Electronic Journal of Probability

Vol. 14 (2009), Paper no. 63, pages 1863–1883.

Journal URL
http://www.math.washington.edu/~ejpecp/

Nonlinear filtering with signal dependent observation noise

Dan Crisan[*], Michael Kouritzin[†] and Jie Xiong[‡]

Abstract

The paper studies the filtering problem for a non-classical framework: we assume that the observation equation is driven by a signal dependent noise. We show that the support of the conditional distribution of the signal is on the corresponding level set of the derivative of the quadratic variation process. Depending on the intrinsic dimension of the noise, we distinguish two cases: In the first case, the conditional distribution has discrete support and we deduce an explicit representation for the conditional distribution. In the second case, the filtering problem is equivalent to a classical one defined on a manifold and we deduce the evolution equation of the conditional distribution. The results are applied to the filtering problem where the observation noise is an Ornstein-Uhlenbeck process.

Key words: Nonlinear Filtering, Ornstein Uhlenbeck Noise, Signal-Dependent Noise.

AMS 2000 Subject Classification: Primary 60G35; Secondary: 60G30.

Submitted to EJP on October 9, 2007, final version accepted August 10, 2009.

[*]Department of Mathematics, Imperial College London, 180 Queen's Gate, London SW7 2BZ, UK (d.crisan@imperial.ac.uk)

[†]Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, T6G 2G1, Canada (mkouritz@math.ualberta.ca). Research supported by NSERC.

[‡]Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA, and Department of Mathematics, Hebei Normal University, Shijiazhuang 050016, PRC (jxiong@math.utk.edu). The research of this author is supported by NSF DMS-0906907.

1 Introduction

Let $(\Omega, \mathscr{F}, P)$ be a probability space on which we have defined a homogeneous Markov process X. We can obtain information on X by observing an associated process y which is a function of X plus random noise n:

$$y_t = h(X_t) + n_t. \tag{1.1}$$

In most of the literature n_t is modeled by white noise, which does not exist in the ordinary sense, but rather as a distribution of a generalized derivative of a Brownian motion. That is, the process W defined formally as

$$W_t = \int_0^t n_s ds, \quad t \geq 0,$$

is a Brownian motion and the (integrated) observation model turns into one of the (classical) form

$$\begin{aligned} Y_t &= \int_0^t y_s ds \\ &= \int_0^t h(X_s) ds + W_t, \end{aligned}$$

where Y is the accumulated observation. In this case, the observation σ-field is $\mathscr{F}_t^Y = \sigma^{\text{Aug}}\{Y_s,\ s \leq t\}$ and the desired conditional distribution

$$\pi_t(\cdot) \doteq P(X_t \in \cdot | \mathscr{F}_t^Y),\ t \geq 0,$$

satisfies the Kushner-Stratonovich or Fujisaki-Kallianpur-Kunita equation

$$d\pi_t(f) = \pi_t(Lf)dt + (\pi_t(fh^T) - \pi_t(f)\pi_t(h^T))d\hat{v}_t, \tag{1.2}$$

where L is the generator of X and $\hat{v}$, defined as

$$\hat{v}_t = Y_t - \int_0^t \pi_s(h) ds,\ t \geq 0,$$

is a Brownian motion called the innovation process.

Balakrishnan (see Kallianpur and Karandikar [8], p. 3) states that this (integrated) approach is not suitable for application since the results obtained cannot be instrumented. Kunita [9], Mandal and Mandrekar [11] and Gawarecki and Mandrekar [4] studied the model (1.1) when n is a general Gaussian process. The most important example is the case when n is an Ornstein-Uhlenbeck process given by

$$dO_t = -\beta O_t dt + \beta dW_t, \tag{1.3}$$

and W is a standard m-dimensional Wiener process. The observation model becomes

$$y_t = h(X_t) + O_t,\ t \geq 0 \tag{1.4}$$

or the integral form

$$Y_t = \int_0^t h(X_s) ds + \int_0^t O_s ds \ \ t \geq 0.$$

To fix the ideas, we let the signal X be a $\mathbb{R}^d$-valued process given by

$$dX_t = b(X_t)dt + c(X_t)dB_t, \ \ t \geq 0$$

where b, c are $\mathbb{R}^d$, respectively $\mathbb{R}^{d\times d}$-valued functions and B is a standard d-dimensional Wiener process independent of W. The optimal filter π is given by

$$\pi_t f = \mathbb{E}(f(X_t)|\mathscr{F}_t^Y).$$

In the aforementioned papers, a Kallianpur-Striebel formula is given, the filtering equation for π is derived and it is proved that π converges to the solution of the (classical) FKK equation as β tends to infinity. However, the conditions imposed in these papers [9], [11], [4] are very restrictive. Most notably, the authors assume that the map $t \mapsto h(X_t)$ is differentiable. To remove this restrictive condition, Bhatt and Karandikar [1] consider the variant observation model

$$y_t = \alpha \int_{(t-\alpha^{-1})\vee 0}^{t} h(X_s)ds + O_t$$

for $\alpha > 0$ and obtain the same results for this modified model.

