1991-23

ET

Faculteit der Economische Wetenschappen en Econometrie

05348

Serie Research Memoranda

NEW RESULTS ON THE PROJECTION FILTER

B. Hanzon

R. Hut

Research Memorandum 1991-23

March 1991

vrije Universiteit amsterdam

BIBL. VRIJE UNIVERSITEIT AMSTERDAM
ET.
05348.-

New Results on the Projection Filter*

B. Hanzon
Dept. Econometrics, Free University Amsterdam
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
E-mail: bhnz@sara.nl

R. Hut
Tuinstraat 8, Zaandijk, The Netherlands

Abstract

The projection filter is an approximate nonlinear filter based on orthogonal projection in the tangent space of a manifold of densities. The Riemannian metric used is the Fisher information metric. In this paper it is shown that if one uses Gaussian densities the projection filter equals a McShane-Fisk-Stratonovich version of the so-called Assumed Density Filter.

Keywords Nonlinear filtering, Fisher information metric, Assumed Density Filters, stochastic differential equations

1 Introduction

Consider a dynamical system of the following form (cf. [8,2])

$$\begin{aligned}(ITO)dx(t) &= f(x(t),t)dt + G(x(t),t)d\beta(t)\\ (ITO)dy(t) &= h(x(t),t)dt + d\eta(t)\end{aligned} \quad (1)$$

Note. We will write "(ITO)" if an Itô stochastic differential equation is meant, while we will write "(MFS)" if a stochastic differential equation is to be interpreted as being in McShane-Fisk-Stratonovich form.

In (1) the symbols have the following meaning: $x(t) \in \mathbf{R}^n$ is the state vector at time t; f is an n-vector function, $G(x,t)$ is an $n \times s$ matrix and β with $\beta \in \mathbf{R}^s$ a Brownian motion process with expectation zero and an $s \times s$ diffusion matrix $Q(t)$; $y(t) \in \mathbf{R}^k$ is the stochastic measurement process, h is a k-vector function and $\eta(t) \in \mathbf{R}^k$ a Brownian motion process with expectation zero and a $k \times k$ diffusion matrix $R(t)$,independent of $\beta(t)$.

The nonlinear filtering problem for such a system is to find the conditional probability density $p(x,t)$ of $x(t)$, given the measurements up till time t (β and η are not observed). Assume f, G, h are twice continuously differentiable. Let $A(x,t) := G(x,t)Q(t)G^T(x,t)$. As is well-known (see e.g. [8]) under certain mild conditions (for which we refer to the literature) $p(x,t)$ satisfies the Kushner-Stratonovich equation (KS equation),

$$\begin{aligned}(ITO) \qquad dp(x,t) = \mathcal{L}(p(x,t))dt +\\ p(x,t)[h(x,t) - E\{h(x(t),t)\}]^T R^{-1}(t) \times\\ [dy(t) - E\{h(x(t),t)\}]\end{aligned} \quad (2)$$

where both the caret symbol ˆ and the expectation symbol E denote the conditional expectation of the corresponding variable given the observations up till time t, and where $\mathcal{L}$ is the differential operator given by

$$\begin{aligned}\mathcal{L}p(x,t) &= -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(f_i(x,t)p(x,t)) +\\ &\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}(a_{ij}p(x,t)).\end{aligned} \quad (3)$$

The corresponding MFS equation, written with separate terms for dt and dy and without arguments of the various functions in order to shorten the formulas, is:

$$\begin{aligned}(MFS) \qquad dp =\\ [\mathcal{L}p - \frac{1}{2}(h^T R^{-1} h - E\{h^T R^{-1} h\})p]dt +\\ p(h - \hat{h})^T R^{-1} dy\end{aligned} \quad (4)$$

An alternative to the Kushner-Stratonovich equation is the Duncan-Mortenson-Zakai equation (DMZ equation) for an unnormalized version $q = q(x,t)$ of $p(x,t)$. It is given by (see e.g. [2,9]),

$$(ITO)dq = \mathcal{L}qdt + qh^T R^{-1} dy \quad (5)$$

Clearly this equation has a simpler structure then the Kushner-Stratonovich equation. The corresponding MFS equation is:

$$\begin{aligned}(MFS) \qquad dq = (\mathcal{L}q - \frac{1}{2}h^T R^{-1} hq)dt +\\ +qh^T R^{-1} dy\end{aligned} \quad (6)$$

Closed form solutions of this equation are rarely found (for a discussion see e.g. [7]). Instead many possible schemes for approximate nonlinear filters have been constructed, like the Extended Kalman Filter (EKF) or the Assumed Density Filters (ADF). In [4] one of us proposed a class of approximate nonlinear filters, since

*Paper to be presented at the ECC-91, Grenoble, July 1991

then called Projection Filters, based on the differential geometric approach to statistics (see e.g. [1]).

