Nonlinear Filtering with Continuous Time Perfect Observations and Noninformative Quadratic Variation*

Marc Joannides and François LeGland
IRISA / INRIA
Campus de Beaulieu
35042 Rennes Cédex, France
e-mail : {mjoannid,legland}@irisa.fr

Abstract

We consider the problem of estimating the state of a diffusion process, based on continuous time observations in singular noise. As long as the observations are regular values of the observation function, we derive an equation for the density (w.r.t. the canonical Lebesgue measure on the corresponding level set) of the conditional probability distribution of the state, given the past observations. The proof is based on the idea of decomposition of solutions of SDE's, as introduced by Kunita.

1 Introduction

One major limiting assumption in the nonlinear filtering literature is the nondegeneracy of the observation noise covariance matrix. However, there are numerous situations of practical interest, where some perfect noise–free information is available about the unknown state, and yet the problem of state estimation with degenerate observation noise has received little attention, except in the linear case, see Kwakernaak and Sivan [8, Section 4.3.4].

An additional motivation for studying this problem, is the existing connection with various problems of state estimation for non classical dynamical systems, including : hybrid systems, i.e. systems with state constraints, stochastic differential–algebraic systems, systems with colored noise, see Korezlioglu and Runggaldier [6], systems with state–dependent observation noise, see Takeuchi and Akashi [9], etc.

Finally, a better understanding of the nonlinear filtering problem with noise–free observations should help designing robust and efficient numerical approximation schemes in the important case where the observation noise is *small.*

*This work was partially supported by the Army Research Office, under grant DAAH04–95–1–0164.

To be more specific, we consider the following state equation in $\mathbf{R}^m$

$$dX_t = b(X_t)\,dt + dW_t \ , \qquad (1)$$

where $\{W_t \, , \, t \geq 0\}$ is a Wiener process with identity covariance matrix, and the noise–free d–dimensional continuous time observations

$$z_t = h(X_t) \ .$$

The objective of nonlinear filtering is to compute the conditional probability distribution

$$\mu_t(dx) = \mathbf{P}[X_t \in dx \mid \mathcal{Z}_t] \ ,$$

where $\mathcal{Z}_t = \sigma(z_s \, , \, 0 \leq s \leq t)$ is the σ–algebra generated by the observations up to time t.

Remark 1.1 In the case where the state equation also is noise–free, i.e. reduces to an ordinary differential equation, the state estimation problem has been investigated by James [3], and numerical approximation schemes have been proposed by James and LeGland [4].

What makes this problem singular is that for any $t \geq 0$ the state X_t is known *exactly* to belong to the level set M_{z_t}, where for all $z \in \mathbf{R}^d$

$$M_z = h^{-1}(z) = \{x \in \mathbf{R}^m \, : \, h(x) = z\} \ .$$

Therefore, the conditional probability distribution μ_t is supported by the set M_{z_t}, which in general has zero Lebesgue measure, and μ_t does not have a density w.r.t. the Lebesgue measure on $\mathbf{R}^m$. The question naturally arises whether μ_t has a density w.r.t. some *canonical* measure on the level set M_{z_t}. The objective of this paper is to give conditions under which this holds, and to provide an explicit expression for the density. A similar problem with discrete time observations has been considered in Joannides and LeGland [5].

0-7803-4187-2 1645-1650

If the mapping $x \mapsto h(x)$ is injective, then X_t is known exactly from the observation of z_t. However, this can only happen if $m \leq d$, which is not the most interesting case. Notice that, in principle, the quadratic variation $\langle z \rangle_t$ of the observation process, and its time derivative

$$\frac{d}{dt}\langle z \rangle_t = h'(X_t)\,[h'(X_t)]^* ,$$

are also observed without error. Hence, if the mapping $x \mapsto (h(x), h'(x)\,[h'(x)]^*)$ is injective, then again X_t is known exactly from the observation of $(z_t, \frac{d}{dt}\langle z \rangle_t)$.