In this paper, we deal with the original model (1.4) but no longer assume differentiability of the map $t \mapsto h(X_t)$. As we will see, this will lead to an observation model with signal dependent observation noise

$$dY_t = h(X_t)dt + \sigma(X_t)dW_t.$$

In turn, the filtering problem with signal dependent observation noise will be converted into a classical one (with signal independent observation noise) via a suitably chosen stochastic flow mapping. This article is organized as follows: In Section 2 we set the filtering problem with signal dependent observation noise and the framework for transforming this singular filtering problem into the classical one. Then, we discuss this transformation in two cases. In Section 3 we consider the case when the signal dimension is small so (under mild regularity) level set $M_z = \{x \in \mathbb{R}^d : \ \sigma(x) = z\}$ is discrete for each positive definite matrix z. In Section 4 we study the transformation when M_z are manifolds. In this case, we decompose the vectors in $\mathbb{R}^d$ into their tangent and normal components, and study the signal according to this decomposition. In Section 5, we convert the filtering with OU-noise to a special case of the general singular filtering model.

The methods and results presented here benefited a lot from the work of Ioannides and LeGland (see [6] and [7]). In particular, in [7], they study the filtering problem with perfect observations. That is, in their set-up, the observation process Y is a *deterministic* function of the signal X. Here we show that the filtering problem that we are interested in can be reduced to one where the observation process has two components: one that is perfect (in the language of Ioannides and LeGland) and one that is of the classical form (see Lemma 2.3 below).

2 A general singular filtering model

Motivated by the filtering problem with OU-process as noise, we will consider a general filtering problem with signal and observation given by

$$\begin{cases} dX_t = b(X_t)dt + c(X_t)dW_t + \tilde{c}(X_t)dB_t \\ dY_t = h(X_t)dt + \sigma(X_t)dW_t \end{cases} \tag{2.1}$$

where B and W are two independent Brownian motions in $\mathbb{R}^d$ and $\mathbb{R}^m$ respectively, b, c, $\tilde{c}$, h, σ are functions defined on $\mathbb{R}^d$ with values in $\mathbb{R}^d$, $\mathbb{R}^{d\times m}$, $\mathbb{R}^{d\times d}$, $\mathbb{R}^m$, $\mathbb{R}^{m\times m}$, respectively. We will assume that $Y_0 = 0$ and that X_0 is independent of B and W. We will denote the law of X_0 by π_0 and assume that it is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$ and will denote by $\tilde{\pi}_0$ its density.

We will also assume that $\sigma(x)$ is a symmetric for each x and positive definite matrix for each x. If not, we can define the m-dimensional process $\tilde{W}$

$$d\tilde{W}_t = \sqrt{\sigma\sigma^T(X_t)}^{-1}\sigma(X_t)dW_t,\ t \geq 0.$$

Then the pair process $(\tilde{W}, B)$ is a standard $m+d$-dimensional Brownian motion and the system (2.1) is equivalent to the following

$$\begin{cases} dX_t = b(X_t)dt + c(X_t)d\tilde{W}_t + \tilde{c}(X_t)dB_t \\ dY_t = h(X_t)dt + \sqrt{\sigma\sigma^T(X_t)}d\tilde{W}_t. \end{cases}$$

The analysis can be easily extended to cover the case when all terms in (2.1) depend on both X and Y. We make the following assumptions throughout the rest of the paper.

Condition (BC): The functions b, c, $\tilde{c}$ are Lipschitz continuous and h is bounded and measurable.

Condition (ND): For any $x \in \mathbb{R}$, the $m \times m$-matrix $\sigma(x)$ in invertible.

Condition (S): The partial derivatives of the function σ up to order 2 are continuous.

Condition (X0): The law of X_0 has a continuous density $\tilde{\pi}_0$ with respect to the Lebesgue measure on $\mathbb{R}^d$.