Consider a differential manifold of densities, e.g. the Gaussian densities. The solution of the DMZ equation will in general not remain within this manifold. However, if one projects the right-hand side of the DMZ-equation orthogonally, using the Fisher information metric, onto the tangent space of the manifold, the solution of the resulting stochastic partial differential equation will lie on the manifold - *if* one uses the MFS-form of stochastic differential equations. The resulting approximate filter is the Projection Filter. Because Fisher information is used in the projection the Projection Filter has the interpretation of being a locally most informative filter on the chosen set of densities (see Section 3).

In this paper it will be shown that if one uses the manifold of Gaussian densities, then the Projection Filter comes very close to the Assumed Density Filter mentioned before. In fact we argue that there are *two* types of ADF, one that follows from the Itô equations for the conditional first and second moments, we will call this the ITO-ADF, and one that follows from the corresponding MFS-equations, the MFS-ADF. The ITO-ADF is the classical one, that is found in the literature (see Section 2). The main result of this paper is that the Projection Filter is equal to the MFS-ADF for the Gaussian case. This provides us with a much simpler way to calculate the equations of the Projection Filter for the Gaussian case. Remarks will be made about generalizations.(See Section 4).

The paper finishes with conclusions and directions of further research.

2 Assumed Density Filters

In the literature on approximate filters one finds among many others, the so-called Assumed Density Filters (ADF); see e.g. Section 12.7 of [8]. In order to explain what we mean by an ITO-ADF and a MFS-ADF and in order to compare these with the Projection Filter, the construction of an ADF will now be presented in some detail.

Consider the (exact) conditional moment equations (cf. equations (12-108),(12-109) from [8]):

$$\begin{aligned}
(ITO)\quad & d\hat{x} = \\
& [\hat{f} - E\{(x-\hat{x})h^T\}R^{-1}\hat{h}]dt + \\
& E\{(x-\hat{x})h^T\}R^{-1}dy \\
(ITO)\quad & dP = \\
& [E\{(x-\hat{x})f^T\} + E\{f(x-\hat{x})^T\} + \\
& E\{GQG^T\} + \\
& -E\{(x-\hat{x})h^T\}R^{-1}E\{h(x-\hat{x}^T)\} + \\
& -E\{((x-\hat{x})(x-\hat{x})^T - P)h^T\}R^{-1}\hat{h}]dt + \\
& E\{((x-\hat{x})(x-\hat{x})^T - P)h^T\}R^{-1}dy \qquad (7)
\end{aligned}$$

where x, P, f, G, h, Q, R represent $x(t)$, $P(t)$, $f(x(t),t)$, $G(x(t),t)$, $h(x(t),t)$, $Q(t)$, $R(t)$ respectively. These conditional moments equations require knowledge of the entire density function, i.e. of all higher moments as well as the first two. However if we assume that the density is Gaussian, which is admittedly a *false* assumption in general, then of course the first two moments determine the conditional density and (7) forms a closed system of ITO stochastic differential equations. This set of SDE's is called the (ITO-) Assumed Density Filter (Gaussian case). This will be abbreviated to ITO-ADF. It can be generalized of course to incorporate non-Gaussian densities. For simplicity of exposition we will concentrate on the Gaussian case here. In Section 4 we will make a remark about generalization. We stress the fact that equations (7) are Itô equations; the *recipe to obtain the ADF is applied to these Itô equations* and therefore we speak about the ITO-ADF. The resulting SDE's are in Itô form, but they can also be put in MFS form, while still describing the ITO-ADF.