We consider here the extreme opposite case where $m \geq d$, and where for any $x \in \mathbf{R}^m$ the matrix $h'(x)\,[h'(x)]^*$ depends only on $h(x)$, i.e. the quadratic variation $\langle z \rangle_t$ would not bring any additional information, even if it would be available. We define the following $\{\mathcal{Z}_t\,,\, t \geq 0\}$–stopping time

$$\tau = \inf\{t \geq 0 \,:\, z_t \notin \mathcal{R}\} ,$$

where $\mathcal{R} \subset \mathbf{R}^d$ denotes the set of regular values of the mapping h. Notice that if $z \in \mathcal{R}$, then the level set M_z is an $(m-d)$–dimensional submanifold of $\mathbf{R}^m$, and for any $x \in M_z$ the matrix $h'(x)\,[h'(x)]^*$ is invertible.

For any $0 \leq t < \tau$, we derive an explicit expression, see equation (12) below, for the density (w.r.t. the canonical Lebesgue measure λ_{z_t} on M_{z_t}) of the conditional probability distribution μ_t. The proof is based on the idea of decomposition of solutions of SDE's, as introduced by Kunita in [7].

2 Differentiating the observations

Recall that our model is

$$\begin{aligned} dX_t &= b(X_t)\,dt + dW_t\ , \qquad X_0 \sim p_0(x)\,dx \\ z_t &= h(X_t) \end{aligned} \tag{2}$$

where $\{W_t\,,\, t \geq 0\}$ is a Wiener process with identity covariance matrix, and where the initial condition X_0 has an absolutely continuous probability distribution $p_0(x)\,dx$ w.r.t. the Lebesgue measure on $\mathbf{R}^m$, with a continuous density p_0. The first step is to write the observation equation in differential form, and the model (2) can be written equivalently as

$$\begin{aligned} dX_t &= b(X_t)\,dt + dW_t\ , \qquad X_0 \sim p_0(x)\,dx \\ dz_t &= L\,h(X_t)\,dt + h'(X_t)\,dW_t\ , \\ z_0 &= h(X_0)\ , \end{aligned} \tag{3}$$

where L is the backward second order partial differential operator associated with equation (1), i.e.

$$L = \sum_{i=1}^{m} b^i(\cdot)\,\frac{\partial}{\partial x_i} + \tfrac{1}{2}\sum_{i,j=1}^{m} \frac{\partial^2}{\partial x_i \partial x_j}\ .$$

Notice that the prior distribution of the r.v. X_0 is $p_0(x)\,dx$. The posterior distribution satisfies

$$\mu_0(dx) = \mathbf{P}[X_0 \in dx \mid \mathcal{Z}_0] = \mu_{z_0}(dx)\ ,$$

where for any $z \in \mathbf{R}^d$

$$\mu_z(dx) = \mathbf{P}[X_0 \in dx \mid h(X_0) = z]\ ,$$

by definition. If z_0 is a regular value, i.e. if $\tau > 0$, then M_{z_0} is a $(m-d)$–dimensional submanifold of $\mathbf{R}^m$, with canonical Lebesgue measure λ_{z_0}. Moreover, if the density p_0 is continuous, and does not vanish identically on M_{z_0}, then

$$\mu_{z_0}(dx) = c\,\frac{p_0(x)}{J_h(x)}\,\lambda_{z_0}(dx)$$

see Joannides and LeGland [5], i.e. the posterior distribution is absolutely continuous w.r.t. λ_{z_0}, where c is a normalization constant, and where J_h denotes the Jacobian determinant of the mapping h, defined by

$$J_h(x) = \sqrt{\det\left(h'(x)\,[h'(x)]^*\right)}\ ,$$

for any $x \in \mathbf{R}^m$, see Evans and Gariepy [1, Section 3.2].

For any $t \geq s$, let $\mathcal{Z}_t^s = \sigma(z_r - z_s\,,\, s \leq r \leq t)$ denote the σ–algebra generated by the increments of the observation between time s and time t, so that $\mathcal{Z}_t = \mathcal{Z}_s \vee \mathcal{Z}_t^s$.

Proposition 2.1 *For any $t \geq 0$, and any test function f defined on $\mathbf{R}^m$*

$$\langle \mu_t, f \rangle = \mathbf{E}[f(X_t) \mid \mathcal{Z}_t] = \mathbf{E}_{\mu_0}[f(X_t) \mid \mathcal{Z}_t^0]\ ,$$

where

$$\mu_0(dx) = \mathbf{P}[X_0 \in dx \mid \mathcal{Z}_0]\ .$$

As a result, the model (3) can be written equivalently as

$$\begin{aligned} dX_t &= b(X_t)\,dt + dW_t\ , \qquad X_0 \sim \mu_0(dx) \\ dz_t &= L\,h(X_t)\,dt + h'(X_t)\,dW_t\ , \end{aligned} \tag{4}$$

where $\{W_t\,,\, t \geq 0\}$ is a Wiener process with identity covariance matrix, and where the initial condition X_0 has the probability distribution $\mu_0(dx)$ supported by M_{z_0}. Notice that this is a *non–standard* nonlinear filtering model, since the covariance matrix of the observation noise depends on the state.