Let $\langle Y\rangle$ be the quadratic covariation process of the m-dimensional semimartingale Y. From (2.1) it follows that

$$\langle Y\rangle_t = \int_0^t \sigma^2\left(X_t\right)ds.$$

Hence, the process $\sigma^2(X_t) = \frac{d}{dt}\langle Y\rangle_t$ is $\mathscr{F}_t^Y$-measurable for any $t > 0$.

Remark 2.1. *$\sigma^2(X_0)$ may not be $\mathscr{F}_0^Y$-measurable, but is $\mathscr{F}_{0+}^Y$-measurable.*

We decompose the observation σ-field $\mathscr{F}_t^Y$ into two parts: One is generated by $\sigma(X_t)$ and another by an observation process of the classical form.

Denote by $\mathscr{S}_m^+$ the set of symmetric positive definite $m \times m$-matrices. Then for $x \in \mathbb{R}^d$, we have $\sigma(x)$ and $\sigma^2(x) \in \mathscr{S}_m^+$. Next, we define the mapping a from $\mathscr{S}_m^+$ to $\mathbb{R}^{\tilde{m}}$ with $\tilde{m} = \frac{m(m+1)}{2}$ as the list of the diagonal entries and those above the diagonal in lexicographical order, i.e., for any $r \in \mathscr{S}_m^+$, $a(r)$ is defined as

$$\begin{aligned} &a(r)^1 = r^{11},\ a(r)^2 = r^{12},\ \ldots,\ a(r)^m = r^{1m}, \\ &a(r)^{m+1} = r^{22},\ \ldots,\ a(r)^{2m-1} = r^{2m}, \\ &a(r)^{2m} = r^{33},\ \ldots,\ a(r)^{\tilde{m}} = r^{mm}. \end{aligned}$$

It is clear that a is one-to-one from $\mathscr{S}_m^+$ onto $a(\mathscr{S}_m^+) \subset \mathbb{R}^{\tilde{m}}$. Further, a and a^{-1} are continuous.

Lemma 2.2. *$a(\mathscr{S}_m^+)$ is open in $\mathbb{R}^{\tilde{m}}$.*

Proof: Suppose $\sigma \in \mathbb{R}^{m\times m}$ is symmetric. Then, $\sigma \in \mathscr{S}_m^+$ if and only if $det(\sigma_k) > 0$, $k = 1, 2, \cdots, m$, where σ_k is the $k \times k$ sub-matrix obtained from σ by removing the last $m-k$ rows and $m-k$ columns. Note that $det(\sigma_k)$ is a polynomial of the entries in σ. Thus the image $a(\mathscr{S}_m^+)$ consists of points in $\mathbb{R}^{\tilde{m}}$ such that these polynomials of its coordinates are positive. This implies that $a(\mathscr{S}_m^+)$ is open. ∎

Now let s be the square root mapping $s : \mathscr{S}_m^+ \to \mathscr{S}_m^+$ such that for any $r \in \mathscr{S}_m^+$, $s(r)$ is the unique matrix belonging to $\mathscr{S}_m^+$ such that $s(r)^2 = r$. It is easy to see that s is a continuous and, in particular, a Borel measurable mapping. Hence, since $\sigma^2(X_t)$ is $\mathscr{F}_t^Y$-measurable for any $t > 0$, then

$$Z_t = \sigma(X_t) = s(\sigma^2(X_t)),\ t > 0$$

is $\mathscr{F}_t^Y$-measurable for any $t > 0$, too. We have the following

Lemma 2.3. *Let $\hat{Y}$ be the stochastic process defined by*

$$d\hat{Y}_t = \sigma^{-1}(X_t)dY_t = \tilde{h}(X_t)dt + dW_t,$$

where $\tilde{h} = \sigma^{-1}h$. Then,

$$\mathscr{F}_t^Y = \mathscr{F}_t^{\hat{Y}} \vee \mathscr{F}_t^Z,\ t > 0.$$

Proof: From the above $\mathscr{F}_t^Z \subset \mathscr{F}_t^Y$. Also $\sigma^{-1}(X_t)$ is $\mathscr{F}_t^Z$-measurable, hence $\hat{Y}_t$ is $\mathscr{F}_t^Y$-measurable, i.e., $\mathscr{F}_t^{\hat{Y}} \subset \mathscr{F}_t^Y$. So

$$\mathscr{F}_t^{\hat{Y}} \vee \mathscr{F}_t^Z \subset \mathscr{F}_t^Y.$$

On the other hand, as $dY_t = Z_t d\hat{Y}_t$, we have indeed,

$$\mathscr{F}_t^Y \subset \mathscr{F}_t^{\hat{Y}} \vee \mathscr{F}_t^Z.$$

∎

By Lemma 2.3, Z can be considered as part of the observations and the signal-observation pair can be written as

$$\begin{cases} dX_t = b(X_t)dt + c(X_t)dW_t + \tilde{c}(X_t)dB_t \\ dY_t = h(X_t)dt + Z_t dW_t \\ Z_t = \sigma(X_t) \end{cases}. \tag{2.2}$$

