begin{aligned} \operatorname{div}(\mathbf{F}) &= \frac{\partial F^k(\mathbf{x})}{\partial x^k} + \\ &+ F^k(\mathbf{x}) \frac{1}{2g(\mathbf{x})} \frac{\partial g(\mathbf{x})}{\partial x^k} = \frac{1}{\sqrt{g(\mathbf{x})}} \frac{\partial}{\partial x^i} \left(F^i(\mathbf{x}) \sqrt{g(\mathbf{x})} \right). \end{aligned} \quad (8.40)$$

This is the formula for the divergence of a vector field in general curvilinear coordinates.

The Laplacian of a scalar field is nothing more than the divergence of the gradient. In symbols,

$$\Delta \phi \equiv \nabla \cdot \nabla \phi$$

From 8.40 and 8.37 it follows

$$\Delta \phi(\mathbf{x}) = \frac{1}{\sqrt{g(\mathbf{x})}} \frac{\partial}{\partial x^i} \left(g^{ik}(\mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial x^k} \sqrt{g(\mathbf{x})} \right). \quad (8.41)$$

Summarizing all this yields the following theorem.

Theorem H.7.1 *The following formulas hold for the gradient, divergence and Laplacian in general curvilinear coordinates.*

$$(\nabla \phi(\mathbf{x}))_r = \frac{\partial \phi(\mathbf{x})}{\partial x^r}, \quad (8.42)$$

$$(\nabla \phi(\mathbf{x}))^r = g^{rk}(\mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial x^k}, \quad (8.43)$$

$$\operatorname{div}(\mathbf{F}) = \frac{1}{\sqrt{g(\mathbf{x})}} \frac{\partial}{\partial x^i} \left(F^i(\mathbf{x}) \sqrt{g(\mathbf{x})} \right), \quad (8.44)$$

$$\Delta \phi(\mathbf{x}) = \frac{1}{\sqrt{g(\mathbf{x})}} \frac{\partial}{\partial x^i} \left(g^{ik}(\mathbf{x}) \frac{\partial \phi(\mathbf{x})}{\partial x^k} \sqrt{g(\mathbf{x})} \right). \quad (8.45)$$

H.8 Exercises

1. Let $y^1 = x^1 + 2x^2$, $y^2 = x^2 + 3x^3$, $y^3 = x^1 + x^3$. Let

$$\mathbf{F}(\mathbf{x}) = x^1 \mathbf{e}_1(\mathbf{x}) + x^2 \mathbf{e}_2(\mathbf{x}) + (x^3)^2 \mathbf{e}(\mathbf{x}).$$

Find $\operatorname{div}(\mathbf{F})(\mathbf{x})$.

2. For the coordinates of the preceding problem, and ϕ a scalar field, find

$$(\nabla\phi(\mathbf{x}))^3$$

in terms of the partial derivatives of ϕ taken with respect to the variables x^i .

3. Let $y^1 = 7x^1 + 2x^2$, $y^2 = x^2 + 3x^3$, $y^3 = x^1 + x^3$. Let ϕ be a scalar field. Find $\nabla^2\phi(\mathbf{x})$.
4. Derive $\nabla^2 u$ in cylindrical coordinates, r, θ, z , where u is a scalar field on $\mathbb{R}^3$.

$$x = r \cos \theta, \quad y = r \sin \theta, \quad z = z.$$

5. $\uparrow$ Find all solutions to $\nabla^2 u = 0$ which depend only on r where $r \equiv \sqrt{x^2 + y^2}$.
6. Let u be a scalar field on $\mathbb{R}^3$. Find all solutions to $\nabla^2 u = 0$ which depend only on

$$\rho \equiv \sqrt{x^2 + y^2 + z^2}.$$

7. The temperature, u , in a solid satisfies $\nabla^2 u = 0$ after a long time. Suppose in a long pipe of inner radius 9 and outer radius 10 the exterior surface is held at 100° while the inner surface is held at 200° find the temperature in the solid part of the pipe.
8. Show

$$\left\{ \begin{matrix} l \\ ij \end{matrix} \right\} = \frac{\partial \mathbf{e}_i}{\partial x^j} \cdot \mathbf{e}^l.$$

Find the Christoffel symbols of the second kind for spherical coordinates in which $x^1 = \phi$, $x^2 = \theta$, and $x^3 = \rho$. Do the same for cylindrical coordinates letting $x^1 = r$, $x^2 = \theta$, $x^3 = z$.

