Proceedings of the 34th
Conference on Decision & Control
New Orleans, LA - December 1995

A Differential Geometric Approach to Nonlinear Filtering : the Projection Filter*

Damiano Brigo, Bernard Hanzon
Department of Econometrics
Free University
1081 HV Amsterdam, The Netherlands
{dbrigo,bhanzon}@econ.vu.nl

Francois Le Gland
IRISA / INRIA
Campus de Beaulieu
35042 Rennes Cédex, France
legland@irisa.fr

Abstract : We present a new and systematic method of approximating exact nonlinear filters with finite dimensional filters, using the differential geometric approach to statistics. We define rigorously the projection filter in the case of exponential families. We propose a convenient exponential family, which allows to simplify the projection filter equation, and to define an a posteriori measure of the performance of the projection filter.

1 INTRODUCTION

The filtering problem consists in estimating the state of a stochastic differential system from noisy observations. In the linear Gaussian case the problem was solved by Kalman, who introduced the well known Kalman filter, a finite dimensional system of equations for the first two conditional moments of the state given the observations. In the general nonlinear case, the filtering problem consists in computing the conditional density of the state given the observations. This density is the solution of a stochastic partial differential equation, the Kushner–Stratonovich equation. The general nonlinear problem is far more complicated because the resulting nonlinear filter is not finite dimensional in general.

In [4] Hanzon introduced the projection filter (PF), which is a finite dimensional nonlinear filter based on the differential geometric approach to statistics. The projection filter is obtained by projecting the Kushner–Stratonovich equation onto the tangent space of a finite dimensional manifold of probability densities, according to the Fisher information metric and its extension to the infinite dimensional space of square roots of densities, the Hellinger distance.

The purpose of this paper is to provide an introduction to the projection filter. We provide a rigorous definition of the PF in the case of a manifold of exponential probability densities. It should be noticed that the resulting finite–dimensional equations can be easily implemented, and do not require any complicated differential geometric operations. We also present some formulae concerning auxiliary quantities, such as the projection residual, the purpose of which is to provide a local measure of the quality of the filter behaviour, and we work out the case of discrete–time observations. The filters are derived by using the geometric approach, but in principle the reader can rederive them by using the assumed density idea without using any Riemannian geometry, see Brigo, Hanzon and LeGland [2]. In [2], we also develop explicit formulae for the particular example of the cubic sensor, and we present some numerical simulations, and comparisons between the projection filter and optimal filter obtained by the numerical solution of the nonlinear filtering equation.

2 STATISTICAL MANIFOLDS

Let $\mathcal{M}(\lambda)$ be the set of all non–negative and finite measures on $\mathbf{R}^n$, which are absolutely continuous w.r.t. the Lebesgue measure λ, and whose density is positive λ–a.e. On the set of square roots of all the densities of measures in $\mathcal{M}(\lambda)$, we define the following metric, called the Hellinger metric : $d(\sqrt{p_1},\sqrt{p_2}) := \|\sqrt{p_1}-\sqrt{p_2}\|$, where $\|\cdot\|$ denotes the norm in $L_2(\lambda)$.

Let S denote the following family of probability densities :

$$S = \{p(\cdot,\theta) \,:\, \theta \in \Theta\},$$

where $\Theta \subseteq \mathbf{R}^m$ is open, and the corresponding set of square roots of densities

$$S^{1/2} = \{\sqrt{p(\cdot,\theta)} \,:\, \theta\in\Theta\} \subset L_2(\lambda) \ .$$

We assume that, for all $\theta \in \Theta$

$$\{\frac{\partial\sqrt{p(\cdot,\theta)}}{\partial\theta_1},\cdots,\frac{\partial\sqrt{p(\cdot,\theta)}}{\partial\theta_m}\} \ ,$$

are linearly independent vectors, hence $S^{1/2}$ is a finite–dimensional submanifold of $L_2(\lambda)$, and a basis for the tangent vector space at $\sqrt{p(\cdot,\theta)}$ to $S^{1/2}$ is given by :

$$L_{\sqrt{p(\cdot,\theta)}}S^{1/2} = \mathrm{span}\{\frac{\partial\sqrt{p(\cdot,\theta)}}{\partial\theta_1},\cdots,\frac{\partial\sqrt{p(\cdot,\theta)}}{\partial\theta_m}\} \ .$$

The inner product of any two basis elements is defined, according to the L_2 inner product

$$\langle\frac{\partial\sqrt{p(\cdot,\theta)}}{\partial\theta_i},\frac{\partial\sqrt{p(\cdot,\theta)}}{\partial\theta_j}\rangle =$$

*This work was partially supported by the Commission of the European Communities, under the SCIENCE project *System Identification*, project number SC1*-CT92-0779, and by the Army Research Office, under grant DAAH04-95-1-0164. Damiano Brigo was also supported by an *Advanced Studying Fellowship* of the University of Padua.