We see now that the framework is truly non-classical as part of the observation process is noiseless. It follows that, given the observation, X_t takes values in the level set M_{Z_t} as defined in the introduction. Hence, π_t has support on M_{Z_t}. Therefore, π_t will not have total support (unless σ is constant) and will be singular with respect to the Lebesgue measure on $\mathbb{R}^d$.

As was seen, only the diagonal entries and those above the diagonal of the process Z_t (in other words, $a(Z_t)$) are required to generate $\mathscr{F}_t^Z$. Hence, we only need to take into account the properties of the mapping a_σ

$$a_\sigma : \mathbb{R}^d \longrightarrow \mathbb{R}^{\tilde{m}},$$

defined as $a_\sigma(x) = a(\sigma(x))$ for all $x \in \mathbb{R}^d$. Then, defining the usual gradient operator for $\mathbb{R}^{\tilde{m}}$-vectors f by

$$\left(\nabla f\right)(x)^{ij} = \partial_j f^i(x); \quad i = 1,...,\tilde{m}; \quad j = 1,...,d; \quad x \in \mathbb{R}^d;$$

$q(x) = \nabla a_\sigma(x)$ is a linear mapping from $\mathbb{R}^d$ to $\mathbb{R}^{\tilde{m}}$, i.e., an $\tilde{m} \times d$ matrix.

Definition 2.4. *The vector $x \in \mathbb{R}^d$ is a regular point for a_σ if and only if the matrix $q(x)$ has full rank* $\min(d, \tilde{m})$*. We shall denote the collection of all regular points by $\mathscr{R}$.*

We will study the optimal filter π_t in the next two sections according to the type of the level set M_z.

3 The case $d \leq \tilde{m}$

In this section, we consider the case when $d \leq \tilde{m}$. We will assume that M_z consists of countably many points and that its connected components do not branch or coalesce. Namely, we assume

Condition (R1): There exists a countable set I such that for any $z \in \mathbb{R}^{\tilde{m}}$,

$$M_z = \{x^i(z) : \ i \in I\},$$

where $x^i : \mathbb{R}^{\tilde{m}} \to \mathbb{R}^d$, $i \in I$ are continuous functions such that $x^i(z) \neq x^j(z)$, for any $i, j \in I$, $i \neq j$.

This condition holds true if a_σ satisfies the following assumption

Condition (R): Every point in $\mathbb{R}^d$ is regular for a_σ, in other words, $\mathscr{R} = \mathbb{R}^d$.

In this case, for $x \in \mathbb{R}^d$, by the inverse function theorem (see, for example, Rudin [12]) there is a continuous bijection between an open neighborhood of $w(a_\sigma(x))$ and an open neighborhood of x, where $w : \mathbb{R}^{\tilde{m}} \to \mathbb{R}^d$ is the projection of $\mathbb{R}^{\tilde{m}}$ onto $\mathbb{R}^d$ corresponding to those coordinates which give a minor of maximal rank for the matrix $q(x)$. The composition between this continuous bijection and the projection w is, in effect, one of the continuous functions x^i appearing in Condition (R1). In particular, if $x \in M_z$, then there is no other element of M_z within the corresponding open neighborhood. In other words, M_z contains only isolated points. Thus, Condition (R) implies that all the level sets of σ (respectively, a_σ) are *discrete* and must be finite on any compact set and therefore countable overall. Hence there is a countable set of continuous functions describing the level sets. These continuous functions do not coalesce or branch as that would contradict the existence of the bijection at the point of coalescence/branching (in topological language, the number of connected components of the level sets is locally constant). Hence condition (R) implies Condition (R1).