Example. *The cubic sensor.*

Consider the system

$$\begin{aligned}
dx &= g d\beta \\
dy &= x^3 dt + d\eta \qquad (8)
\end{aligned}$$

In other words, consider (1) with $f(x,t) = 0; G(x,t) = g(\text{constant}); h(x,t) = x^3$. Furthermore let $R = 1; Q = 1$. For this case the Itô equations for the conditional mean and covariance of the state are

$$\begin{aligned}
(ITO)\quad & d\hat{x} = [E\{x^4\} - \hat{x}E\{x^3\}][dy - E\{x^3\}dt] \\
(ITO)\quad & dP = [g^2 - (E\{(x-\hat{x})x^3\})^2]dt + \\
& E\{(x-\hat{x})^2(x^3 - E\{x^3\}\} \times \\
& [dy - E\{x^3\}dt] \qquad (9)
\end{aligned}$$

In the Gaussian case one has

$$\begin{aligned}
& E\{(x-\hat{x})^2\} = P \\
& E\{(x-\hat{x})^3\} = 0 \\
& E\{(x-\hat{x})^4\} = 3P^2 \\
& E\{(x-\hat{x})^5\} = 0 \\
& E\{(x-\hat{x})^6\} = 15P^3 \qquad (10)
\end{aligned}$$

To obtain the ITO-ADF one replaces in the r.h.s. of (9) each higher order central moment by the corresponding expression in P from (10). One obtains:

$$\begin{aligned}
(ITO)\quad & d\hat{x}_A = (-3\hat{x}_A^5 P_A - 12\hat{x}_A^3 P_A^2 - 9\hat{x}_A P_A^3)dt + \\
& +(3\hat{x}_A^2 P_A + 3P_A^2)dy \\
(ITO)\quad & dP_A = (g^2 - 15\hat{x}_A^4 P_A^2 - 36\hat{x}_A^2 P_A^3 - 9P_A^4)dt + \\
& +(6\hat{x}_A P_A^2)dy \qquad (11)
\end{aligned}$$

Putting these Itô equations in MFS form one obtains

$$\begin{aligned}
(MFS)\quad & d\hat{x}_A = (-3\hat{x}_A^5 P_A - 30\hat{x}_A^3 P_A^2 - 36\hat{x}_A P_A^3)dt \\
& +(3\hat{x}_A^2 P_A + 3P_A^2)dy \\
(MFS)\quad & dP_A = (g^2 - 15\hat{x}_A^4 P_A^2 - 81\hat{x}_A^2 P_A^3 - 18P_A^4)dt \\
& +(6\hat{x}_A P_A^2)dy \qquad (12)
\end{aligned}$$

□

The ITO-ADF is the "classical" ADF as it is found in the literature . However one can apply the same idea also to the conditional mean and covariance equations in *MFS form*. They are (cf.[6],App.3B; these equations can be derived from [2],equation (40))

$$(MFS) \quad d\hat{x} = [\hat{f} - \frac{1}{2}E\{(x-\hat{x})h^T R^{-1} h\}]dt + E\{(x-\hat{x})h^T\}R^{-1}dy$$

$$(MFS) \quad dP = [E\{(x-\hat{x})f^T\} + E\{f(x-\hat{x})^T\} + E\{GQG^T\} + -\frac{1}{2}E\{h^T R^{-1} h((x-\hat{x})(x-\hat{x}^T) - P)\}]dt + E\{((x-\hat{x})(x-\hat{x})^T - P)h\}R^{-1}dy \quad (13)$$

Just as for the corresponding Itô equations these conditional expectations equations require knowledge of the entire density function, i.e. of all higher moments as well as the first two. However if we make the assumption (which is *false* in general) that the density is Gaussian then the first two moments determine the conditional density and (13) forms a closed system of MFS stochastic differential equations. This set of SDE's is called the MFS-Assumed Density Filter (Gaussian case). This will be abbreviated to MFS-ADF. It can be generalized to incorporate non-Gaussian densities.

Example*(continued)*. The MFS-ADF for the cubic sensor is given by the equations

$$(MFS) \quad d\hat{x}_B = (-3\hat{x}_B^5 P_B - 30\hat{x}_B^3 P_B^2 - 45\hat{x}_B P_B^3)dt + (3\hat{x}_B^2 P_B + 3P_B^2)dy$$

$$(MFS) \quad dP_B = (g^2 - 15\hat{x}_B^4 P_B^2 - 90\hat{x}_B^2 P_B^3 - 45P_B^4)dt + +(6\hat{x}_B P_B^2)dy \quad (14)$$

Comparing this with (12) it follows that these filters are *unequal*. So in general the ITO-ADF and the MFS-ADF are two different filters! There are systems for which the ITO-ADF and the MFS-ADF are the same; characterization of the class of systems for which they are the same is an open research problem.