0-7803-2685-7/95 $4.00 © 1995 IEEE

$$= \tfrac{1}{4} \int \frac{1}{p(x,\theta)} \frac{\partial p(x,\theta)}{\partial \theta_i} \frac{\partial p(x,\theta)}{\partial \theta_j} \, d\lambda(x) = \tfrac{1}{4} \, g_{ij}(\theta) \; .$$

This is, up to the numeric factor $\frac{1}{4}$, the Fisher information metric, and the matrix $g(\theta) = (g_{ij}(\theta))$ is called the Fisher information matrix, see [1].

We conclude this section with a lemma on exponential families, which will be used throughout the paper, see [1].

Definition 2.1 *Let* $\{c_1, \cdots, c_m\}$ *be* linearly independent *scalar functions defined on* $\mathbf{R}^n$, *and assume that the convex set* Θ_0

$$\{\theta \in \mathbf{R}^m \; : \; \psi(\theta) = \log \int \exp[\theta^T c(x)] \, d\lambda(x) < \infty\} \; ,$$

has non–empty interior. *Then*

$$S = \{p(\cdot,\theta) \, , \, \theta \in \Theta\}, \quad p(x,\theta) := \exp[\theta^T c(x) - \psi(\theta)] \; ,$$

where $\Theta \subseteq \Theta_0$ *is open, is called an* exponential family *of probability densities.*

Lemma 2.2 *The function* ψ *is infinitely differentiable in* Θ, *and*

$$E_{p(\cdot,\theta)}\{c_i\} = \partial_i \psi(\theta) =: \eta_i(\theta) \; ,$$

$$E_{p(\cdot,\theta)}\{c_i c_j\} = \partial^2_{ij} \psi(\theta) + \partial_i \psi(\theta) \, \partial_j \psi(\theta) \; .$$

In addition, the Fisher information matrix satisfies

$$g_{ij}(\theta) = \partial^2_{ij} \psi(\theta) = \partial_i \eta_j(\theta) \; .$$

3 THE NONLINEAR FILTERING PROBLEM

On the probability space $(\Omega, \mathcal{F}, P)$ with the filtration $\{\mathcal{F}_t \, , \, t \geq 0\}$ we consider the following state and observation equations

$$\begin{aligned} dX_t &= f_t(X_t) \, dt + \sigma_t(X_t) \, dW_t \; , \\ dY_t &= h_t(X_t) \, dt + dV_t \; . \end{aligned} \tag{1}$$

These equations are Itô stochastic differential equations. In (1), the unobserved state process $\{X_t \, , \, t \geq 0\}$ and the observation process $\{Y_t \, , \, t \geq 0\}$ are taking values in $\mathbf{R}^n$ and $\mathbf{R}^d$ respectively, the noise processes $\{W_t \, , \, t \geq 0\}$ and $\{V_t \, , \, t \geq 0\}$ are two Brownian motions, taking values in $\mathbf{R}^p$ and $\mathbf{R}^d$ respectively, with covariance matrices Q_t and R_t respectively. We assume that R_t is invertible for all $t \geq 0$, which implies that, without loss of generality, we can assume that $R_t = I$ for all $t \geq 0$. Finally, the initial state X_0 and the noise processes $\{W_t \, , \, t \geq 0\}$ and $\{V_t \, , \, t \geq 0\}$ are mutually independent.

We assume that the initial state X_0 has a density p_0 w.r.t. the Lebesgue measure λ on $\mathbf{R}^n$, and has finite moments of any order, and we make the following assumptions on the coefficients f_t, $a_t := \sigma_t \, Q_t \, \sigma_t^T$, and h_t of the system (1)

(A) Local Lipschitz continuity : for all $R > 0$, there exists $K_R > 0$ such that

$$|f_t(x) - f_t(x')| \leq K_R \, |x - x'| \; ,$$

$$\|a_t(x) - a_t(x')\| \leq K_R \, |x - x'| \; ,$$

for all $t \geq 0$, and for all $x, x' \in B_R$, the ball of radius R.

(B) Non–explosion : there exists $K > 0$ such that

$$x^T f_t(x) \leq K \, (1 + |x|^2) \; ,$$

$$\text{trace } a_t(x) \leq K \, (1 + |x|^2) \; ,$$

for all $t \geq 0$, and for all $x \in \mathbf{R}^m$.

(C) Polynomial growth : there exist $K > 0$ and $r \geq 0$ such that

$$|h_t(x)| \leq K \, (1 + |x|^r) \; ,$$

for all $t \geq 0$, and for all $x \in \mathbf{R}^m$.