9. Show that velocity can be expressed as $\mathbf{v} = v_i(\mathbf{x}) \mathbf{e}^i(\mathbf{x})$, where

$$v_i(\mathbf{x}) = \frac{\partial r_i}{\partial x^j} \frac{dx^j}{dt} - r_p(\mathbf{x}) \left\{ \begin{matrix} p \\ ik \end{matrix} \right\} \frac{dx^k}{dt}$$

and $r_i(\mathbf{x})$ are the covariant components of the displacement vector,

$$\mathbf{r} = r_i(\mathbf{x}) \mathbf{e}^i(\mathbf{x}).$$

10. $\uparrow$ Using problem 8 and 9, show that the covariant components of velocity in spherical coordinates are

$$v_1 = \rho^2 \frac{d\phi}{dt}, \quad v_2 = \rho^2 \sin^2(\phi) \frac{d\theta}{dt}, \quad v_3 = \frac{d\rho}{dt}.$$

Hint: First observe that if $\mathbf{r}$ is the position vector from the origin, then $\mathbf{r} = \rho \mathbf{e}_3$ so $r_1 = 0 = r_2$, and $r_3 = \rho$. Now use 9.

H.9 Curl And Cross Products

What about the curl and cross product in general curvilinear coordinates in $\mathbb{R}^3$. In all of this, assume that for $\mathbf{x}$ a set of curvilinear coordinates,

$$\det \left(\frac{\partial y^i}{\partial x^j} \right) > 0 \quad (8.46)$$

Where the $\mathbf{y}_i$ are the usual coordinates in which $\mathbf{e}_k(\mathbf{y}) = \mathbf{i}_k$. This is a statement about orientation.

Theorem H.9.1 *Let 8.46 hold. Then*

$$\det \left(\frac{\partial y^i}{\partial x^j} \right) = \sqrt{g(\mathbf{x})} \quad (8.47)$$

and

$$\det \left(\frac{\partial x^i}{\partial y^j} \right) = \frac{1}{\sqrt{g(\mathbf{x})}}. \quad (8.48)$$

Proof:

$$\mathbf{e}_i(\mathbf{x}) = \frac{\partial y^k}{\partial x^i} \mathbf{i}_k$$

and so

$$g_{ij}(\mathbf{x}) = \frac{\partial y^k}{\partial x^i} \mathbf{i}_k \cdot \frac{\partial y^l}{\partial x^j} \mathbf{i}_l = \frac{\partial y^k}{\partial x^i} \frac{\partial y^k}{\partial x^j}.$$

Therefore, $g = \det(g_{ij}(\mathbf{x})) = \left(\det \left(\frac{\partial y^k}{\partial x^i} \right) \right)^2$. By 8.46, $\sqrt{g} = \det \left(\frac{\partial y^k}{\partial x^i} \right)$ as claimed. Now

$$\frac{\partial y^k}{\partial x^i} \frac{\partial x^i}{\partial y^r} = \delta_r^k$$

and so

$$\det \left(\frac{\partial x^i}{\partial y^r} \right) = \frac{1}{\sqrt{g(\mathbf{x})}}.$$

This proves the theorem.