3 Projection Filters

In this section we present a short recapitulation of the definition of the Projection Filter (cf. [4]). As stated in the introduction, the Projection Filter is an approximate filter that is obtained by orthogonal projection of the r.h.s. of the Kushner-Stratonovich equation, in MFS form (or equivalently, projection of the r.h.s. of the DMZ equation in MFS form) on the tangent space of the chosen approximating manifold of densities.

The metric that is used in the orthogonal projection is the Fisher information metric, and its generalization to nonparametric (i.e. infinite dimensional) classes of densities: the Hellinger metric (cf. [5,1]). This metric is defined on the set of finite non-negative measures on a measure space $(\Omega, \mathcal{F})$ by the formula

$$d(\mu_1, \mu_2) := \frac{1}{2}\|r_1 - r_2\|_{L^2(\lambda)} = \frac{1}{2}(\int_\Omega (r_1 - r_2)^2 d\lambda)^{\frac{1}{2}} \quad (15)$$

where λ is any probability measure on $(\Omega, \mathcal{F})$ such that $\mu_1, \mu_2 \ll \lambda$, with densities $p_1 = \frac{d\mu_1}{d\lambda}$ and $p_2 = \frac{d\mu_2}{d\lambda}$ and $r_1 = \sqrt{p_1}, r_2 = \sqrt{p_2}$. (Note: in [5] we called $2d(\mu_1, \mu_2)$ the Hellinger distance). The metric space of all nonnegative measures on $(\Omega, \mathcal{F})$ with the Hellinger distance will be denoted by $\mathcal{H}$. The subspace of all nonnegative measures which are absolutely continuous w.r.t. to a given nonnegative measure λ and have a *positive density function* on the support of λ, is denoted by $\mathcal{H}(\lambda)$. From (15) it follows that it is isometric to the subset of all positive functions in the Hilbert space $L^2(\lambda)$. We will use this Hilbert space structure and we will consider $\mathcal{H}(\lambda)$ to be an infinite dimensional open submanifold of the Hilbert space $L^2(\lambda)$. Let $\mathcal{S}$ denote a smoothly embedded finite dimensional submanifold of $\mathcal{H}(\lambda)$. Then $\mathcal{H}(\lambda)$ induces a Riemannian metric on $\mathcal{S}$. The corresponding Riemannian metric tensor is equal to the Fisher information matrix (cf.e.g. [1,5]). The tangent space of $\mathcal{S}$ at some point of $\mathcal{S}$ can be regarded as a linear subspace of the tangent space of $\mathcal{H}(\lambda)$ at that point. The tangent space of $\mathcal{H}(\lambda)$ is isometric to the Hilbert space $L^2(\lambda)$. Therefore the orthogonal projection Π of the elements of the tangent space of $\mathcal{H}(\lambda)$ onto the corresponding tangent space of $\mathcal{S}$ is well defined. Consider instead of the DMZ equation its "projected version"

$$(MFS) \quad dq_\Pi = \Pi(\mathcal{L}q_\Pi - \frac{1}{2}h^T R^{-1} h q_\Pi)dt + \Pi(q_\Pi h^T R^{-1})dy \quad (16)$$

where Π applied to a vector is to be interpreted as: Π applied to each of the components of the vector.

If the initial density $q_\Pi(x, t_0)$ lies in the manifold $\mathcal{S}$ then the solution of (16) will remain in the manifold, *because (16) is in MFS form!* (cf. [3],p.123). If one would use the corresponding Itô equations this essential property would not hold; this is one of the reasons that in stochastic differential geometry the MFS form of stochastic differential equations is often preferred over the Itô form (cf.[3]).

Remarks.

(a) If one chooses $\mathcal{S}$ to be the space of all *probability* densities in $\mathcal{H}(\lambda)$ the projection filter only normalizes the densities: equation (16) is equal to the Kushner-Stratonovich equation (4) in that case! More generally speaking, if $\mathcal{S}$ consists only of probability densities then it does not matter whether the projection takes place in the DMZ equation or in the Kushner-Stratonovich equation; the same (approximate) filter equations result. In fact the normalization factor plays only a trivial role here: It does not matter for the results of the Projection Filter whether one uses normalized densities and the KS equation, or unnormalized densities and the DMZ equation, provided that normalization is then taken care of afterwards of course.

