The Exponential Projection Filter and the Selection of the Exponential Family*

Damiano Brigo, Francois Le Gland
IRISA / INRIA
Campus de Beaulieu
35042 Rennes Cédex, France
{dbrigo,legland}@irisa.fr

Bernard Hanzon
Department of Econometrics
Free University
1081 HV Amsterdam, The Netherlands
bhanzon@econ.vu.nl

Abstract : We present the projection filter, an approximate finite–dimensional filter based on the differential geometric approach to statistics. We recall the definition of the projection filter in the case of exponential families, and we give some hints about the selection of the coefficients in the exponential family.

1 INTRODUCTION

The filtering problem consists in estimating the state of a stochastic differential system from noisy observations. In the linear Gaussian case the solution consists of the 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 [10] 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 for the square–root of the conditional density onto the tangent space of a finite dimensional manifold of (square–root 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.

A rigorous definition of the projection filter has been given in Brigo, Hanzon and LeGland [5] for the case of a manifold of (square–root of) densities from an exponential family. In the general case of (square–root of) densities from a non–exponential family, a rigorous definition is still possible, see Brigo, Hanzon and LeGland [4], but the computations are less simple in this case.

The purpose of this paper is to give some hints about the selection of the m–dimensional exponential family

$$
\begin{aligned}
S &= \{p(\cdot,\theta)\,,\ \theta\in\Theta\} \\
p(x,\theta) &:= \exp[\theta^T c(x) - \psi(\theta)]\ ,
\end{aligned}
\tag{1}
$$

i.e. about the choice of the coefficient $c = (c_1, \cdots, c_m)$.

We will show that a convenient choice is to incorporate the components of the observation function h (see the model in Section 3 below) and possibly also the scalar function $\frac{1}{2}\,|h|^2$, into the coefficients of the family, i.e. to make sure that the functions $\{h_1, \cdots, h_d\}$ or $\{h_1, \cdots, h_d, \frac{1}{2}\,|h|^2\}$ belong to span$\{c_1, \cdots, c_m\}$. Let the corresponding exponential families be denoted as S_* and $S_\bullet$ respectively. We recall from Brigo, Hanzon and LeGland [5], see also [6] and [7] for details, that :

- with the choice $S = S_*$, the diffusion coefficient in the stochastic differential equation for the PF parameter $\{\theta_t\,,\ t \geq 0\}$ is constant, i.e. the equation has a very simple form,
- with the choice $S = S_*$, it is possible to define an estimate of the local error resulting from the projection filter approximation,
- with the choice $S = S_\bullet$, the correction step in the nonlinear filtering algorithm with discrete–time observations, is handled exactly, without any error.

Another approximation method in nonlinear filtering is the assumed density filter (ADF) obtained by closing the equations for a few conditional moments, under the assumption that the exact conditional density is of a given form, see Kushner [11]. In the case of the exponential family (1), one would

(i) write the equation for the evolution of the conditional c–moment $\eta_t = \mathbf{E}[c(X_t) \mid \mathcal{Y}_t]$, and

(ii) evaluate all the conditional expectations appearing in the r.h.s. of the equation as if the conditional density was the density in the exponential family uniquely defined by the expectation parameter η_t.

*This work was partially supported by the Commission of the European Communities, under the SCIENCE project *System Identification*, project number SC1*–CT92–0779, and under the HCM project *Statistical Inference for Stochastic Processes*, project number CHRX–CT92–0078, and by the Army Research Office, under grant DAAH04–95–1–0164.

The method is logically inconsistent in the sense that the resulting ADF is different whether the moment equations are written in Itô or Stratonovich form. We will show below, see Brigo, Hanzon and LeGland [6] and [7] for details, that for any exponential family, the Stratonovich–based ADF coincides with the PF. We will also show that :

- with the choice $S = S_*$, the Itô–based ADF coincides with the Stratonovich–based ADF.

2 STATISTICAL MANIFOLDS ASSOCIATED WITH AN EXPONENTIAL FAMILY

We begin with the definition of an exponential family, and a lemma which will be used throughout the paper, see Amari [1].