To get the curl and cross product in curvilinear coordinates, let ϵ^{ijk} be the usual permutation symbol. Thus,

$$\epsilon^{123} = 1$$

and when any two indices in ϵ^{ijk} are switched, the sign changes. Thus

$$\epsilon^{132} = -1, \epsilon^{312} = 1, \text{ etc.}$$

Now define

$$\varepsilon^{ijk}(\mathbf{x}) \equiv \epsilon^{ijk} \frac{1}{\sqrt{g(\mathbf{x})}}.$$

Then for $\mathbf{x}$ and $\mathbf{z}$ satisfying 8.46,

$$\begin{aligned} \varepsilon^{ijk}(\mathbf{x}) \frac{\partial z^r}{\partial x^i} \frac{\partial z^s}{\partial x^j} \frac{\partial z^t}{\partial x^k} &= \epsilon^{ijk} \det \left(\frac{\partial x^p}{\partial y^q} \right) \frac{\partial z^r}{\partial x^i} \frac{\partial z^s}{\partial x^j} \frac{\partial z^t}{\partial x^k} \\ &= \epsilon^{rst} \det \left(\frac{\partial x^p}{\partial y^q} \right) \det \left(\frac{\partial z^i}{\partial x^k} \right) = \epsilon^{rst} \det(MN) \end{aligned}$$

where N is the matrix whose pq^{th} entry is $\frac{\partial x^p}{\partial y^q}$ and M is the matrix whose ik^{th} entry is $\frac{\partial z^i}{\partial x^k}$. Therefore, from the definition of matrix multiplication and the chain rule, this equals

$$= \epsilon^{rst} \det \left(\frac{\partial z^i}{\partial y^p} \right) \equiv \epsilon^{rst}(\mathbf{z})$$

from the above discussion.

Now $\epsilon^{ijk}(\mathbf{y}) = \epsilon^{ijk}$ and for a vector field, $\mathbf{F}$,

$$\text{curl}(\mathbf{F}) \equiv \epsilon^{ijk}(\mathbf{y}) F_{k,j}(\mathbf{y}) \mathbf{e}_i(\mathbf{y}).$$

Therefore, since it is known how these things transform, assuming 8.46, it is routine to write this in terms of $\mathbf{x}$.

$$\begin{aligned} \text{curl}(\mathbf{F}) &= \epsilon^{rst}(\mathbf{x}) \frac{\partial y^i}{\partial x^r} \frac{\partial y^j}{\partial x^s} \frac{\partial y^k}{\partial x^t} F_{p,q}(\mathbf{x}) \frac{\partial x^p}{\partial y^k} \frac{\partial x^q}{\partial y^j} \mathbf{e}_m(\mathbf{x}) \frac{\partial x^m}{\partial y^i} \\ &= \epsilon^{rst}(\mathbf{x}) \delta_r^m \delta_s^q \delta_t^p F_{p,q}(\mathbf{x}) \mathbf{e}_m(\mathbf{x}) = \epsilon^{mnp}(\mathbf{x}) F_{p,q}(\mathbf{x}) \mathbf{e}_m(\mathbf{x}). \end{aligned} \quad (8.49)$$

More simplification is possible. Recalling the definition of $F_{p,q}(\mathbf{x})$,

$$\begin{aligned} \frac{\partial \mathbf{F}}{\partial x^q} &\equiv F_{p,q}(\mathbf{x}) \mathbf{e}^p(\mathbf{x}) = \frac{\partial}{\partial x^q} [F_p(\mathbf{x}) \mathbf{e}^p(\mathbf{x})] \\ &= \frac{\partial F_p(\mathbf{x})}{\partial x^q} \mathbf{e}^p(\mathbf{x}) + F_p(\mathbf{x}) \frac{\partial \mathbf{e}^p}{\partial x^q} = \frac{\partial F_p(\mathbf{x})}{\partial x^q} \mathbf{e}^p(\mathbf{x}) - F_r(\mathbf{x}) \left\{ \begin{matrix} r \\ pq \end{matrix} \right\} \mathbf{e}^p(\mathbf{x}) \end{aligned}$$