Under assumptions (A) and (B), there exists a unique solution $\{X_t \, , \, t \geq 0\}$ to the state equation, see Khasminskii [6], and X_t has finite moments of any order. For all $t \geq 0$, the associated backward diffusion operator $\mathcal{L}_t$ is defined by

$$\mathcal{L}_t = \sum_{i=1}^{n} f_t^i \, \frac{\partial}{\partial x_i} + \tfrac{1}{2} \sum_{i,j=1}^{n} a_t^{ij} \, \frac{\partial^2}{\partial x_i \partial x_j} \; .$$

Under the additional assumption (C) the following *finite energy* condition holds

$$E \int_0^T |h_t(X_t)|^2 \, dt < \infty \; , \qquad \text{for all } T \geq 0.$$

The nonlinear filtering problem consists in finding the conditional probability distribution π_t of the state X_t given the observations up to time t, i.e. $\pi_t(dx) := P[X_t \in dx \mid \mathcal{Y}_t]$, where $\mathcal{Y}_t := \sigma(Y_s \, , 0 \leq s \leq t)$ — see [3] for an introduction. We assume that for all $t \geq 0$, the probability distribution π_t has a density p_t w.r.t. the Lebesgue measure on $\mathbf{R}^n$. Then $\{p_t \, , \, t \geq 0\}$ satisfies

$$\begin{aligned} dp_t &= \mathcal{L}_t^* \, p_t \, dt - \tfrac{1}{2} \, p_t \, [|h_t|^2 - E_{p_t}\{|h_t|^2\}] \, dt \\ &\quad + \sum_{k=1}^{d} p_t \, [h_t^k - E_{p_t}\{h_t^k\}] \circ dY_t^k \; . \end{aligned}$$

in a suitable functional space, where $E_{p_t}\{\cdot\}$ denotes the expectation w.r.t. the probability density p_t, i.e. the conditional expectation given the observations up to time t, and where for all $t \geq 0$, the forward diffusion operator $\mathcal{L}_t^*$ is defined by

$$\mathcal{L}_t^* \phi = - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \, [f_t^i \, \phi] + \tfrac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} \, [a_t^{ij} \, \phi] \; ,$$

for any test function ϕ defined on $\mathbf{R}^n$. We shall frequently work with square roots of densities, rather than densities themselves. Then, we compute by formal rules, using the Stratonovich form :

$$d\sqrt{p_t} = \frac{1}{2\sqrt{p_t}} \circ dp_t \;=\; \mathcal{P}_t(\sqrt{p_t})\, dt - \mathcal{Q}_t^0(\sqrt{p_t})\, dt + \sum_{k=1}^{d} \mathcal{Q}_t^k(\sqrt{p_t}) \circ dY_t^k \,, \tag{2}$$

where the nonlinear time dependent operators $\mathcal{P}_t$, $\mathcal{Q}_t^0$ and $\mathcal{Q}_t^k$ for $k = 1, \cdots, d$ are defined by

$$\begin{aligned} \mathcal{P}_t(\sqrt{p}) &:= \frac{\mathcal{L}_t^* \, p}{2\sqrt{p}} \,, \\ \mathcal{Q}_t^0(\sqrt{p}) &:= \tfrac{1}{4}\sqrt{p}\,[|h_t|^2 - E_p\{|h_t|^2\} \,, \\ \mathcal{Q}_t^k(\sqrt{p}) &:= \tfrac{1}{2}\sqrt{p}\,[h_t^k - E_p\{h_t^k\} \,, \end{aligned} \tag{3}$$

respectively.

4 THE EXPONENTIAL PROJECTION FILTER

In this section we present the rigorous definition of an exponential projection filter. We will show that if we choose $S^{1/2}$ as the set of square roots of probability densities of a finite dimensional exponential family, then under an additional assumption, see Theorem 4.1 below, the operators $\mathcal{P}_t$ and $\mathcal{Q}_t^k$ for $k = 0, 1, \cdots, d$, introduced in (3) map elements of $S^{1/2}$ into $L_2(\lambda)$. This is important because in general the operator $\mathcal{P}_t$ is unbounded, i.e. does not map $L_2(\lambda)$ into $L_2(\lambda)$, and the projection of the coefficients in the right hand side of the Kushner–Stratonovich equation is not possible. Let us consider the following exponential family of probability densities

$$S := \{p(\cdot, \theta)\,,\, \theta \in \Theta\}, \quad p(x, \theta) := \exp[\theta^T c(x) - \psi(\theta)] \,.$$

For all $\theta \in \Theta$, and all $t \geq 0$, we define

$$\begin{aligned} \alpha_{t,\theta} := \frac{\mathcal{L}_t^* \, p(\cdot, \theta)}{p(\cdot, \theta)} &= -\sum_{i=1}^{n} [\, f_t^i \, \frac{\partial(\theta^T c)}{\partial x_i} + \frac{\partial f_t^i}{\partial x_i} \,] \\ &+ \tfrac{1}{2} \sum_{i,j=1}^{n} [\, a_t^{ij} \, \frac{\partial(\theta^T c)}{\partial x_i} \, \frac{\partial(\theta^T c)}{\partial x_j} \\ &+ a_t^{ij} \, \frac{\partial^2(\theta^T c)}{\partial x_i \partial x_j} + 2\, \frac{\partial a_t^{ij}}{\partial x_j} \, \frac{\partial(\theta^T c)}{\partial x_i} + \frac{\partial^2 a_t^{ij}}{\partial x_i \partial x_j} \,] \,. \end{aligned} \tag{4}$$