3 State–dependent observation noise, and main assumption

For the model (4) obtained at the end of the previous section, the following decomposition of the observation has been shown by Takeuchi and Akashi [9]. On the time interval $0 \leq t < \tau$, the $d \times d$ matrix $r(X_t) = h'(X_t)\ [h'(X_t)]^*$ is invertible, and we define the two processes

$$\bar{z}_t = \int_0^t r^{-1/2}(X_s)\, dz_s \ ,$$

$$\langle z \rangle_t = z_t\, z_t^* - (\int_0^t z_s\, dz_s^*) - (\int_0^t z_s\, dz_s^*)^* \ .$$

Notice that $\mathcal{R}_t^0 \triangleq \sigma(\langle z \rangle_s - \langle z \rangle_0\, ,\, 0 \leq s \leq t) \subset \mathcal{Z}_t^0$ for any $t \geq 0$. It holds

$$d\bar{z}_t = r^{-1/2}(X_t)\, L\, h(X_t)\, dt + d\bar{V}_t \ ,$$

where

$$d\bar{V}_t = r^{-1/2}(X_t)\, h'(X_t)\, dW_t \ ,$$

hence $\{\bar{V}_t\, ,\, 0 \leq t < \tau\}$ is a d-dimensional Wiener process with unit covariance matrix, and

$$\frac{d}{dt} \langle z \rangle_t = r(X_t) \ .$$

The observation z_t is decomposed into a noisy component $\bar{z}_t$ with additive nonsingular Gaussian white noise, and a noise–free component $\langle z \rangle_t$. Takeuchi and Akashi have shown in [9, Theorem 1] the following equivalence : $\mathcal{Z}_t^0 = \bar{\mathcal{Z}}_t^0 \vee \mathcal{R}_t^0$, where $\bar{\mathcal{Z}}_t^0 = \sigma(\bar{z}_s - \bar{z}_0\, ,\, 0 \leq s \leq t)$ denotes the σ-algebra associated with the increments of the noisy component of the observation.

This illustrate the connexion, mentioned in the Introduction, between models with noise–free observations, and models with state–dependent observation noise. However, this connexion is rather in the form of a circling argument : the original model (2) had noise–free observations, and was transformed, after time–differentiation, into the model (4) with state–dependent observation noise. This new model was in turn transformed into the following model

$$\begin{aligned} dX_t &= b(X_t)\, dt + dW_t \ , \qquad X_0 \sim \mu_0(dx) \\ d\bar{z}_t &= r(X_t)^{-1/2}\, L\, h(X_t)\, dt + d\bar{V}_t \ , \\ \frac{d}{dt} \langle z \rangle_t &= r(X_t) \ , \end{aligned} \tag{5}$$

with noise–free observations. In principle, either the mapping $x \mapsto (h(x), r(x))$ is injective, or we should differentiate the noise–free component of the observation, etc.

To avoid these difficulties, and because it is questionnable whether the quadratic variation process $\{\langle z \rangle_t\, ,\, t \geq 0\}$ can be computed in practice, we make the following assumption, which will hold throughout the end of the paper, and which states that the quadratic variation derivative $\frac{d}{dt} \langle z \rangle_t$, even if it could be computed, would not bring any additional information, that was not already available from the original observation z_t.

Assumption A : For any $x \in \mathbf{R}^m$

$$r(x) = h'(x)\ [h'(x)]^* = R(h(x)) \ ,$$

for some measurable function $R(\cdot)$ defined on $\mathbf{R}^d$, and taking values in the space of symmetric positive $d \times d$ matrices.

Example 3.1 [distance measurements] The Assumption A is satisfied in the case where $m \geq d = 1$, and where $h(x) = |x|^2$ for all $x \in \mathbf{R}^m$. Indeed $h'(x) = 2\, x^*$ hence $r(x) = 4\, |x|^2 = 4\, h(x)$ for all $x \in \mathbf{R}^m$.