by Theorem H.6.2. Therefore,

$$F_{p,q}(\mathbf{x}) = \frac{\partial F_p(\mathbf{x})}{\partial x^q} - F_r(\mathbf{x}) \left\{ \begin{matrix} r \\ pq \end{matrix} \right\}$$

and so

$$\text{curl}(\mathbf{F}) = \epsilon^{mnp}(\mathbf{x}) \frac{\partial F_p(\mathbf{x})}{\partial x^q} \mathbf{e}_m(\mathbf{x}) - \epsilon^{mnp}(\mathbf{x}) F_r(\mathbf{x}) \left\{ \begin{matrix} r \\ pq \end{matrix} \right\} \mathbf{e}_m(\mathbf{x}).$$

However, because $\left\{ \begin{matrix} r \\ pq \end{matrix} \right\} = \left\{ \begin{matrix} r \\ qp \end{matrix} \right\}$, the second term in this expression equals 0. To see this,

$$\epsilon^{mnp}(\mathbf{x}) \left\{ \begin{matrix} r \\ pq \end{matrix} \right\} = \epsilon^{mpq}(\mathbf{x}) \left\{ \begin{matrix} r \\ qp \end{matrix} \right\} = -\epsilon^{mnp}(\mathbf{x}) \left\{ \begin{matrix} r \\ pq \end{matrix} \right\}.$$

Therefore, by 8.49,

$$\text{curl}(\mathbf{F}) = \epsilon^{mnp}(\mathbf{x}) \frac{\partial F_p(\mathbf{x})}{\partial x^q} \mathbf{e}_m(\mathbf{x}). \quad (8.50)$$

What about the cross product of two vector fields? Let $\mathbf{F}$ and $\mathbf{G}$ be two vector fields. Then in terms of standard coordinates, $\mathbf{y}$,

$$\begin{aligned} \mathbf{F} \times \mathbf{G} &= \epsilon^{ijk}(\mathbf{y}) F_j(\mathbf{y}) G_k(\mathbf{y}) \mathbf{e}_i(\mathbf{y}) \\ &= \epsilon^{rst}(\mathbf{x}) \frac{\partial y^i}{\partial x^r} \frac{\partial y^j}{\partial x^s} \frac{\partial y^k}{\partial x^t} F_p(\mathbf{x}) \frac{\partial x^p}{\partial y^k} G_q(\mathbf{x}) \frac{\partial x^q}{\partial y^j} \mathbf{e}_l(\mathbf{x}) \frac{\partial x^l}{\partial y^i} \\ &= \epsilon^{rst}(\mathbf{x}) \delta_r^p \delta_s^q \delta_t^l F_p(\mathbf{x}) G_q(\mathbf{x}) \mathbf{e}_l(\mathbf{x}) = \epsilon^{lpq}(\mathbf{x}) F_p(\mathbf{x}) G_q(\mathbf{x}) \mathbf{e}_l(\mathbf{x}). \end{aligned} \quad (8.51)$$

This implies

Theorem H.9.2 Suppose $\mathbf{x}$ is a system of curvilinear coordinates in $\mathbb{R}^3$ such that

$$\det \left(\frac{\partial y^i}{\partial x^j} \right) > 0.$$

Let

$$\varepsilon^{ijk}(\mathbf{x}) \equiv \epsilon^{ijk} \frac{1}{\sqrt{g(\mathbf{x})}}.$$

Then the following formulas for curl and cross product hold in this system of coordinates.

$$\operatorname{curl}(\mathbf{F}) = \varepsilon^{mqp}(\mathbf{x}) \frac{\partial F_p(\mathbf{x})}{\partial x^q} \mathbf{e}_m(\mathbf{x}),$$

and

$$\mathbf{F} \times \mathbf{G} = \varepsilon^{lpq}(\mathbf{x}) F_p(\mathbf{x}) G_q(\mathbf{x}) \mathbf{e}_l(\mathbf{x}).$$