We assume that for all $\theta \in \Theta$, and all $t \geq 0$

$$E_{p(\cdot,\theta)}\{|\alpha_{t,\theta}|^2\} < \infty \quad \text{and} \quad E_{p(\cdot,\theta)}\{|h_t|^4\} < \infty \,,$$

which implies that $\mathcal{P}_t(\sqrt{p(\cdot, \theta)})$ and $\mathcal{Q}_t^k(\sqrt{p(\cdot, \theta)})$ for $k = 0, 1, \cdots, d$ are vectors of $L_2(\lambda)$ for all $\theta \in \Theta$, and all $t \geq 0$.

Let us consider the equation (2) for $\{\sqrt{p_t}\,,\, t \geq t_0\}$, starting at time t_0 from the initial condition $\sqrt{p_{t_0}} = \sqrt{p(\cdot, \theta_0)} \in S^{1/2}$ for some $\theta_0 \in \Theta$, i.e.

$$d\sqrt{p_t} \;=\; \mathcal{P}_t(\sqrt{p_t})\, dt - \mathcal{Q}_t^0(\sqrt{p_t})\, dt + \sum_{k=1}^{d} \mathcal{Q}_t^k(\sqrt{p_t}) \circ dY_t^k \,.$$

Under our assumptions, $\mathcal{P}_{t_0}(\sqrt{p_{t_0}})$ and $\mathcal{Q}_{t_0}^k(\sqrt{p_{t_0}})$ for $k = 0, 1, \cdots, d$ are vectors of $L_2(\lambda)$, which we can project onto the finite dimensional tangent vector space $L_{\sqrt{p(\cdot,\theta_0)}} S^{1/2}$. For this purpose, we define for all $\theta \in \Theta$ the orthogonal projection Π_θ

$$\begin{aligned} L_2(\lambda) &\longrightarrow L_{\sqrt{p(\cdot,\theta)}} S^{1/2} \\ v &\longmapsto \sum_{i=1}^{m} [\sum_{j=1}^{m} 4 g^{ij}(\theta) \; (v, \frac{\partial\sqrt{p(\cdot, \theta)}}{\partial \theta_j})] \; \frac{\partial\sqrt{p(\cdot, \theta)}}{\partial \theta_i} \,. \end{aligned}$$

The exponential projection filter for the exponential family S is defined as the solution of the following stochastic differential equation on the manifold $S^{1/2}$:

$$\begin{aligned} d\sqrt{p(\cdot, \theta_t)} \;&=\; \Pi_{\theta_t} \circ \mathcal{P}_t(\sqrt{p(\cdot, \theta_t)})\, dt \\ &- \Pi_{\theta_t} \circ \mathcal{Q}_t^0(\sqrt{p(\cdot, \theta_t)})\, dt \\ &+ \sum_{k=1}^{d} \Pi_{\theta_t} \circ \mathcal{Q}_t^k(\sqrt{p(\cdot, \theta_t)}) \circ dY_t^k \,. \end{aligned} \tag{5}$$

We can now state the following theorem :

Theorem 4.1 *Assume that, in addition to (A), (B) and (C), the coefficients f_t, a_t and h_t of the system (1), and the coefficients c of the exponential family S are such that :*

$$E_{p(\cdot,\theta)}\{|\alpha_{t,\theta}|^2\} < \infty \quad \text{and} \quad E_{p(\cdot,\theta)}\{|h_t|^4\} < \infty \,,$$

holds for all $\theta \in \Theta$, and all $t \geq 0$, where the expression of $\alpha_{t,\theta}$ is given in (4).

Then, for all $\theta \in \Theta$, and all $t \geq 0$, $\Pi_\theta \circ \mathcal{P}_t$ and $\Pi_\theta \circ \mathcal{Q}_t^k$ for $k = 0, 1, \cdots, d$ are vector fields on the exponential manifold $S^{1/2}$.

The projection filter density $p_t^\pi = p(\cdot, \theta_t)$ is described by equation (5), and the projection filter parameters satisfy the following stochastic differential equation :

$$\begin{aligned} g(\theta_t) \circ d\theta_t &= E_{p(\cdot,\theta_t)}\{\mathcal{L}_t \, c\}\, dt \\ &- E_{p(\cdot,\theta_t)}\{\tfrac{1}{2}\, |h_t|^2 \, [c - \eta(\theta_t)]\}\, dt \\ &+ \sum_{k=1}^{d} E_{p(\cdot,\theta_t)}\{h_t^k \, [c - \eta(\theta_t)]\} \circ dY_t^k \,, \end{aligned} \tag{6}$$

where the expectation parameters $\eta_1(\theta), \cdots, \eta_m(\theta)$ are defined in Lemma 2.2.