By Condition (R1) and the continuity of the process X_t, we see that if X_0 is non-random and $X_0 = x^i(Z_0)$ for some $i \in I$, then $X_t = x^i(Z_t)$ for the same value of $i \in I$ (the process does not 'jump' between connected components. In other words, the noiseless component of the observation uniquely identifies the conditional distribution of the signal given the observation:

$$\pi_t = \delta_{X_t}, \quad t > 0.$$

Next, we consider the case that X_0 is not constant. We need to take into account the additional information that may arise from observing the quadratic variation of the process $a\left(Z_t\right)$ and the covariation process between $a(Z_t)$ and Y_t. This will not influence the trajectory of π: From above we already know that it is deterministic given its initial value π_0 and the process $a\left(Z_t\right)$. However, this additional information may restrict π_0.

Applying Itô's formula to $a_\sigma(X_t)$ for X_t being given by (2.1), we get

$$da\left(Z_t\right) = La_\sigma(X_t)dt + q\left(X_t\right)\left(c(X_t)dW_t + \tilde{c}(X_t)dB_t\right), \tag{3.1}$$

where L is the second order differential operator given by

$$Lf = \frac{1}{2}\sum_{i,j=1}^{d} \tilde{a}_{ij}\partial^2_{ij} f + \sum_{i=1}^{d} b_i \partial_i f$$

with $\tilde{a} = cc^T + \tilde{c}\tilde{c}^T$. Thus, the quadratic covariation process of $a\left(Z_t\right)$ is

$$\int_0^t q(X_s)\left(cc^T(X_s) + \tilde{c}\tilde{c}^T\left(X_s\right)\right) q^T\left(X_s\right) ds.$$

It follows from (3.1) and (2.1) that the quadratic covariation process between $a(Z_t)$ and Y_t is

$$\int_0^t qc(X_s) Z_s^T ds.$$

Therefore,

$$\tilde{Z}_0 \equiv \left(q\left(cc^T + \tilde{c}\tilde{c}^T\right) q^T\left(X_0\right), qc(X_0)Z_0^T\right) \tag{3.2}$$

is $\mathscr{F}^Y_{0+}$ measurable.

The analysis that follows hinges upon an explicit characterization of the information contained in $\mathscr{F}^Y_{0+}$. Such a characterization may not be available in general. However, we present two cases where it is possible and note that other cases may be deduced in an inductive manner. From Remark 2.1 we know that $\sigma^2(X_0)$, which is the derivative of the quadratic variation of the process Y at 0, is $\mathscr{F}^Y_{0+}$-measurable. Moreover, from the discussion following Lemma 2.2 $\sigma(X_0)$ and $a_\sigma(X_0)$ are also $\mathscr{F}^Y_{0+}$-measurable. In Case 1, no more additional information is available from $\tilde{Z}_0$. This means that the $\mathscr{F}^Y_{0+}$-measurable random variables obtained by differentiating at zero the quadratic variation of $a_\sigma(X)$ and the quadratic covariation process between $a_\sigma(X)$ and Y are functions of $a_\sigma(X_0)$. In Case 2, these two new processes offer new information (they are not functions of $a_\sigma(X_0)$), but their corresponding quadratic variations and quadratic covariation processes do not have informative derivatives. In subsequent cases, which can be treated in a similar manner as the first two cases, more and more of the processes constructed by computing quadratic variations and quadratic covariations and differentiating at zero offer information (they are not functions of the already computed $\mathscr{F}^Y_{0+}$-measurable random variables).

Case 1: The matrices $q\left(cc^T + \tilde{c}\tilde{c}^T\right)q^T$ and qc are functions of a_σ, in other words there exist two Borel measurable functions H_1 and H_2 from $a(\mathscr{S}^+_m)$ to $\mathbb{R}^{\tilde{m}\times\tilde{m}}$ and $\mathbb{R}^{\tilde{m}\times m}$ respectively such that

$$q\left(cc^T + \tilde{c}\tilde{c}^T\right)q^T = H_1\left(a_\sigma\right) \text{ and } qc = H_2(a_\sigma). \tag{3.3}$$

Let us note that $x \in \mathbb{R}^d$ is a regular point if and only if

$$J(x) = \sqrt{\det q^T q(x)} > 0.$$

Theorem 3.1. *Suppose that the conditions (R1, S, BC, X0) and (3.3) are satisfied and $x^i(Z_0) \in \mathscr{R}$ for each $i \in I$. If $\sum_{x \in M_{Z_0}} \frac{\tilde{\pi}_0(x)}{J(x)} < \infty$, then*

$$\pi_t = \sum_{x \in M_{Z_0}} p_x \delta_{\hat{X}_t^x}, \ \ t > 0 \text{ and } \pi_0 = \tilde{\pi}_0,$$

where $\hat{X}_t^x = x^i(Z_t)$ with $i \in I$ is such that $x = x^i(Z_0)$, and

$$p_x = \frac{\frac{\tilde{\pi}_0(x)}{J(x)}}{\sum_{x \in M_{Z_0}} \frac{\tilde{\pi}_0(x)}{J(x)}}.$$

Proof: We consider the filter $\hat{\pi}_t = P\left(X_t \in \cdot \middle| \mathscr{F}_t^{\hat{Y}} \vee \mathscr{F}_t^Z\right)$. Note that $\pi_0 = \tilde{\pi}_0$, the law of X_0, and $\pi_t = \hat{\pi}_t$ for $t > 0$ by Lemma 2.3.