Definition 2.1 *Let $\{c_1, \cdots, c_m\}$ be scalar functions defined on $\mathbf{R}^n$, such that $\{1, c_1, \cdots, c_m\}$ are* linearly independent, *and assume that the convex set Θ_0*

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

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.*

Associated with the exponential family S, is the corresponding set of square roots of densities

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

It is easily checked 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 m–dimensional submanifold of L_2, and 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

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

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

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 Amari [1].

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

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

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

More generally

$$E_{p(\cdot,\theta)}\{c_{i_1} \cdots c_{i_k}\} =$$

$$= \exp[-\psi(\theta)] \; \frac{\partial^k}{\partial \theta_{i_1} \cdots \partial \theta_{i_k}} \; \exp[\psi(\theta)] \; .$$

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

$$\begin{aligned} dX_t &= f_t(X_t) \, dt + \sigma_t(X_t) \, dW_t \; , \\ dY_t &= h_t(X_t) \, dt + dV_t \; . \end{aligned} \tag{2}$$

These equations are Itô stochastic differential equations. In (2), 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 I respectively. 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 classical assumptions on the coefficients of the system (2) : local Lipschitz continuity and non–explosion for f_t, and $a_t := \sigma_t \, Q_t \, \sigma_t^T$, and polynomial growth for h_t. For more detailed statements and references, see Brigo, Hanzon and LeGland [6]. For all $t \geq 0$, the associated backward diffusion operator $\mathcal{L}_t$ is defined by

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

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 Davis and Marcus [8] for an introduction. For all $t \geq 0$, the probability distribution π_t has a density p_t w.r.t. the Lebesgue measure on $\mathbf{R}^n$, and $\{p_t \, , \, t \geq 0\}$ satisfies

$$\begin{aligned} dp_t &= \mathcal{L}_t^* \, p_t \, dt - \tfrac{1}{2} \, p_t \, [|h_t|^2 - E_{p_t}\{|h_t|^2\}] \, dt \\ &\quad + \sum_{k=1}^d p_t \, [h_t^k - E_{p_t}\{h_t^k\}] \circ dY_t^k \; . \end{aligned} \tag{3}$$

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 $\mathbf{R}^n$. Using the Stratonovich chain rule, we obtain that $\{\sqrt{p_t} \, , \, t \geq 0\}$ satisfies

$$\begin{aligned} d\sqrt{p_t} \; &= \; \frac{1}{2\sqrt{p_t}} \circ dp_t = \mathcal{P}_t(\sqrt{p_t}) \, dt - \mathcal{Q}_t^0(\sqrt{p_t}) \, dt \\ &\quad + \sum_{k=1}^d \mathcal{Q}_t^k(\sqrt{p_t}) \circ dY_t^k \; , \end{aligned} \tag{4}$$

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

$$\begin{aligned} \mathcal{P}_t(\sqrt{p}) &:= \frac{\mathcal{L}_t^* \, p}{2\sqrt{p}} \; , \\ \mathcal{Q}_t^0(\sqrt{p}) &:= \tfrac{1}{4} \sqrt{p} \, [|h_t|^2 - E_p\{|h_t|^2\} \; , \\ \mathcal{Q}_t^k(\sqrt{p}) &:= \tfrac{1}{2} \sqrt{p} \, [h_t^k - E_p\{h_t^k\} \; , \end{aligned} \tag{5}$$

respectively.

4 THE EXPONENTIAL PROJECTION FILTER

In this section we recall 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 from a finite–dimensional exponential family, then under an additional assumption, see Theorem 4.3 below, the operators $\mathcal{P}_t$ and $\mathcal{Q}_t^k$ for $k = 0, 1, \cdots, d$, introduced in (5) map elements of $S^{1/2}$ into L_2. This is important because in general the operator $\mathcal{P}_t$ is unbounded, i.e. does not map L_2 into L_2, 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