Bibliography

- [1] **Apostol, T. M.**, *Calculus second edition*, Wiley, 1967.
- [2] **Apostol T.** *Calculus Volume II Second edition*, Wiley 1969.
- [3] **Apostol, T. M.**, *Mathematical Analysis*, Addison Wesley Publishing Co., 1974.
- [4] **Baker, Roger**, *Linear Algebra*, Rinton Press 2001.
- [5] **Bartle R.G.**, *A Modern Theory of Integration*, Grad. Studies in Math., Amer. Math. Society, Providence, RI, 2000.
- [6] **Chahal J. S.** , *Historical Perspective of Mathematics* 2000 B.C. - 2000 A.D.
- [7] **Davis H. and Snider A.**, *Vector Analysis* Wm. C. Brown 1995.
- [8] **D'Angelo, J. and West D.** *Mathematical Thinking Problem Solving and Proofs*, Prentice Hall 1997.
- [9] **Edwards C.H.** *Advanced Calculus of several Variables*, Dover 1994.
- [10] **Euclid**, *The Thirteen Books of the Elements*, Dover, 1956.
- [11] **Fitzpatrick P. M.**, *Advanced Calculus a course in Mathematical Analysis*, PWS Publishing Company 1996.
- [12] **Fleming W.**, *Functions of Several Variables*, Springer Verlag 1976.
- [13] **Greenberg, M.** *Advanced Engineering Mathematics*, Second edition, Prentice Hall, 1998
- [14] **Gurtin M.** *An introduction to continuum mechanics*, Academic press 1981.
- [15] **Hardy G.**, *A Course Of Pure Mathematics*, Tenth edition, Cambridge University Press 1992.
- [16] **Horn R. and Johnson C.** *matrix Analysis*, Cambridge University Press, 1985.
- [17] **Karlin S. and Taylor H.** *A First Course in Stochastic Processes*, Academic Press, 1975.
- [18] **Kuttler K. L.**, *Basic Analysis*, Rinton
- [19] **Kuttler K.L.**, *Modern Analysis* CRC Press 1998.
- [20] **Lang S.** *Real and Functional analysis* third edition Springer Verlag 1993. Press, 2001.
- [21] **Nobel B. and Daniel J.** *Applied Linear Algebra*, Prentice Hall, 1977.

- [22] **Rose, David, A.**, The College Math Journal, vol. 22, No.2 March 1991.
- [23] **Rudin, W.**, *Principles of mathematical analysis*, McGraw Hill third edition 1976
- [24] **Rudin W.**, *Real and Complex Analysis*, third edition, McGraw-Hill, 1987.
- [25] **Salas S. and Hille E.**, *Calculus One and Several Variables*, Wiley 1990.
- [26] **Sears and Zemansky**, *University Physics, Third edition, Addison Wesley 1963*.
- [27] **Tierney John**, *Calculus and Analytic Geometry*, fourth edition, Allyn and Bacon, Boston, 1969.
- [28] **Yosida K.**, *Functional Analysis, Springer Verlag, 1978*.

Index

- C^1 , 502
- C^k , 502
- Δ , 588
- Σ , 19
- $\cap$, 18
- $\cup$, 18
- ∇^2 , 588
- n^{th} term test, 263

- Abel's formula, 343, 661
- absolute convergence, 260
- adjugate, 333, 654
- agony, pain and suffering, 547
- alternating series, 265
- alternating series test, 265
- amplitude, 28, 108
- angle between planes, 398
- angle between vectors, 370
- angular velocity, 387
- angular velocity vector, 710
- antiderivatives, 148
- arc length, 439
- area of a parallelogram, 381
- arithmetic mean, 538

- balance of momentum, 600
- barallelepiped
 - volume, 389
- basis, 733
- bezier curves, 439
- binomial series, 280
- binomial theorem, 27
- binormal, 459
- bounded, 424
- box product, 389
- box product, 734