Under the assumptions on the coefficients, this equation has a unique solution, up to the a.s. positive time $\tau := \inf\{t > 0 \,:\, \theta_t \notin \Theta\}$.

The proof of the theorem can be found in [2].

5 THE PROJECTION RESIDUAL AND A CONVENIENT EXPONENTIAL FAMILY

In this section, we are interested in defining quantities which will provide a local measure of the quality of the projection filter approximation. Compare equation (2) for the (square root of the) true density p_t, i.e.

$$d\sqrt{p_t} = \mathcal{P}_t(\sqrt{p_t})\,dt - \mathcal{Q}_t^0(\sqrt{p_t})\,dt + \sum_{k=1}^{d} \mathcal{Q}_t^k(\sqrt{p_t}) \circ dY_t^k \,, \tag{7}$$

and equation (5) for the (square root of the) projection filter density $p_t^\pi = p(\cdot, \theta_t)$, i.e.

$$d\sqrt{p_t^\pi} = \Pi_{\theta_t} \circ \mathcal{P}_t(\sqrt{p_t^\pi})\,dt - \Pi_{\theta_t} \circ \mathcal{Q}_t^0(\sqrt{p_t^\pi})\,dt + \sum_{k=1}^{d} \Pi_{\theta_t} \circ \mathcal{Q}_t^k(\sqrt{p_t^\pi}) \circ dY_t^k \,. \tag{8}$$

Two steps are involved in using the projection filter density p_t^π as an approximation of the true density p_t : We make a first approximation by evaluating the right–hand side of equation (7) at the current projection filter density p_t^π and not at the true density p_t. Even with this approximation, the resulting coefficients $\mathcal{P}_t(\sqrt{p_t^\pi})$ and $\mathcal{Q}_t^k(\sqrt{p_t^\pi})$ for $k = 0, 1, \cdots, d$ would make the solution leave the manifold $S^{1/2}$, and we make a second approximation by projecting these coefficients on the linear space $L_{\sqrt{p_t^\pi}} S^{1/2}$ via the projection mapping Π_{θ_t}. In order to express the error occurring in the second approximation step at time t, we define the prediction residual operator $\mathcal{R}_t^\bullet$ and the correction residual operators $\mathcal{R}_t^k$ for $k = 0, 1, \cdots, d$ as follows :

$$\mathcal{R}_t^\bullet := \mathcal{P}_t - \Pi_{\theta_t} \circ \mathcal{P}_t$$

$$\mathcal{R}_t^k := \mathcal{Q}_t^k - \Pi_{\theta_t} \circ \mathcal{Q}_t^k \,.$$

These operators, when applied to the square root of density $\sqrt{p_t^\pi} = \sqrt{p(\cdot, \theta_t)} \in S^{1/2}$ yield vectors of $L_2(\lambda)$. We call such vectors *projection residuals* : they give a local measure of the quality of the approximation involved in the projection filter. We can compute the norm of such vectors according to the norm $\|\cdot\|$ in $L_2(\lambda)$, and we define the prediction residual norm $r_t^\bullet$ and correction residual norms r_t^k for $k = 0, 1, \cdots, d$ as follows :

$$r_t^\bullet := \|\mathcal{R}_t^\bullet(\sqrt{p_t^\pi})\|$$

$$r_t^k := \|\mathcal{R}_t^k(\sqrt{p_t^\pi})\| \,.$$

However, we are still missing a single measure of the local error resulting from the projection. We define below a single residual operator, only in the case where $\mathcal{R}_t^k = 0$ for all $t \geq 0$, and all $k = 1, \cdots, d$. In this case, we define the total residual operator $\mathcal{R}_t^*$ as :

$$\mathcal{R}_t^* := \mathcal{R}_t^\bullet - \mathcal{R}_t^0 \,,$$

and the corresponding total residual norm r_t^* as :

$$r_t^* := \|\mathcal{R}_t^*(\sqrt{p_t^\pi})\| \,.$$

Notice that if in addition $\mathcal{R}_t^0 = 0$, then r_t^* reduces to $r_t^\bullet$. In the next section we will introduce manifolds $S_\bullet^{1/2}$ for which such a definition is applicable. Now we try to give some intuition for the above definition. Suppose we replace in equations (7) and (8) the observation $\{Y_t\,,\, t \geq 0\}$ with some smooth process $\{u_t\,,\, t \geq 0\}$, e.g. a regularized approximation, i.e. we consider the equations

$$\frac{d}{dt}\sqrt{p_t} = \mathcal{P}_t(\sqrt{p_t}) - \mathcal{Q}_t^0(\sqrt{p_t}) + \sum_{k=1}^{d} \mathcal{Q}_t^k(\sqrt{p_t})\, \dot{u}_t^k \,, \tag{9}$$

and

$$\frac{d}{dt}\sqrt{p_t^\pi} = \Pi_{\theta_t} \circ \mathcal{P}_t(\sqrt{p_t^\pi}) - \Pi_{\theta_t} \circ \mathcal{Q}_t^0(\sqrt{p_t^\pi}) + \sum_{k=1}^{d} \Pi_{\theta_t} \circ \mathcal{Q}_t^k(\sqrt{p_t^\pi})\, \dot{u}_t^k \,. \tag{10}$$