$$\begin{aligned} \alpha_{t,\theta} := \frac{\mathcal{L}_t^* \, p(\cdot, \theta)}{p(\cdot, \theta)} &= - \sum_{i=1}^n [\, f_t^i \, \frac{\partial(\theta^T c)}{\partial x_i} + \frac{\partial f_t^i}{\partial x_i} \,] \\ &\quad + \tfrac{1}{2} \sum_{i,j=1}^n [\, a_t^{ij} \, \frac{\partial(\theta^T c)}{\partial x_i} \, \frac{\partial(\theta^T c)}{\partial x_j} \\ &\quad + a_t^{ij} \, \frac{\partial^2(\theta^T c)}{\partial x_i \partial x_j} + 2 \, \frac{\partial a_t^{ij}}{\partial x_j} \, \frac{\partial(\theta^T c)}{\partial x_i} + \frac{\partial^2 a_t^{ij}}{\partial x_i \partial x_j} \,] \; . \end{aligned} \tag{6}$$

We make the following additional assumption on the coefficients f_t, a_t and h_t of the system (2), and on the coefficient c of the exponential family (1) :

(E) 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 for all $\theta \in \Theta$, and all $t \geq 0$.

Remark 4.1 Sufficient explicit conditions for Assumption (E) to hold can be easily given. For example, Assumption (E) will hold if the coefficients f_t and its first derivatives, a_t and its first and second derivatives, h_t, c and its first and second derivatives, have at most polynomial growth, and if densities in the exponential family integrate any polynomial.

The idea now is to project the L_2 vectors $\mathcal{P}_t(\sqrt{p(\cdot, \theta)})$ and $\mathcal{Q}_t^k(\sqrt{p(\cdot, \theta)})$ for $k = 0, 1, \cdots, d$, onto the finite–dimensional tangent vector space $L_{\sqrt{p(\cdot,\theta)}} S^{1/2}$. For this purpose, we define for all $\theta \in \Theta$ the orthogonal projection Π_θ

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

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

$$\begin{aligned} d\sqrt{p(\cdot, \theta_t)} \; &= \; \Pi_{\theta_t} \circ \mathcal{P}_t(\sqrt{p(\cdot, \theta_t)}) \, dt \\ &\quad - \Pi_{\theta_t} \circ \mathcal{Q}_t^0(\sqrt{p(\cdot, \theta_t)}) \, dt \\ &\quad + \sum_{k=1}^d \Pi_{\theta_t} \circ \mathcal{Q}_t^k(\sqrt{p(\cdot, \theta_t)}) \circ dY_t^k \; . \end{aligned} \tag{7}$$

Remark 4.2 Notice that although at first sight (7) may look like a stochastic PDE, it is just a finite dimensional SDE which can be equivalently written using different coordinates as an equation in $\Theta \subset \mathbf{R}^m$ for the parameter θ_t. The explicit form of this SDE is given in the following theorem.

Theorem 4.3 *Assume that, in addition to the assumptions of Section 3, the coefficients f_t, a_t and h_t of the system (2), and the coefficients c of the exponential family S satisfy Assumption (E), i.e.*

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

holds for all $\theta \in \Theta$, and all $t \geq 0$, where the expression of $\alpha_{t,\theta}$ is given in (6).

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 (7), and the projection filter param-

eters satisfy the following stochastic differential equation :

$$
\begin{aligned}
g(\theta_t) \circ d\theta_t &= E_{p(\cdot,\theta_t)}\{\mathcal{L}_t\, c\}\, dt \\
&- E_{p(\cdot,\theta_t)}\{\tfrac{1}{2}\, |h_t|^2\, [c - \eta(\theta_t)]\}\, dt \\
&+ \sum_{k=1}^{d} E_{p(\cdot,\theta_t)}\{h_t^k\, [c - \eta(\theta_t)]\} \circ dY_t^k \ ,
\end{aligned} \tag{8}
$$

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 Brigo, Hanzon and LeGland [6].

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_* below.

Theorem 4.4 *Let $s := \mathrm{rank}\{h^1, \cdots, h^d\}$. There exist s linearly independent functions $\{c_1, \cdots, c_s\}$ defined on $\mathbf{R}^n$, such that for all $x \in \mathbf{R}^n$*

$$
h^k(x) = \sum_{i=1}^{s} \lambda_i^k\, c_i(x) \ , \tag{9}
$$

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

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

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

Assume that, in addition to the assumptions of Section 3, the coefficients f_t and a_t of the system (2), and the coefficients c of the exponential family S_* 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 (6).