- capacitance, 212
- carbon dating, 208
- cardioid, 465
- Cartesian coordinates, 294
- catenary, 206
- Cauchy, 50
- Cauchy condensation test, 262
- Cauchy mean value theorem, 98
- Cauchy product, 271
- Cauchy Schwarz, 302
- Cauchy Schwarz inequality, 368, 377
- Cauchy sequence, 426
- Cauchy sequence, 66, 426
- Cauchy stress, 602
- Cavendish, 470
- center of mass, 386, 720
- central force, 462
- central force field, 469
- centrifugal acceleration, 724
- centripetal acceleration, 724
- centripetal force, 469
- chain rule, 514
- change of variables formula, 573
- characteristic equation, 347
- characteristic value, 346
- Christoffel symbols, 743
- circular functions, 105
- circular helix, 460
- circular shells, 189
- circulation density, 632
- classical adjoint, 333
- closed set, 358
- coefficient of thermal conductivity, 518
- cofactor, 326, 328, 652
- cofactor matrix, 328
- column vector, 310
- compact, 430
- comparison test, 260
- complement, 358
- completeness, 66
- completeness axiom, 19
- complex conjugate, 38
- complex numbers, 37
- component, 363, 379
- component of a force, 372
- components of a matrix, 308
- concave down, 101
- concave up, 101
- conformable, 311
- conservation of linear momentum, 718

- conservation of mass, 599
- conservative, 625
- constitutive laws, 605
- contented set, 681
- continuity
 - limit of a sequence, 428
- continuous function, 50, 413
- continuous functions
 - properties, 418
- converge, 426
- Coordinates, 293
- Coriolis acceleration, 724
- Coriolis acceleration
 - earth, 726
- Coriolis force, 469, 724
- Cramer's rule, 336, 654
- critical point, 524
- cross product, 381
 - area of parallelogram, 381
 - coordinate description, 382
 - distributive law, 384
 - geometric description, 381
 - limits, 417
- cross product
 - general curvilinear coordinates, 749
- curl, 587
- curl
 - general curvilinear coordinates, 749
- curvature, 454, 458
- curvilinear coordinates, 741
- cycloid, 631

- D'Alembert, 493
- damped oscillation, 116
- Darboux, 231
- Darboux integral, 231
- deformation gradient, 600
- density and mass, 548
- derivative, 497
 - intermediate value property, 100
 - mean value theorem, 97
- derivative of a function, 86, 433
- determinant, 325, 647
 - Laplace expansion, 328
 - product, 650
 - product of matrices, 330
 - transpose, 649
- diameter, 424
- difference quotient, 86, 433
- differentiability and continuity, 87
- differentiable, 495
- differentiable matrix, 707
- differential, 499
- differential equation, 180
- differential equations, 740
- differentiation rules, 88, 436
- directed line segment, 298
- direction vector, 298
- directional derivative, 486
- directrix, 380
- Dirichlet function, 44
- Dirichlet test, 265
- distance, 637
- distance formula, 300
- divergence, 587
- divergence, 745
 - general curvilinear coordinates, 746
- divergence theorem, 592
- domain, 43
- donut, 583
- dot product, 368
- dual basis, 734
- dual basis, 738

- eigenspace, 348
- eigenvalue, 346, 538
- eigenvector, 346
- Einstein summation convention, 391
- ellipse, 32
- entries of a matrix, 308
- equality of mixed partial derivatives, 491
- Eulerian coordinates, 600
- extreme value theorem, 57, 424

- Fermat's principle, 141
- Fibonacci sequence, 47, 426
- Fick's law, 518, 611
- first order linear differential equations, 241
- flip, 669
- focus, 32, 304
- force, 361
- force field, 442, 469
- Foucault pendulum, 726
- Fourier law of heat conduction, 518
- Frenet Serret formulas, 459
- frequency, 108
- Fresnel integral, 248
- frustum of a cone, 196
- function
 - even, 96
 - odd, 96
 - uniformly continuous, 69
- fundamental theorem line integrals, 625
- fundamental theorem of algebra, 641
- fundamental theorem of calculus, 226