(b) Calculation of the equations of the Projection Filter can in principle take place by constructing an orthonormal basis of the tangent space of S at each point and calculating the inner product of each of the basis elements with all the functions in the r.h.s. of the DMZ equation (or the KS equation) that are to be projected. However in the next section we show an alternative and much simpler way to do the calculations for the Gaussian case (and more generally for the case of exponential densities).

(c) Because the Fisher information metric is infinitesimally equivalent to Kullback-Leibler information, one can say that within the tangent space of the manifold S the projection $\Pi[V]$ of a vector V in the tangent space of $\mathcal{H}(\lambda)$, is the vector which contains maximal information about the vector V, if V and $\Pi[V]$ are interpreted as variations in a probability density.

4 The relationship between the Projection Filter and the Assumed Density Filters

In this section we want to show that the projection filter is equal to the MFS-ADF in the Gaussian case, i.e. if the manifold of densities is taken to be the manifold of Gaussian densities on the state space. In fact the result holds more generally for exponential densities, i.e. probability densities of the form

$$p(x;\theta) = \exp(\sum_{k=1}^{d} \theta_k c_k(x) - \psi(\theta)), \tag{17}$$

where $x = (x_1, x_2, \ldots, x_n)^T \in \mathbf{R}$, $\theta = (\theta_1, \theta_2, \ldots, \theta_d)^T \in \Theta \subset \mathbf{R}^d$ and $\exp(-\psi(\theta))$ is the normalization factor and d the dimension of the parameter space Θ, i.e. of the manifold of densities. Of course the Gaussian densities belong to this class. In that case the functions $c_k(x), k = 1, 2, \ldots, d$ are the first and second degree monomials in $x_i, i = 1, 2, \ldots, n$ and $d = n + \frac{1}{2}n(n+1) = \frac{1}{2}n^2 + \frac{3}{2}n$.

The $\theta_k, k = 1, 2 \ldots, d$ are called the *natural* parameters of the exponential family. Consider

$$\eta_k = E[c_k(x)], k = 1, 2, \ldots, d \tag{18}$$

It can be shown that $\eta_1, \eta_2, \ldots, \eta_d$, which are called the *expectation* parameters, define indeed a parametrization of the manifold of exponential densities (cf. [1],p.107,108). Clearly the domain of the expectation parameters is a subset of $\mathbf{R}^d$. For example in the case of Gaussian densities, one has the restriction that $E(x - Ex)(x - Ex)^T = (E(x_i x_j) - E(x_i)E(x_j))_{ij}$ has to be positive definite. This is clearly the only restriction on the parameter vector $\eta = (\eta_1, \eta_2, \ldots, \eta_d)$ in the Gaussian case.

A crucial result from the theory of "information geometry" is that if one chooses the Fisher information metric as a Riemannian metric on manifolds of densities, then the bases of tangent vectors $\{\frac{\partial}{\partial \theta_k}\}_{k=1}^d$ and $\{\frac{\partial}{\partial \eta_k}\}_{k=1}^d$ are *biorthogonal*, i.e.

$$< \frac{\partial}{\partial \theta_k}, \frac{\partial}{\partial \eta_l} >= \delta_{kl}, \tag{19}$$

where δ_{kl} is the Kronecker delta and $<,>$ is the Riemannian inner product on the tangent space at a point of the manifold of densities (cf. [1], p.79 etc). From this we will deduce the following result on projections:

Theorem 4.1 *Let $\tilde{S}$ be a $\tilde{d}$-dimensional manifold of exponential densities that contains S as a submanifold. Consider a tangent vector T of the manifold $\tilde{S}$ and suppose*

$$T = \sum_{k=1}^{\tilde{d}} a_k \frac{\partial}{\partial \eta_k} = \sum_{k=1}^{\tilde{d}} b_k \frac{\partial}{\partial \theta_k} \tag{20}$$

at a point of $S \subset \tilde{S}$. The orthogonal projection $\Pi(T)$ of T on the tangent space of S, i.e. on the linear subspace $span\{\frac{\partial}{\partial \theta_1}, \ldots, \frac{\partial}{\partial \theta_d}\}$ of the tangent space of $\tilde{S}$ at that point, is given by the formula

$$\Pi(T) = \sum_{k=1}^{d} a_k \frac{\partial}{\partial \eta_k} + \sum_{k=d+1}^{\tilde{d}} \tilde{a}_k \frac{\partial}{\partial \eta_k} \tag{21}$$

where the $\tilde{a}_k$ are uniquely determined by the requirement $\Pi(T) \in span\{\frac{\partial}{\partial \theta_1}, \ldots, \frac{\partial}{\partial \theta_d}\}$