In this case, we can define a single residual operator expressing the difference between the rate of change in the smooth Kushner–Stratonovich equation (9) and the rate of change in the smooth projection filter equation (10), i.e.

$$\mathcal{R}_t^u := \mathcal{R}_t^\bullet - \mathcal{R}_t^0 + \sum_{k=1}^{d} \mathcal{R}_t^k\, \dot{u}_t^k \,.$$

Of course, if we return to the original situation, e.g. letting the regularized approximation $\{u_t\,,\, t \geq 0\}$ converge to the observation $\{Y_t\,,\, t \geq 0\}$, there is no limit to the smooth residual operator $\mathcal{R}_t^u$, *unless* $\mathcal{R}_t^k = 0$ for all $t \geq 0$, and all $k = 1, \cdots, d$. In this case only, we define the total residual operator $\mathcal{R}_t^*$ as above.

From now on, and throughout the paper, we assume for simplicity that $h_t(x) = h(x)$ does not depend explicitly on time. This is necessary in order to define the simplifying *time invariant* exponential family $S_\bullet$ below.

We can now state the following theorem :

Theorem 5.1 *Let* $s := \text{rank}\{h^1, \cdots, h^d, \frac{1}{2}|h|^2\}$. *There exist s linearly independent functions $\{c_1, \cdots, c_s\}$ defined on $\mathbf{R}^n$, such that for all $x \in \mathbf{R}^n$*

$$\tfrac{1}{2}\,|h(x)|^2 = \sum_{i=1}^{s} \lambda_i^0\, c_i(x) \,, \quad h^k(x) = \sum_{i=1}^{s} \lambda_i^k\, c_i(x) \,, \tag{11}$$

for $k = 1, \cdots, d$. Remaining functions $\{c_{s+1}, \cdots, c_m\}$ are chosen such that

$$S_\bullet := \{p(\cdot, \theta)\,,\, \theta \in \Theta\} \,, \quad p(x, \theta) := \exp[\theta^T c(x) - \psi(\theta)] \,,$$

where $\Theta \subseteq \mathbf{R}^m$ is open, is an exponential family of probability densities.

Assume that, in addition to (A), (B) and (C), the coefficients f_t and a_t of the system (1), and the coefficients c of the exponential family $S_\bullet$ are such that :

$$E_{p(\cdot,\theta)}\{|\alpha_{t,\theta}|^2\} < \infty \,,$$

holds for all $\theta \in \Theta$, and all $t \geq 0$, where the expression of $\alpha_{t,\theta}$ is given in (4).

Then, for the projection filter associated with the exponential family $S_\bullet$, the correction residual norms r_t^k are identically zero for all $t \geq 0$, and all $k = 0, 1, \cdots, d$, and the stochastic differential equation for the parameters reduces to :

$$d\theta_t = [g(\theta_t)]^{-1} E_{p(\cdot,\theta_t)}\{\mathcal{L}_t c\} dt - \lambda_\bullet^0 dt + \sum_{k=1}^{d} \lambda_\bullet^k dY_t^k , \qquad (12)$$

where for all $k = 0, 1, \cdots, d$ the m-dimensional vector $\lambda_\bullet^k$ is defined by

$$\lambda_\bullet^k = [\lambda_1^k \cdots \lambda_s^k \ 0 \cdots 0]^T .$$

Under the assumptions on the coefficients, this equation has a unique solution, up to the a.s. positive exit time $\tau := \inf\{t > 0 : \theta_t \notin \Theta\}$.

PROOF : All the assumptions of Theorem 4.1 are satisfied, and therefore the solution of the stochastic differential equation for the projection filter with manifold $S_\bullet^{1/2}$ exists and is unique up to the a.s. positive exit time τ.

Next, we prove that the correction residual norms vanish. Indeed, it follows from (11) that

$$\mathcal{Q}_t^0(\sqrt{p(\cdot,\theta_t)}) = \tfrac{1}{4}[|h|^2 - E_{p(\cdot,\theta_t)}\{|h|^2\}] \sqrt{p(\cdot,\theta_t)} = \tfrac{1}{2} \sum_{i=1}^{s} \lambda_i^0 [c_i - E_{p(\cdot,\theta_t)}\{c_i\}] \sqrt{p(\cdot,\theta_t)} ,$$