- gamma function, 247
- Gauss's theorem, 592
- geometric mean, 538
- geometric series, 258
- gradient, 488
- gradient
 - contravariant components, 745
 - covariant components, 745
- gradient vector, 517
- Grammian, 395
- Green's theorem, 618
- grid, 542, 550
- grids, 673
- hanging chain, 205
- harmonic, 492
- heat equation, 492
- Heine Borel, 70, 448
- Heine Borel theorem, 430
- Hessian matrix, 527, 539
- Holder's inequality, 144
- Hooke's law, 209
- hyperbola, 34
- implicit differentiation, 118
- implicit function theorem, 119, 663, 667
- improper integrals, 244
- impulse, 717
- increment of volume
 - increment of area, 572
- indefinite integral, 148
- index
 - lowering, 735
 - raising, 735
- inductance, 212
- inflection point, 101
- initial value problem, 147, 180
- inner product, 368
- inscribed angle, 30
- integral, 148
- integral test, 264
- integrand, 148
- integration by parts, 154, 240
- intercepts, 401
- intercepts of a surface, 403
- interior point, 357
- intermediate value theorem, 56, 83
- intervals, 18
- inverse
 - left, 653
 - right, 653
- inverse function theorem, 119, 667
- inverse image, 45
- inverses and determinants, 335, 653
- invertible, 317
- isocles triangle, 639
- iterated integrals, 544
- Jacobian determinant, 572
- Jordan content, 678
- Jordan set, 681
- joule, 374
- Kepler's first law, 471
- Kepler's laws, 470
- Kepler's third law, 473
- kilogram, 386
- kinetic energy, 715
- kinetic energy, 739
- Kroneker delta, 391
- Lagrange multipliers, 535, 668, 669
- Lagrange remainder, 254
- Lagrangian coordinates, 600
- Lagrangian formalism, 740
- Laplace expansion, 328, 652
- Laplacian, 492
- Laplacian
 - general curvilinear coordinates, 746
- law of cosines, 639
- law of sines, 29
- least squares regression, 492
- Lebesgue number, 430, 678
- Lebesgue's theorem, 681
- length of smooth curve, 441
- limit comparison test, 261
- limit of a function, 58, 414, 431
- limit point, 361, 483
- limits and continuity, 416
- line integral, 444
- linear approximation, 499
- linear combination, 650
- linear momentum, 717
- linear transformation, 319, 498
- lines
 - parametric equation, 298
- Lipschitz, 70, 418, 419
- lizards
 - surface area, 580
- local extremum, 523
- local maximum, 93, 523
- local minimum, 93, 523
- logarithmic differentiation, 122
- Lotka Volterra equations, 412
- lower sum, 551, 674
- main diagonal, 329

- mass balance, 599
- material coordinates, 600
- matrix, 307
 - inverse, 317
 - left inverse, 654
 - lower triangular, 329, 654
 - right inverse, 654
 - upper triangular, 328, 654
- matrix multiplication
 - entries, 313
 - properties, 315
- matrix transpose, 316
- matrix transpose properties, 316
- max. min.theorem, 57
- mean value theorem
 - for integrals, 233
- Merten's theorem, 272
- method of bisection, 58
- metric tensor, 395, 736
- metric tensor, 744
- midpoint sum, 230
- minor, 326, 328, 652
- mixed partial derivatives, 490
- moment of a force, 386
- motion, 600
- moving coordinate system, 707, 723
 - acceleration , 724
- multi-index, 414

- natural logarithms, 114
- Navier, 613
- nested interval lemma, 57, 77, 423
- Newton, 364
 - second law, 710
- Newton Raphson method, 144
- Newton's laws, 710
- Newton's second law, 740
- nilpotent, 341
- nonlinear spring, 251
- nonremovable discontinuity, 50
- normal vector to plane, 397

- Ohm's law, 212
- one to one, 319
- onto, 320
- open cover, 429
- open set, 357
- orientable, 624
- orientation, 442
- oriented curve, 442
- origin, 293
- orthogonal matrix, 340, 657
- orthonormal, 657