4 The Zakai equation

Under the Assumption A, $r(X_t) = h'(X_t)\ [h'(X_t)]^* = R(z_t)$, for any $t \geq 0$. In particular, the *noise* process

$$V_t = \int_0^t h'(X_s)\, dW_s \ , \tag{6}$$

in the differentiated observation equation is a d–dimensional continuous martingale, with quadratic variation

$$\langle V \rangle_t = \int_0^t R(z_s)\, ds \ ,$$

depending on the observation only, and quadratic covariation

$$\langle V, W \rangle_t = \int_0^t h'(X_s)\, ds \ .$$

The Jacobian determinant $J_h(x) = \sqrt{\det R(z_0)}$ is independent of $x \in M_{z_0}$, hence the conditional probability distribution at time 0 reduces to

$$\mu_0(dx) = c\, p_0(x)\, \lambda_{z_0}(dx) \ ,$$

where c is another normalization constant.

As a result, the model (3) can be written equivalently, under the Assumption A, as

$$\begin{aligned} dX_t &= b(X_t)\, dt + dW_t \ , \qquad X_0 \sim \mu_0(dx) \\ dz_t &= L\, h(X_t)\, dt + dV_t \ , \end{aligned} \tag{7}$$

where $\{W_t\, ,\, t \geq 0\}$ is a Wiener process with identity covariance matrix, where the observation noise $\{V_t\, ,\, t \geq 0\}$ is a d-dimensional continuous martingale, with quadratic variation depending on the observation only, and where the initial condition X_0 has the probability distribution $\mu_0(dx)$ supported by M_{z_0}. Notice that this is now a standard nonlinear filtering model, with correlated noise.

We introduce the following notations : for any $x \in \mathbf{R}^m$

$$r(x) = h'(x)\ [h'(x)]^* \ ,$$

and for any regular point $x \in M_Z$ with $z \in \mathcal{R}$

$$\rho(x) = [h'(x)]^* \, r^{-1}(x) \ ,$$

$$\pi(x) = [h'(x)]^* \, r^{-1}(x) \, h'(x) = \rho(x) \, h'(x) \ .$$

Notice that, for any regular point x the rows of the $d \times m$ Jacobian matrix $h'(x)$ generate the normal space $N_x M_Z$ to the manifold M_Z at point x, and for any $v \in \mathbf{R}^m$ the vector $\pi(x) v$ is the orthogonal projection of v on the linear space $N_x M_Z$. Notice also that $h'(x) \rho(x) = I$ for any regular point x, where I denotes the $d \times d$ identity matrix. For any $0 \leq t < \tau$, we can decompose the noise process in the state equation as

$$W_t = W_t^{\|} + W_t^{\perp} \ ,$$

with

$$W_t^{\perp} = \int_0^t \pi(X_s) \, dW_s = \int_0^t \rho(X_s) \, dV_s \ ,$$

and we notice that

$$\langle W^{\perp} \rangle_t = \langle W^{\perp}, W \rangle_t = \int_0^t \pi(X_s) \, ds \ ,$$

hence $\langle W^{\perp}, W^{\|} \rangle_t = 0$ and $\langle V, W^{\|} \rangle_t = 0$.

The model (7) can be written equivalently as

$$\begin{aligned} dX_t \;&=\; b(X_t)\, dt + \rho(X_t)\, dV_t + [I - \pi(X_t)]\, dW_t \ , \\ &\qquad\qquad X_0 \sim \mu_0(dx) \qquad\qquad (8) \\ dz_t \;&=\; L\, h(X_t)\, dt + dV_t \ , \end{aligned}$$

where $\{W_t \, , \, t \geq 0\}$ is a Wiener process with identity covariance matrix, where $\{V_t \, , \, t \geq 0\}$ is a continuous martingale defined by (6), and where the initial condition X_0 has the probability distribution $\mu_0(dx)$ supported by M_{z_0}. The standard approach to this problem would be to use the reference probability method, and to compute

$$\langle \mu_t, f \rangle = \frac{\langle \sigma_t, f \rangle}{\langle \sigma_t, \mathbf{1} \rangle} \ ,$$