Then, for the projection filter associated with the exponential family S_*, the stochastic differential equation for the parameters reduces to :

$$
\begin{aligned}
d\theta_t &= [g(\theta_t)]^{-1}\, E_{p(\cdot,\theta_t)}\{\mathcal{L}_t\, c\}\, dt \\
&- [g(\theta_t)]^{-1}\, E_{p(\cdot,\theta_t)}\{\tfrac{1}{2}\, |h_t|^2\, [c - \eta(\theta_t)]\}\, dt \\
&+ \sum_{k=1}^{d} \lambda_*^k\, dY_t^k \ ,
\end{aligned} \tag{10}
$$

where for all $k = 0, 1, \cdots, d$ the m–dimensional vector λ_*^k is defined by

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

5 ASSUMED DENSITY FILTERS

If one tries to develop approximation schemes starting from the nonlinear filtering equations, one is confronted with the problem that the conditional densities do not belong in general to any finite dimensional class of densities. One heuristic way to deal with this problem is to consider the moment equations and to assume arbitrarily that the conditional densities belong to some finite dimensional class of densities, even if this is known to be wrong. The resulting moment equations will in general be *inconsistent*, but by selecting carefully a limited number of moment equations one can obtain a consistent definition of an approximate filter, which is called an assumed density filter in the literature, see Kushner [11], and Maybeck [12, Section 12.7]. As will be shown, it also matters whether the selected moment equations are taken in Itô or in Stratonovich form.

We give now a more formal definition of assumed density filters, associated with the exponential family S. Throughout the remaining part of the paper we assume that, in addition to the assumptions of Section 3, the coefficient f_t of the system (2) has at most polynomial growth when $|x|$ goes to infinity. Consider any twice differentiable function c which, together with its derivatives up to order 2, have at most polynomial growth when $|x|$ goes to infinity. Then the conditions given in Fujisaki, Kallianpur and Kunita [9] are fulfilled for the c–moment to satisfy

$$
\begin{aligned}
d\,\pi_t(c) &= \pi_t(\mathcal{L}_t\, c)\, dt \\
&- \tfrac{1}{2}\, [\pi_t(|h_t|^2\, c) - \pi_t(|h_t|^2)\, \pi_t(c)]\, dt \\
&+ \sum_{k=1}^{d} [\pi_t(h_t^k\, c) - \pi_t(h_t^k)\, \pi_t(c)] \circ dY_t^k \ ,
\end{aligned} \tag{11}
$$

see equation (3) above. The Itô version of this c–moment equation is obtained from the Itô version of equation (3), and holds under the conditions just described.

The following is a generalization of the concept of assumed conditional probability density filters as introduced in Kushner [11].

Definition 5.1 *In accordance with the ADF principle, the Itô–based ADF is defined by the Itô stochastic differential equations*

$$
\begin{aligned}
d\eta_t &= E_{\eta_t}\{\mathcal{L}_t\, c\}\, dt \\
&+ \sum_{k=1}^{d} E_{\eta_t}\{h_t^k\, [c - \eta_t]\}\, [\, dY_t^k - E_{\eta_t}\{h_t^k\}\, dt\,] \ .
\end{aligned} \tag{12}
$$

Similarly the Stratonovich–based ADF is defined by the Stratonovich stochastic differential equations

$$
\begin{aligned}
d\eta_t &= E_{\eta_t}\{\mathcal{L}_t\, c\}\, dt - \tfrac{1}{2}\, E_{\eta_t}\{|h_t|^2\, [c - \eta]\}\, dt \\
&+ \sum_{k=1}^{d} E_{\eta_t}\{h_t^k\, [c - \eta_t]\} \circ dY_t^k \ .
\end{aligned} \tag{13}
$$

Although this may be surprising at first, the Itô–based ADF and the Stratonovich–based ADF are *different* filters in general. This can be shown by working out the Itô–based and Stratonovich–based Gaussian assumed density filters for the cubic sensor problem. The fact that they are different is due to the *inconsistency* that is inherent to the ADF–concept : selecting a different set of equations to which it is applied leads to different results. Examples are given in Brigo, Hanzon and LeGland [7].