and similarly

$$\mathcal{Q}_t^k(\sqrt{p(\cdot,\theta_t)}) = \tfrac{1}{2}[h^k - E_{p(\cdot,\theta_t)}\{h^k\}] \sqrt{p(\cdot,\theta_t)} = \tfrac{1}{2} \sum_{i=1}^{s} \lambda_i^k [c_i - E_{p(\cdot,\theta_t)}\{c_i\}] \sqrt{p(\cdot,\theta_t)} ,$$

for $k = 1, \cdots, d$. We remark that

$$\tfrac{1}{2}[c_i - E_{p(\cdot,\theta_t)}\{c_i\}] \sqrt{p(\cdot,\theta_t)} = \frac{\partial \sqrt{p(\cdot,\theta_t)}}{\partial \theta_i} ,$$

hence $\mathcal{Q}_t^k(\sqrt{p(\cdot,\theta_t)}) \in L_{\sqrt{p(\cdot,\theta_t)}} S_\bullet^{1/2}$ for $k = 0, 1, \cdots, d$. Therefore, the projection does not modify these vectors since they already lie in the tangent space of $S_\bullet^{1/2}$.

Finally, the equation for the parameters is obtained via straightforward calculations, see [2]. □

What the above theorem shows is that the projection residuals are greatly simplified if we make use of the functions $\{h^1, \cdots, h^d, \frac{1}{2}|h|^2\}$ in the definition of the exponential manifold, i.e. if we choose the functions $\{c_1, \cdots, c_m\}$ in such a way that the functions $\{h^1, \cdots, h^d, \frac{1}{2}|h|^2\}$ belong to span$\{c_1, \cdots, c_m\}$. Indeed, $\mathcal{R}_t^k(\sqrt{p_t^\pi}) = 0$ for all $t \geq 0$, and all $k = 0, 1, \cdots, k$, whereas

$$\frac{1}{\sqrt{p_t^\pi}} \mathcal{R}_t^\bullet(\sqrt{p_t^\pi}) = \tfrac{1}{2} \frac{\mathcal{L}_t^* p_t^\pi}{p_t^\pi} - \tfrac{1}{2}[c - \eta(\theta_t)]^T [g(\theta_t)]^{-1} E_{p(\cdot,\theta_t)}\{\mathcal{L}_t c\} .$$

The diffusion coefficient in the stochastic differential equation (12) for the parameters θ_t is constant. This implies that (12) can be seen as either an Itô or a Stratonovich stochastic differential equation, so that it satisfies the formal rules of calculus. Moreover, for the numerical solution of such an equation, the simpler Euler scheme coincides with the Milshtein scheme, which is a strongly convergent scheme of order 1, see Kloeden and Platen [7].

Notice also that we have still some freedom left, and we may wonder whether one can use this to select m and the remaining functions $\{c_{s+1}, \cdots, c_m\}$ in order to reduce the total residual norm $r_t^* = r_t^\bullet$. However, a great prudence is needed, because the filter may become complicated and numerical problems may arise. In general, a trade–off is necessary in order to obtain an accurate, but still not too involved, exponential family and the associated projection filter.

6 THE CASE OF DISCRETE–TIME OBSERVATIONS

In this section we present the effect of choosing the exponential family $S_\bullet$, in the case of a nonlinear filtering problem with discrete–time observations. In this model, the state process is as in equation (1), i.e.

$$dX_t = f_t(X_t) dt + \sigma_t(X_t) dW_t ,$$

but only discrete–time observations are available

$$z_n = h(X_{t_n}) + v_n ,$$

at times $0 = t_0 < \cdots < t_n < \cdots$, where $\{v_n , n \geq 0\}$ is a Gaussian white noise sequence independent of $\{X_t , t \geq 0\}$.

The nonlinear filtering problem consists in finding the conditional density $p_n(x)$ of the state X_{t_n} given the observations up to time t_n, i.e. such that $P[X_{t_n} \in dx \mid \mathcal{Z}_n] = p_n(x) dx$, where $\mathcal{Z}_n := \sigma(z_0, \cdots, z_n)$. We define also the prediction conditional density $p_n^-(x) dx = P[X_{t_n} \in dx \mid \mathcal{Z}_{n-1}]$. The sequence $\{p_n , n \geq 0\}$ satisfies a recurrent equation, and the transition from p_{n-1} to p_n is decomposed in two steps, as explained in [5], [10] :

□ **Prediction step.** Between time t_{n-1} and t_n, we solve the Fokker–Planck equation

$$\frac{\partial p_t^n}{\partial t} = \mathcal{L}_t^* p_t^n , \quad p_{t_{n-1}}^n = p_{n-1} .$$

The solution at final time t_n defines the prediction conditional density $p_n^- = p_{t_n}^n$.