From (3.2,3.3), we get that $\tilde{Z}_0 = \left(H_1\left(a\left(Z_0\right)\right), H_2\left(a\left(Z_0\right)\right)\right)$ brings no new knowledge, hence we can ignore it. Following the proof of Theorem 2.8 in [6], we let μ_z be the conditional probability distribution of X_0 given $Z_0 = z$, i.e.,

$$\mu_z(dx) = P\left(X_0 \in dx | Z_0 = z\right).$$

For any $B \in \mathscr{B}(\mathbb{R}^{\tilde{m}})$ and $\phi \in L^1(\mathbb{R}^d, \tilde{\pi}_0(x)dx)$, let

$$g(x) = \phi(x) 1_{a_\sigma(x) \in B} \frac{\tilde{\pi}_0(x)}{J(x)}.$$

By the area formula (cf. Evans and Gariepy [2], p. 99, Theorem 2), we have

$$\begin{aligned}
\mathbb{E}[\phi(X_0) 1_{a_\sigma(X_0) \in B}] &= \int_{\mathbb{R}^d} \phi(x) 1_{a_\sigma(x) \in B} \tilde{\pi}_0(x) dx \\
&= \int_{\mathbb{R}^d} g(x) J(x) dx \\
&= \int_{\mathbb{R}^{\tilde{m}}} \sum_{x \in M_{a^{-1}(u)}} g(x) \mathscr{H}^d(du) \\
&= \int_B \sum_{x \in M_{a^{-1}(u)}} \phi(x) \frac{\tilde{\pi}_0(x)}{J(x)} \mathscr{H}^d(du), \qquad (3.4)
\end{aligned}$$

where $\mathscr{H}^d$ is the d-dimensional Haussdorff measure (cf. Evans and Gariepy [2], p60). Taking $\phi = 1$, we get

$$P(a_\sigma(X_0) \in B) = \int_B \sum_{x \in M_{a^{-1}(u)}} \frac{\tilde{\pi}_0(x)}{J(x)} \mathscr{H}^d(du).$$

Thus

$$\begin{aligned}\mathbb{E}[\phi(X_0)1_{a_\sigma(X_0)\in B}] &= \mathbb{E}[1_{a_\sigma(X_0)\in B}\int \phi(x)\mu_{Z_0}(dx)] \\ &= \int_B\int \phi(x)\mu_{a^{-1}(u)}(dx)\sum_{x'\in M_{a^{-1}(u)}}\frac{\tilde{\pi}_0(x')}{J(x')}\mathscr{H}^d(du).\end{aligned}$$

Comparing with (3.4), we get

$$\int_B\int \phi(x)\mu_{a^{-1}(u)}(dx)\sum_{x'\in M_{a^{-1}(u)}}\frac{\tilde{\pi}_0(x')}{J(x')}\mathscr{H}^d(du)=\int_B\sum_{x\in M_{a^{-1}(u)}}\frac{\tilde{\pi}_0(x)}{J(x)}\phi(x)\mathscr{H}^d(du).$$

Therefore,

$$\int \phi(x)\mu_z(dx)\sum_{x'\in M_z}\frac{\tilde{\pi}_0(x')}{J(x')}=\sum_{x\in M_z}\frac{\tilde{\pi}_0(x)}{J(x)}\phi(x).$$

Hence μ_z has the support on the set M_z and

$$\mu_z=\sum_{x\in M_z}p_x\delta_x.$$

Following the case with constant X_0, we then have

$$\pi_t(f)=\mathbb{E}_{\mu_{Z_0}}(f(X_t)|\mathscr{F}_t^Y)=\sum_{x\in M_z}p_xf(X_t^x),\qquad \forall t\geq 0.$$

■

Before we consider the second case, we give an example for which the conditions of Theorem 3.1 are satisfied.