The main theorem of this second part of the paper can now be stated.

Theorem 5.2 *For any exponential family S, the projection filter coincides with the Stratonovich–based assumed density filter.*

We will give a concise proof based on stochastic calculus. A second proof is given in Brigo, Hanzon and LeGland [7], which makes explicit use of projections on the tangent spaces, and relies on a crucial result from the theory of *information geometry*, i.e. the biorthogonality relations between the tangent vectors corresponding to the canonical parameters, and the tangent vectors corresponding to the expectation parameters, see Amari [1].

PROOF. We start from equation (8) for the projection filter canonical parameters, i.e.

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

According to Lemma 2.2, the expectation parameters can be expressed in terms of the canonical parameters as

$$\eta_i = \eta_i(\theta) = E_{p(\cdot,\theta)}\{c_i\} = E_{p_E(\cdot,\eta)}\{c_i\} \ ,$$

with derivatives

$$\frac{\partial}{\partial\theta_i} \eta_j(\theta) = g_{ij}(\theta) \ .$$

The chain rule for Stratonovich integrals immediately gives

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

which is exactly equation (13) obtained using the assumed density filter idea. □

The equivalence between the Stratonovich–based ADF and the PF is shown to hold for exponential families. In general, for other families of distributions such equivalence does *not* hold. In Brigo, Hanzon and LeGland [7], a simple example is given in which a particular *curved* (Gaussian) exponential family is considered.

One of the striking features of the Theorem 5.2 is that it yields a characterization of the projection filters for exponential families in terms of assumed density filters, which are not intrinsically based on differential geometry, and can be understood without using geometric concepts. Indeed, in Brigo [3] and [2] the Gaussian projection filter is derived by using the assumed–density approach, and its efficiency is studied in the case of small observation noise. In those papers use is made of the manifold $S_\bullet$, which coincides with the Gaussian manifold, since h is taken to be the identity function.

Since the Itô–based and Stratonovich–based ADF are different for a general exponential family S, the equivalence with the projection filter holds only for the Stratonovich–based ADF. However, it can be shown that the Stratonovich–based and the Itô–based ADF coincide for special choices of the exponential family, such as the family S_* introduced in Section 4 above. This is another evidence for the choice of the exponential family S_* (which contains the exponential family $S_\bullet$ as a particular case).

Theorem 5.3 *For the exponential family S_*, the Itô–based assumed density filter coincides with the Stratonovich–based assumed density filter.*

PROOF. It follows from (9) that

$$\tfrac{1}{2}\, |h|^2 = \tfrac{1}{2} \sum_{k=1}^{d} |h^k|^2 = \tfrac{1}{2} \sum_{k=1}^{d} \sum_{l,l'=1}^{s} \lambda_l^k\, \lambda_{l'}^k\, c_l\, c_{l'} \ .$$

By specializing to the exponential family S_* the general equation (13) for the Stratonovich–based ADF, and using Lemma 2.2, we obtain

$$
\begin{aligned}
d\eta_t^i &= E_{\eta_t}\{\mathcal{L}_t\, c_i\}\, dt \\
&\quad -\tfrac{1}{2} \sum_{k=1}^{d} \sum_{l,l'=1}^{s} \lambda_l^k\, \lambda_{l'}^k\, [\, E_{\eta_t}\{c_l c_{l'} c_i\} - E_{\eta_t}\{c_l c_{l'}\}\, \eta_t^i\,]\, dt \\
&\quad + \sum_{k=1}^{d} \sum_{l=1}^{s} \lambda_l^k\, [\, E_{\eta_t}\{c_l c_i\} - E_{\eta_t}\{c_l\}\, \eta_t^i\,] \circ dY_t^k \\
&= E_{\eta_t}\{\mathcal{L}_t\, c_i\}\, dt - \sum_{k=1}^{d} \sum_{l,l'=1}^{s} g_{il}(\eta_t)\, \lambda_l^k\, \lambda_{l'}^k\, \eta_t^{l'}\, dt \\
&\quad - \tfrac{1}{2} \sum_{k=1}^{d} \sum_{l,l'=1}^{s} \frac{\partial}{\partial\theta_{l'}} g_{il}(\eta_t)\, \lambda_l^k\, \lambda_{l'}^k\, dt \\
&\quad + \sum_{k=1}^{d} \sum_{l=1}^{s} g_{il}(\eta_t)\, \lambda_l^k \circ dY_t^k \ ,
\end{aligned}
$$