Corollary 4.2 *The tangent vector $\Pi(T)$ lies in the tangent space of S; in terms of the basis $\{\frac{\partial}{\partial \eta_k}\}_{k=1}^d$ of the tangent space of S at the relevant point, $\Pi(T)$ can be expressed as*

$$\Pi(T) = \sum_{k=1}^{d} a_k \frac{\partial}{\partial \eta_k}. \tag{22}$$

Proof. Because $T - \Pi(T)$ is orthogonal to $span\{\frac{\partial}{\partial \theta_1}, \ldots, \frac{\partial}{\partial \theta_d}\}$, one has

$$< \frac{\partial}{\partial \theta_k}, T - \Pi(T) >= 0, k = 1, 2, \ldots, d. \tag{23}$$

Combining this with (19) it follows that the coefficients of $\frac{\partial}{\partial \eta_k}, k = 1, 2, \ldots, d$ in a representation with respect to the basis $\{\frac{\partial}{\partial \eta_1}, \ldots, \frac{\partial}{\partial \eta_{\tilde{d}}}\}$, of T and $\Pi(T)$ must be the same.

Considerations of dimension show that the coefficients $\tilde{a}_k$ in (21) are determined uniquely by the $\{a_k\}_{k=1}^d$ and by the fact that $\Pi(T)$ lies in the space $span\{\frac{\partial}{\partial \theta_1}, \ldots, \frac{\partial}{\partial \theta_d}\}$,

q.e.d.

We can now present the main result of this paper.

Theorem 4.3 *Let S be the manifold of Gaussian densities on $\mathbf{R}^n$. The projection filter w.r.t. S is equal to the Gaussian MFS-ADF.*

Proof. Consider a Gaussian density $p(x;\theta(t))$. As remarked before a Gaussian density is an exponential density and can therefore be written in the form (17). The tangent vectors to the manifold S can be written as

$$\{\sum_{k=1}^{d} \dot{\theta}_k c_k(x) - \psi'(\theta)\dot{\theta}\}p(x;\theta) \qquad (24)$$

where $\psi'(\theta)$ denotes the Jacobian of the function $\psi(\theta)$. For a fixed value of θ let us write the expressions that multiply dy and dt in the Kushner-Stratonovich equation (4) as $z_1(x)p(x;\theta)$ and $z_2(x)p(x;\theta)$. (The functions z_1 and z_2 will depend on θ, but because θ is fixed here, one can consider them as functions of only the variable x.) Clearly $z_1(x)$ and $z_2(x)$ can in general not be written as a linear combination of $c_1(x), c_2(x), \ldots, c_d(x)$. Now for the sake of the argument consider the exponential family $\tilde{S}$ of densities of the form $\exp(\sum_{k=1}^{d} \theta_k c_k(x) + \theta_{d+1} z_1(x) + \theta_{d+2} z_2(x) - \tilde{\psi}(\theta))$, where $(\theta_{d+1}, \theta_{d+2})$ takes its values in some open set in $\mathbf{R}^2$ containing $(0,0)$, independently of $(\theta_1, \ldots, \theta_d)$. The tangent space in $(\theta_1, \ldots, \theta_d, 0, 0)$ at $\tilde{S}$ has the tangent space in $(\theta_1, \ldots, \theta_d)$ at S as a linear subspace, and clearly $z_1(x)p(x;\theta)$ and $z_2(x)p(x;\theta)$ lie in $T\tilde{S}_{(\theta_1,\ldots,\theta_d,0,0)}$.