for any test function f defined on $\mathbf{R}^m$, where the unnormalized conditional probability distribution $\sigma_t(dx)$ satisfies the Zakai equation

$$d\sigma_t = L^* \, \sigma_t \, dt + \sum_{k,l=1}^{d} B_k^* \, \sigma_t \, R^{k,l}(z_t) \, dz_t^l \ , \qquad (9)$$

in weak form, with first order partial differential operators

$$B_k = L\, h_k + \sum_{i=1}^{m} \rho_k^i \, \frac{\partial}{\partial x_i} \ ,$$

for any $k = 1, \cdots, d$, and with initial condition $\sigma_0(dx) = \mu_0(dx)$ supported by M_{z_0}. In (9), the coefficients $R^{k,l}(z_t)$ for $k, l = 1, \cdots, d$ are the entries of the $d \times d$ matrix $R^{-1}(z_t)$.

By definition, for any $0 \leq t < \tau$ the unnormalized conditional probability distribution $\sigma_t(dx)$ has to be supported by M_{z_t}. However, the equation (9) does not give any insight about this fact. The purpose of the next section is to give a more geometric picture, using the idea of decomposition of solutions of SDE's.

5 Decomposition of stochastic flows

Let $\{\phi_{t,s}(\cdot) \, , \, 0 \leq s \leq t\}$ denote the stochastic flow of diffeomorphisms (solution map) associated with the following stochastic differential equation

$$\begin{aligned} dX_t \;&=\; b(X_t)\, dt + \rho(X_t)\, [dz_t - L\, h(X_t)\, dt] \\ &\qquad + [I - \pi(X_t)]\, dW_t \ , \end{aligned} \qquad (10)$$

obtained from the model (8). In particular

$$\langle \mu_t, f \rangle = \mathbf{E}_{\mu_0}[f(X_t) \mid \mathcal{Z}_t^0] = \mathbf{E}_{\mu_0}[f \circ \phi_{t,0}(X_0) \mid \mathcal{Z}_t^0] \ ,$$

for any test function f defined on $\mathbf{R}^m$.

Proposition 5.1 *On the time interval $0 \leq t < \tau$, the solution of (10) satisfies*

$$\begin{aligned} dX_t \;&=\; [I - \pi(X_t)] \circ [b(X_t)\, dt + dW_t] \\ &\qquad + \rho(X_t) \circ dz_t \ . \end{aligned} \qquad (11)$$

From this equation, and using the Stratonovich chain rule, we have

$$dh(X_t) = h'(X_t) \circ dX_t = dz_t \ ,$$

hence, integrating from s to t

$$h[\phi_{t,s}(x)] - h(x) = z_t - z_s \ ,$$

hence $\phi_{t,s}(x) \in M_{z_t}$ if $x \in M_{z_s}$, i.e. the diffeomorphism $\phi_{t,s}(\cdot)$ maps M_{z_s} to M_{z_t}.

Let $\{\xi_{t,s}(\cdot) \, , \, 0 \leq s \leq t\}$ denote the stochastic flow of diffeomorphisms associated with the following SDE

$$d\xi_t = \rho(\xi_t) \circ dz_t \ .$$

The diffeomorphism $\xi_{t,s}(\cdot)$ is measurable w.r.t. $\mathcal{Z}_t^s$, and can be computed directly from the observations. Notice that the same remark as above holds, i.e.

$$dh(\xi_t) = h'(\xi_t) \circ d\xi_t = dz_t \ ,$$

hence $\xi_{t,s}(x) \in M_{z_t}$ if $x \in M_{z_s}$, i.e. the diffeomorphism $\xi_{t,s}(\cdot)$ maps M_{z_s} to M_{z_t}.

The purpose of the decomposition method is to find another stochastic flow of diffeomorphisms $\{\kappa_{t,s}(\cdot) \, , \, 0 \leq s \leq t\}$ such that the identity

$$\phi_{t,s} = \xi_{t,s} \circ \kappa_{t,s}$$

holds for any $0 \leq s \leq t$. Necessarily

$$\kappa_{t,s} = \xi_{t,s}^{-1} \circ \phi_{t,s} \ ,$$

hence $\kappa_{t,s}(x) \in M_{Z_s}$ if $x \in M_{Z_s}$, i.e. the diffeomorphism $\kappa_{t,s}(\cdot)$ maps M_{Z_s} to itself. This is the situation depicted in Figure 1.