for $i = 1, \cdots, m$. It is easily checked that the Itô–

Stratonovich transformation yields

$$g_{il}(\eta_t)\, dY_t^k = g_{il}(\eta_t) \circ dY_t^k - \tfrac{1}{2} \sum_{l'=1}^{s} \frac{\partial}{\partial \theta_{l'}} g_{il}(\eta_t)\, \lambda_{l'}^k \, dt \ ,$$

for all $k = 1, \cdots, d$ and all $i = 1, \cdots, m$. On the other hand, by specializing to the exponential family S_* the general equation (12) for the Itô–based ADF, and using Lemma 2.2 again, we obtain directly

$$\begin{aligned} d\eta_t^i &= E_{\eta_t}\{\mathcal{L}_t\, c_i\}\, dt \\ &\quad + \sum_{k=1}^{d} \sum_{l=1}^{s} \lambda_l^k \, [\, E_{\eta_t}\{c_l c_i\} - E_{\eta_t}\{c_l\}\, \eta_t^i \,] \\ &\qquad\qquad [\, dY_t^k - \sum_{l'=1}^{s} \lambda_{l'}^k \, E_{\eta_t}\{c_{l'}\}\, dt \,] \\ &= E_{\eta_t}\{\mathcal{L}_t\, c_i\}\, dt - \sum_{k=1}^{d} \sum_{l,l'=1}^{s} g_{il}(\eta_t)\, \lambda_l^k \, \lambda_{l'}^k \, \eta_t^{l'} \, dt \\ &\quad + \sum_{k=1}^{d} \sum_{l=1}^{s} g_{il}(\eta_t)\, \lambda_l^k \, dY_t^k \ , \end{aligned}$$

for $i = 1, \cdots, m$. □

6 CONCLUSION, AND DIRECTION OF FURTHER RESEARCH

Even with the choice $S = S_*$ or $S = S_\bullet$, there is still some freedom left in the choice of the dimension m of the exponential family, and in the choice of the remaining coefficients defining the exponential family. One would like to use this extra freedom to reduce as much as possible the total residual norm, which is our estimate of the local error resulting from the projection filter approximation. The first step in this direction is to provide a convenient expression for the total residual norm, and 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 $\{r_s^*\, ,\, 0 \leq s \leq t\}$.

This problem will be addressed in subsequent work.

7 REFERENCES

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

[2] D. BRIGO. New developments on the Gaussian projection filter. Accepted for publication in *Systems & Control Letters*.

[3] D. BRIGO. On the nice behaviour of the Gaussian projection filter with small observation noise. *Systems & Control Letters*, 26(5):363–370, 1995.

[4] D. BRIGO, B. HANZON, and F. LE GLAND. Approximated filtering by projection on the manifold of exponential densities. (submitted).

[5] D. BRIGO, B. HANZON, and F. LE GLAND. A differential geometric approach to nonlinear filtering : the projection filter. In *Proceedings of the 34th Conference on Decision and Control, New Orleans 1995*, pages 4006–4011. IEEE–CSS, December 1995.

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

[7] D. BRIGO, B. HANZON, and F. LE GLAND. On the relationship between assumed density filters and projection filters. Discussion Paper TI 7–96–18, Tinbergen Institute, February 1996.

[8] 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*, volume C78 of *NATO Advanced Study Institutes Series*, Dordrecht, 1981. D. Reidel.

[9] M. FUJISAKI, G. KALLIANPUR, and H. KUNITA. Stochastic differential equations for the non–linear filtering problem. *Osaka Journal of Mathematics*, 9(1):19–40, 1972.

[10] 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.

[11] H.J. KUSHNER. Approximations to optimal nonlinear filters. *IEEE Transactions on Automatic Control*, AC–12(5):546–556, 1967.

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