Example 3.2. *Let $d=1$ and $m=2$. The coefficients of the system are $b(x)=x$, $c(x)=0$, $\tilde{c}=1$, $h(x)$ bounded, and*

$$\sigma(x)=\begin{pmatrix}\sigma_{11}(x) & \sigma_{12}(x)\\ \sigma_{21}(x) & \sigma_{22}(x)\end{pmatrix}=\begin{pmatrix}2+\sin x & 0\\ 0 & 2+\cos x\end{pmatrix}.$$

Then,

$$M_z=\{x^i(z):\ i\in\mathbb{Z}\},$$

where $\mathbb{Z}$ denotes the collection of integers and the continuous functions $x^i:\mathbb{R}^3\to\mathbb{R}$ are defined as

$$x^i(z)=x^i(z_1,z_2,z_3)=\begin{cases}2\pi i+\arcsin(z_1-2) & z_3\geq 2\\ 2\pi i+\pi+\arcsin(2-z_1) & z_3<2\end{cases}.$$

This proves Condition (R1). Condition (3.3) is also satisfied as

$$q(x)=\begin{pmatrix}\cos x\\ 0\\ -\sin x\end{pmatrix}=\begin{pmatrix}\sigma_{22}(x)-2\\ 0\\ 2-\sigma_{11}(x)\end{pmatrix}.$$

The other conditions are easy to verify.

Now we consider the other case.

Case 2: If $q\left(cc^T+\tilde{c}\tilde{c}^T\right)q^T$ and qc are not functions of a_σ, then $\tilde{Z}_0$ as defined in (3.2) offers new information, that is $\mathscr{F}^Y_{0+}$ is larger than the σ-field generated by Z_0. Hence the support of the distribution of X_0 given $\mathscr{F}^Y_{0+}$ may be smaller than given $\sigma(Z_0)$. To handle the new information, we need to impose additional constraints. Let $\sigma_0=\sigma$, $q_0=q$ and σ_k be the following matrix valued function

$$\sigma_k(x)=\begin{pmatrix} \sigma_{k-1}(x) & c^Tq^T_{k-1}(x) \\ q_{k-1}c(x) & q_{k-1}\left(cc^T+\tilde{c}\tilde{c}^T\right)q^T_{k-1}(x)\end{pmatrix},\ \ x\in\mathbb{R}^d,\ k=1,2,3...,\tag{3.5}$$

$\tilde{m}_k$ be the dimension of the image of the mapping a_{σ_k} and $q_k=\nabla a_{\sigma_k}$. We replace Conditions (S), (R1) and (3.3) with the following:

Condition ($\tilde{S}$): The partial derivatives of σ_1 up to order 2 are continuous.

Condition ($\tilde{R}$1): There exists a countable set $\tilde{I}$ such that for any $z\in\mathbb{R}^{\tilde{m}_1}$,

$$M_z\equiv\{x\in\mathbb{R}^d:\ \sigma_1(x)=z\}=\{x^i(z):\ i\in\tilde{I}\},$$

where $x^i:\mathbb{R}^{\tilde{m}_1}\to\mathbb{R}^d$, $i\in\tilde{I}$ are continuous functions such that $x^i(z)\neq x^j(z)$, for any $i,j\in\tilde{I}$, $i\neq j$.

Condition (IN$_k$): q_kc and $q_k\left(cc^T+\tilde{c}\tilde{c}^T\right)q_k^T$ are functions of a_{σ_k}.

Then, the following analogue of Theorem 3.1 holds true.

Theorem 3.3. *Suppose that the conditions ($\tilde{R}$1, $\tilde{S}$, BC, X0, IN$_1$) are satisfied. If $\sum_{x\in M_{\tilde{Z}_0}}\frac{\tilde{\pi}_0(x)}{J(x)}<\infty$, then*

$$\pi_t=\sum_{x\in M_{\tilde{Z}_0}}p_x\delta_{\hat{X}^x_t},\ \ t>0,$$

where $\hat{X}^x_t=x^i(Z_t)$ with $i\in\tilde{I}$ is such that $x=x^i(Z_0)$, and

$$p_x=\frac{\frac{\tilde{\pi}_0(x)}{J(x)}}{\sum_{x\in M_{\tilde{Z}_0}}\frac{\tilde{\pi}_0(x)}{J(x)}}.$$

We now give an example for which the conditions of Theorem 3.3 are satisfied.