- oscillation
 - critically damped, 211
 - over damped, 211
 - underdamped, 211
- osculating plane, 454, 458

- p series, 262
- parabola, 30
- parallelepiped, 388
- parameter, 298
- parametric equation, 298
- parametrization, 440
- partial derivative, 487
- partial fractions expansion, 164
- partial summation formula, 265
- partition, 216
- partition of unity, 693
- permutation symbol, 391
- perpendicular, 371
- phase shift, 28, 108
- Piola Kirchhoff stress, 604
- plane containing three points, 399
- planes, 397
- polar form complex number, 39
- polynomials in n variables, 414
- position vector, 296, 362
- power series, 275
- precession of a top, 718
- primitive, 669
- principal normal, 454, 458
- product of matrices, 311
- product rule, 88
 - cross product, 436
 - dot product, 436
 - matrices, 707
- profit, 94
- projection of a vector, 373
- properties of integral
 - properties, 225
- Pythagorean theorem, 636

- quadric surfaces, 401
- quotient rule, 88

- Raabe's test, 274
- radius of curvature, 454, 458
- range, 43
- ratio test, 266
- rational function, 46
- rational function of cosines and sines, 168
- raw eggs, 722
- recurrence relation, 47, 426

- recursively defined sequence, 83
- recursively defined sequence, 47, 426
- refinement of a grid, 542, 550
- refinement of grids, 673
- regular Sturm Liouville problem, 243
- removable discontinuity, 50
- resistance, 212
- resultant, 363
- Revenue, 94
- Riemann criterion, 219
- Riemann criterion, 675
- Riemann integrable, 218
- Riemann integral, 231, 543, 550
- Riemann integral, 675
- Riemann sum, 229
- right handed system, 381
- Rolle's theorem, 98
- root test, 267
- rot, 587
- row operations, 330
- row vector, 310
- saddle point, 527
- scalar field, 587
- scalar multiplication, 295
- scalar potential, 626
- scalar product, 368
- scalars, 295, 307
- second derivative test, 540
- separable differential equations, 205
- sequence of partial sums, 257
- sequences, 47, 425
- sequential compactness, 69, 427, 448
- sequentially compact, 427
- Simpson's rule, 232, 242
- singular point, 524
- skew symmetric, 316
- smooth curve, 439
- Snell's law, 142
- spacial coordinates, 600
- span, 650
- spectrum, 346
- speed, 364
- spherical coordinates, 738
- squeezing theorem, 60
- squeezing theorem, 64
- standard matrix, 498
- standard position, 362
- Stoke's theorem, 621
- Stokes, 613
- subtend, 20
- summation notation, 19
- support of a function, 692
- symmetric, 316
- symmetric form of a line, 299
- Taylor series, 275
- Taylor's formula, 242, 253
- torque vector, 386
- torsion, 459
- torus, 583
- trace of a surface, 403
- traces, 401
- transformation rules, 741
- trapezoid rule, 232, 242
- triangle inequality, 144, 302, 369
 - complex numbers, 38
- trigonometric functions, 638
- uniform continuity, 429
- uniformly continuous, 69, 419, 429
- unit tangent vector, 454, 458
- upper and lower sums, 216
- upper sum, 551, 674
- Urysohn's lemma, 430
- vector, 295
- vector
 - contravariant components, 735
 - covariant components, 735
- vector field, 442, 587
- vector fields, 410
- vector potential, 590
- vector valued function
 - continuity, 413
 - derivative, 433
 - integral, 433
 - limit theorems, 415
- vector valued functions, 409
- vectors, 293, 361
- velocity, 364
- volume element, 572, 700
- volume increment, 572
- volume of parallelepiped, 571
- volume of unit ball in n dimensions, 609
- wave equation, 492
- Weierstrass, 50
- work, 443
- Wronskian, 342, 661
- zero matrix, 308