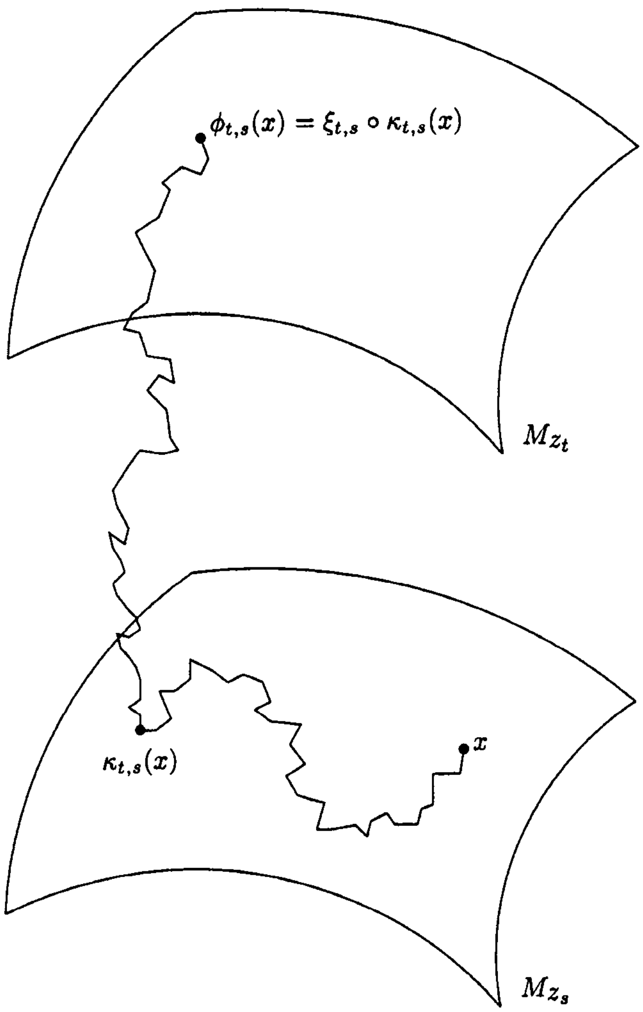

Figure 1: Decomposition of flows

A more explicit description of $\{\kappa_{t,s}(\cdot)\,,\, 0 \leq s \leq t\}$ is provided by the decomposition theorem of Kunita [7]. Rewrite (11) as

$$\begin{aligned} dX_t &= [I - \pi(X_t)] \circ [b(X_t)\, dt + dW_t] + \rho(X_t) \circ dz_t \\ &= \rho(X_t) \circ dz_t + g_0(X_t)\, dt + \sum_{j=1}^{d} g_j(X_t) \circ dW_t^j \ , \end{aligned}$$

where the vector fields are defined, for any regular point $x \in M_Z$ with $z \in \mathcal{R}$, by

$$g_0(x) = [I - \pi(x)]\, b(x) = b_{||}(x) \ ,$$

and for any $j = 1, \cdots, m$

$$g_j(x) = [I - \pi(x)]\, e_j = e_j - \pi_j(x) \ ,$$

where e_j denotes the j–th vector basis in $\mathbf{R}^m$, and $\pi_j(x)$ denotes the j-th column of the $m \times m$ matrix $\pi(x)$. Notice that, for any regular point $x \in \mathbf{R}^m$, and any $j = 0, 1, \cdots, m$

$$h'(x)\, g_j(x) = 0 \ ,$$

hence $g_j(x) \in T_x\, M_Z$ if $x \in M_Z$, i.e. the vector field g_j is tangent to M_Z.

Definition 5.2 *For any vector field g defined on $\mathbf{R}^m$, the random vector field $(\xi_{t,s}^{-1})_*\, g$ is defined by*

$$(\xi_{t,s}^{-1})_*\, g(x) = [\,\xi_{t,s}'(x)\,]^{-1}\, g[\xi_{t,s}(x)] \ ,$$

for any regular point $x \in \mathbf{R}^m$.

Notice that

(i) If $x \in M_{Z_s}$, then $g_j[\xi_{t,s}(x)] \in T_{\xi_{t,s}(x)}\, M_{Z_t}$, for any $j = 0, 1, \cdots, m$.