□ **Correction step.** At time t_n, the observation z_n is combined with the prediction conditional density p_n^- via the Bayes rule

$$p_n(x) = c_n \, \Psi_n(x) \, p_n^-(x) \; , \qquad (13)$$

where c_n is a normalizing constant, and $\Psi_n(x)$ denotes the likelihood function for the estimation of X_{t_n} based on the observation z_n only, i.e.

$$\Psi_n(x) := \exp\{ -\tfrac{1}{2}|z_n - h(x)|^2 \} \; . \qquad (14)$$

If we use the exponential family $S_\bullet$ defined above, then we obtain the projection filter density $p(\cdot, \theta_n)$, and the transition from θ_{n-1} to θ_n is also decomposed in two steps :

□ **Prediction step.** Between time t_{n-1} and t_n, we solve the ODE

$$g(\theta_t^n) \, \dot\theta_t^n = E_{p(\cdot,\theta_t^n)}\{\mathcal{L}_t \, c\} \; , \quad \theta_{t_{n-1}}^n = \theta_{n-1} \; .$$

The solution at final time t_n defines the prediction parameters $\theta_n^- = \theta_{t_n}^n$.

□ **Correction step.** Substituting the approximation $p(\cdot, \theta_n^-)$ into formula (13), we observe that the resulting density does not leave the exponential family $S_\bullet$. Indeed, it follows from (11) and (14) that

$$\begin{aligned}\Psi_n(x) &= \exp\{ -\tfrac{1}{2}|h(x)|^2 + \sum_{k=1}^d h^k(x)\, z_n^k - \tfrac{1}{2}|z_n|^2 \} \\ &\propto \exp\{ -\sum_{l=1}^s \lambda_l^0 \, c_l(x) + \sum_{l=1}^s [\sum_{k=1}^d \lambda_l^k \, z_n^k] \, c_l(x) \} \; ,\end{aligned}$$

and the parameters are updated according to the formula

$$\theta_n = \theta_n^- - \lambda_\bullet^0 + \sum_{k=1}^d \lambda_\bullet^k \, z_n^k \; ,$$

which is *exact*.

7 CONCLUSION

In this paper, we have introduced a new and systematic way of designing approximate finite–dimensional filters.

One major issue left is the choice of the exponential family S. A first answer has been given in Section 5, but this does not completely solve the problem : with the choice of the family $S_\bullet$ there is still some freedom left in the choice of the dimension m and in the choice of the remaining functions $\{c_{s+1}, \cdots, c_m\}$, which could be used to reduce the total residual norm $r_t^* = r_t^\bullet$.

This freedom could also be used to design an adaptive scheme for the choice of the exponential family S.

It would also be useful to obtain for all $t \geq 0$ an estimate of the distance between the optimal filter density p_t and the projection filter density p_t^π, in terms of the total residual norm history $\{r_s^*, \, 0 \leq s \leq t\}$.

Finally, we would like to define projection filters for discrete time systems, and relate this problem with the work of Kulhavý [8], [9]. Another motivation for this study will be to obtain efficient numerical schemes for the solution of the stochastic differential equation satisfied by the projection filter parameters, i.e. equation (6) for a general family S, or equation (12) for the family $S_\bullet$.

Each of these problems requires further investigation, and we hope to address all of them in a subsequent work.

8 REFERENCES

[1] Shun–ichi AMARI. *Differential–Geometrical Methods in Statistics.* Volume 28 of *Lecture Notes in Statistics*, Springer Verlag, Berlin, 1985.

[2] D. BRIGO, B. HANZON, and F. LE GLAND. *A differential geometric approach to nonlinear filtering : the projection filter.* Publication Interne 914, IRISA, June 1995.

[3] M.H.A. DAVIS and S.I. MARCUS. An introduction to nonlinear filtering. In M. Hazewinkel and J.C. Willems, editors, *Stochastic Systems : the Mathematics of Filtering and Identification and Applications, Les Arcs 1980*, D. Reidel, Dordrecht, 1981.

[4] B. HANZON. A differential–geometric approach to approximate nonlinear filtering. In C.T.J. Dodson, editor, *Geometrization of Statistical Theory*, pages 219–223, ULMD Publications, University of Lancaster, 1987.

[5] A.H. JAZWINSKI. *Stochastic Processes and Filtering Theory.* Volume 64 of *Mathematics in Science and Engineering*, Academic Press, New York, 1970.

[6] R.Z. KHASMINSKII. *Stochastic Stability of Differential Equations.* Sijthoff and Noordhoff, Alphen aan den Rijn, 1980.

[7] P.E. KLOEDEN and E. PLATEN. *Numerical Solution of Stochastic Differential Equations.* Volume 23 of *Applications of Mathematics*, Springer Verlag, New York, 1992.

[8] R. KULHAVÝ. Recursive nonlinear estimation : a geometric approach. *Automatica*, 26(3):545–555, 1990.

[9] R. KULHAVÝ. Recursive nonlinear estimation : geometry of a space of posterior densities. *Automatica*, 28(2):313–323, 1992.

[10] P.S. MAYBECK. *Stochastic Models, Estimation, and Control. Volume 2.* Volume 141–2 of *Mathematics in Science and Engineering*, Academic Press, New York, 1979.