Consider the vector of expectation parameters $\eta = (\eta_1, \ldots, \eta_d, \eta_{d+1}, \eta_{d+2})^T$ of the densities in $\tilde{S}$ and let us denote the corresponding densities by $p_E(x;\eta)$. So if η is the vector of expectation parameters of the density $p(x;\theta)$ then $p(x;\theta) = p_E(x;\eta)$. In terms of the corresponding basis $\{\frac{\partial}{\partial \eta_1} p_E(x;\eta), \ldots, \frac{\partial}{\partial \eta_{d+2}} p_E(x;\eta)\}$ of the tangent space of $\tilde{S}$ at $p_E(x;\eta)$ one can write

$$z_1(x)p_E(x;\eta) = \sum_{k=1}^{d+2} a_{1k} \frac{\partial}{\partial \eta_k} p_E(x;\eta)$$
$$z_2(x)p_E(x;\eta) = \sum_{k=1}^{d+2} a_{2k} \frac{\partial}{\partial \eta_k} p_E(x;\eta) \qquad (25)$$

According to theorem 4.1 and corollary 4.2 projection of these elements of $T\tilde{S}_{(\theta_1,\ldots,\theta_d,0,0)}$ gives as a result respectively the elements $\sum_{k=1}^{d} a_{1k} \frac{\partial}{\partial \eta_k} p_E(x;\eta)$ and $\sum_{k=1}^{d} a_{2k} \frac{\partial}{\partial \eta_k} p_E(x;\eta)$ in $TS_(\theta_1, \ldots, \theta_d)$! This implies that a projection filter update corresponds to an update of the expectation parameters $\eta_1, \ldots, \eta_d$, where of course the knowledge of the density $p_E(x;\theta)$ and the fact that this density is completely characterized by the expectation parameters $\eta_1, \ldots, \eta_d$, may be used to calculate this update. In the Gaussian case one has $d = 2$ and one only has to update the mean and covariance, using the fact that the Gaussian density $p(x;\theta) = p_E(x;\eta)$ depends completely on the mean and covariance,

q.e.d.

Remark. It is clear from the proof that the theorem can be generalized to the more general case of exponential families without much difficulty, provided that the parameter vector θ takes its values in an open subset of $\mathbf{R}^d$.

5 Conclusions and further research

The result of this paper that the Projection Filter (Gaussian case) is equal to the MFS-ADF (Gaussian case) makes the derivation of the equations of the Projection Filter much more tractable in a large number of cases. The Assumed Density Filters have the drawback that they are based on an assumption which is known to be false, which makes them logically unacceptable. This drawback is overcome for the MFS-ADF by the new interpretation of this filter as a Projection Filter. Further research is taking place in the following directions:

(i) Comparison in simulations of the Projection Filter with several other approximating filters and with the numerical solution of the Zakai equation.

(ii) Construction of "good" approximating manifolds of densities $\tilde{S}$ and derivation of the corresponding filter equations.

(iii) Construction of quantities which measure to some extend the quality of the approximation that is performed by the Projection Filter.

(iv) Projection Filters for discrete time systems. This is in fact closely related to the problem of having a non-Gaussian initial probability density.

We hope to present an extended version of the present paper which will include a simulation study and observations on some of the other items just mentioned, in the near future.

References

[1] S.-I. Amari, *Differential-Geometrical Methods in Statistics*, LNS 28, Springer Verlag, Berlin, 1985.

[2] M. H. A. Davis, S. I. Marcus, *An introduction to nonlinear filtering*, pp.53-75 in: M. Hazewinkel, J. C. Willems(eds), *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, Reidel, Dordrecht, 1981.

[3] K. D. Elworthy, *Stochastic Differential Equations on Manifolds*, Cambridge University Press, Cambridge, 1982.

[4] B. Hanzon, *A differential-geometric approach to approximate nonlinear filtering*,pp.219-224 in:

C.T.J.Dodson(ed), "Geometrization of Statistical Theory", ULDM Publications, Dept.Math., University of Lancaster, Lancaster, U.K., 1987.

[5] B. Hanzon, *Identifiability, Recursive Identification and Spaces of Linear Dynamical Systems*, CWI Tracts 63,64, CWI, Amsterdam, 1989.

[6] R. Hut, Ph.D. thesis, forthcoming.

[7] S. I. Marcus, *Algebraic and geometric methods in nonlinear filtering*, SIAM J. Control and Optimization,22,(1984),pp.817-844.

[8] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, vol. II, Academic Press, New York, 1982.

[9] E. Pardoux, *Stochastic partial differential equations and filtering of diffusion processes*, Stochastics, 2 (1979), pp. 127-167.