(ii) If $x \in M_{Z_s}$, then $\xi_{t,s}'(x)$ maps linearly $T_x\, M_{Z_s}$ to $T_{\xi_{t,s}(x)}\, M_{Z_t}$. Indeed, let $x \in M_{Z_s}$, and $v \in T_x\, M_{Z_s}$, and consider a continuously differentiable curve $\theta \mapsto \gamma(\theta)$ on M_{Z_s}, such that $\gamma(0) = x$ and $\dot\gamma(0) = v$. Then $\theta \mapsto \xi_{t,s}[\gamma(\theta)]$ defines a continuously differentiable curve on M_{Z_t}, and

$$\frac{d}{d\theta}\, \xi_{t,s}[\gamma(\theta)] = \xi_{t,s}'[\gamma(\theta)]\, \dot\gamma(\theta) \in T_{\xi_{t,s}(x)}\, M_{Z_t} \ .$$

This holds in particular for $\theta = 0$, which proves that $\xi_{t,s}'(x)\, v \in T_{\xi_{t,s}(x)}\, M_{Z_t}$.

The combination of (i) and (ii) yields

$$(\xi_{t,s}^{-1})_*\, g_j(x) \in T_x\, M_{Z_s} \ ,$$

for any $x \in M_{Z_s}$ and any $j = 0, 1, \cdots, m$, i.e. the random vector field $(\xi_{t,s}^{-1})_*\, g_j$ is tangent to M_{Z_s}.

According to Kunita [7], the stochastic flow of diffeomorphisms $\{\kappa_{t,s}(\cdot)\,,\, 0 \leq s \leq t\}$ satisfies the following SDE

$$\begin{aligned} d\kappa_{t,s}(x) &= (\xi_{t,s}^{-1})_*\, g_0[\kappa_{t,s}(x)]\, dt \\ &\quad + \sum_{j=1}^{m} (\xi_{t,s}^{-1})_*\, g_j[\kappa_{t,s}(x)] \circ dW_t^j \ . \end{aligned}$$

The coefficients are random tangent vector fields on M_{Z_s}, adapted to the filtration $\{\mathcal{Z}_t^s\,,\, t \geq s\}$, and the above equation is a SDE on the manifold M_{Z_s}.

6 Solution as reduced order filtering + transport

Making use of the decomposition $\phi_{t,0} = \xi_{t,0} \circ \kappa_{t,0}$, we have

$$\langle \mu_t , f \rangle = \mathbf{E}_{\mu_0}[f \circ \phi_{t,0}(X_0) \mid \mathcal{Z}_t^0] = \mathbf{E}_{\mu_0}[f \circ \xi_{t,0}(\kappa_t) \mid \mathcal{Z}_t^0] ,$$

for any test–function f defined on $\mathbf{R}^m$, with $\kappa_t = \kappa_{t,0}(X_0)$ by definition. Since $\xi_{t,0}$ is measurable w.r.t. $\mathcal{Z}_t^0$, the problem reduces to the computation of the conditional probability distribution of κ_t given $\mathcal{Z}_t^0$, in the following *reduced order* model

$$d\kappa_t \;=\; (\xi_{t,0}^{-1})_* \, g_0(\kappa_t)\, dt + \sum_{j=1}^m (\xi_{t,0}^{-1})_* \, g_j(\kappa_t) \circ dW_t^j \; ,$$

$$\kappa_0 \sim \mu_0(dx)$$

$$dz_t \;=\; L\, h \circ \xi_{t,0}(\kappa_t)\, dt + dV_t \; ,$$

where $\{W_t \, , \, t \geq 0\}$ is a Wiener process with identity covariance matrix, where $\{V_t \, , \, t \geq 0\}$ is a continuous martingale defined by (6), and where the initial condition $\kappa_0 = X_0$ has the probability distribution $\mu_0(dx)$ supported by M_{z_0}.

According to Gyöngy [2], the conditional probability distribution of κ_t given $\mathcal{Z}_t^0$ is absolutely continuous

$$\mathbf{P}_{\mu_0}[\kappa_t \in dx \mid \mathcal{Z}_t^0] = q_t(x)\, \lambda_{z_0}(dx) \; ,$$

and the density q_t satisfies a stochastic partial differential equation on the $(m-d)$–dimensional submanifold M_{z_0} of $\mathbf{R}^m$, hence

$$\begin{aligned} \langle \mu_t , f \rangle \;&=\; \int_{M_{z_0}} f \circ \xi_{t,0}(x)\; \mathbf{P}_{\mu_0}[\kappa_t \in dx \mid \mathcal{Z}_t^0] \\ &=\; \int_{M_{z_0}} f \circ \xi_{t,0}(x)\; q_t(x)\, \lambda_{z_0}(dx) \; . \end{aligned}$$

The image measure $\xi_{t,0}^{-1} \lambda_{z_0}$ of the canonical measure λ_{z_0} on M_{z_0} under the mapping $\xi_{t,0}(\cdot)$ is absolutely continuous w.r.t. the canonical measure λ_{z_t} on M_{z_t}, with density

$$c_{t,0}(x) = \sqrt{\det\left([\xi'_{t,0}(x)]^* \, \xi'_{t,0}(x)\right)} \; ,$$

where the linear mapping $[\xi'_{t,0}(x)]^* \, \xi'_{t,0}(x)$ from $T_x M_{z_0}$ to itself is invertible, hence the change of variable $x \mapsto \xi_{t,0}(x)$ gives

$$\begin{aligned} \langle \mu_t , f \rangle \;&=\; \int_{M_{z_t}} f(x)\; q_t \circ \xi_{t,0}^{-1}(x)\; \xi_{t,0}^{-1} \lambda_{z_0}(dx) \\ &=\; \int_{M_{z_t}} f(x)\; q_t \circ \xi_{t,0}^{-1}(x)\, c_{t,0}(x)\, \lambda_{z_t}(dx) \; . \end{aligned}$$

This proves that μ_t is absolutely continuous w.r.t. the canonical measure λ_{z_t} on M_{z_t}, and

$$\mu_t(dx) = q_t \circ \xi_{t,0}^{-1}(x)\, c_{t,0}(x)\, \lambda_{z_t}(dx) \; . \qquad (12)$$

Remark 6.1 From the computational point of view, we have to solve a SPDE on a $(m-d)$–dimensional submanifold of $\mathbf{R}^m$ with an absolutely continuous solution. This is much more efficient than solving equation (9), which is a SPDE on $\mathbf{R}^m$ with a singular solution.

Acknowledgment

The authors gratefully acknowledge Dan Ocone for his suggestion that the diffeomorphisms between the different level sets should allow to reduce the problem to a single level set.

References

[1] L.C. EVANS and R.F. GARIEPY. *Measure Theory and Fine Properties of Functions.* CRC Press, Boca Raton, FL, 1992.

[2] I. GYÖNGY. Filtering on manifolds. *Acta Applicandae Mathematicae*, 35(1–2):165–177, 1994.

[3] M.R. JAMES. Finite time observer design by probabilistic–variational methods. *SIAM Journal on Control and Optimization*, 29(4):954–967, 1991.

[4] M.R. JAMES and F. LE GLAND. Numerical approximation for nonlinear filtering and finite–time observers. In I. Karatzas and D. Ocone, editors, *Applied Stochastic Analysis, Rutgers University 1991*, volume 177 of *Lecture Notes in Control and Information Sciences*, pages 159–175. Springer Verlag, Berlin, 1992.

[5] M. JOANNIDES and F. LE GLAND. Nonlinear filtering with perfect discrete time observations. In *Proceedings of the 34th Conference on Decision and Control, New Orleans 1995*, pages 4012–4017. IEEE–CSS, December 1995.

[6] H. KOREZLIOGLU and W.J. RUNGGALDIER. Filtering for nonlinear systems driven by nonwhite noises : An approximation scheme. *Stochastics and Stochastics Reports*, 44:65–102, 1993.

[7] H. KUNITA. On the decomposition of solutions of stochastic differential equations. In D. Williams, editor, *Stochastic Integrals, Durham 1980*, volume 851 of *Lecture Notes in Mathematics*, pages 213–255. Springer Verlag, Berlin, 1981.

[8] H. KWAKERNAAK and R. SIVAN. *Linear Optimal Control Systems.* John Wiley & Sons, New York, 1972.

[9] Y. TAKEUCHI and H. AKASHI. Least–squares state estimation of systems with state–dependent observation noise. *Automatica*, 21(3):303–313, 1985.