Example 3.4. *Let $d=m=1$. The coefficients of the system are $b(x)=x$, $c(x)=1$, $\tilde{c}=1$, $h(x)$ bounded, and $\sigma(x)=2+\sin x$. Then, $q(x)=\cos x$ is not a function of σ. However, q' is a function of σ as*

$$q'(x)=-\sin x=2-\sigma(x).$$

In this case $\tilde{m}_1=3$,

$$\sigma_1(x)=\begin{pmatrix} 2+\sin x & \cos x \\ \cos x & 2(\cos x)^2\end{pmatrix},\ \ x\in\mathbb{R},$$

$$q_1(x) = \begin{pmatrix} \cos x \\ -\sin x \\ 4\cos x \sin x \end{pmatrix}, \quad x \in \mathbb{R}$$

is a function of a_{σ_1} so (IN$_1$) holds and the level sets M_z are described by

$$M_z = \{x^i(z) : \ i \in \mathbb{Z}\},$$

where $\mathbb{Z}$ denotes the collection of integers and the continuous functions $x^i : \mathbb{R}^3 \to \mathbb{R}$ are defined as

$$x^i(z) = x^i(z_1, z_2, z_3) = \begin{cases} 2\pi i + \arcsin(z_1 - 2) & z_2 \geq 0 \\ 2\pi i + \pi + \arcsin(2 - z_1) & z_2 < 0 \end{cases}.$$

This proves Condition ($\tilde{R}1$). The other conditions are easy to verify.

The analysis can continue in this manner: if (IN$_1$) is not satisfied by q_1, we can define σ_2 in a similar manner with σ, q in (3.5) replaced by σ_1 and q_1 respectively; and the above procedure is then continued until Condition (IN$_k$) is satisfied.

4 The case $d > \tilde{m}$

In this section, we consider the case when $d > \tilde{m}$. We will show that M_z is no longer a discrete set but rather a surface (manifold) and the optimal filter π_t is a probability measure on the manifold M_z and is absolutely continuous with respect to the surface measure. For this, we follow closely the analysis in [7].

Note that, if $d > \tilde{m}$, then $x \in \mathbb{R}^d$ is a regular point if and only if

$$J(x) = \sqrt{\det qq^T(x)} > 0.$$

We shall use $T_x M_z$ to denote the space consisting of all the tangent vectors to the manifold (surface) M_z at point $x \in M_z$. $T_x M_z$ is called the tangent space of M_z at x. We will use $N_x M_z$ to denote the orthogonal complement of $T_x M_z$ in $\mathbb{R}^d$. We call $N_x M_z$ the normal space of M_z at x. We list in the following lemma some well-known facts about the transformation a_σ without giving their proofs:

Lemma 4.1. *i) For any $u \in \mathbb{R}^{\tilde{m}}$, $M_{a^{-1}(u)}$ is a $\tilde{d}$-dimensional manifold where $\tilde{d} = d - \tilde{m}$. The Haussdorff measure $\mathscr{H}^{\tilde{d}}$ on $M_{a^{-1}(u)}$ is the surface measure.*
ii) For any $x \in M_z$, the rows of the matrix q generate the normal space $N_x M_z$ to the manifold M_z at point x.
iii) Let

$$\rho(x) = q^T(qq^T)^{-1}(x) \text{ and } p(x) = \rho(x)q(x).$$

Then $p(x)$ is the orthogonal projection matrix from $\mathbb{R}^{\tilde{m}}$ to the subspace $N_x M_z$.

For simplicity of the presentation, we make an assumption which is slightly stronger than (3.3).

Condition (IN): There exist two Borel measurable functions H_1 and H_2 from $a(S_m^+)$ to $\mathbb{R}^{\tilde{m}\times m}$ and $\mathbb{R}^{\tilde{m}\times d}$ respectively such that

$$qc = H_1(a_\sigma) \text{ and } q\tilde{c} = H_2(a_\sigma). \tag{4.1}$$

To demonstrate this condition and the lemma above, we give an example.

Example 4.2. *Let $d = 2$ and $m = 1$. Define the coefficients by $b(x) = x$, $c(x) = 0$, $\tilde{c}(x) = I$ and $\sigma(x) = e^{x_1}$. Then $J(x) = e^{x_1} > 0$ for all $x \in \mathbb{R}^2$ and hence, Condition (R) is satisfied. Further,*

$$M_z = \{(x_1, x_2) : \ e^{x_1} = z\}$$

is a (line) manifold of dimension $\tilde{d} = 2 - 1 = 1$. It is clear that the row vector of q generates the normal space

$$N_x M_z = \{\lambda(1,0) : \ \lambda \in \mathbb{R}\}.$$

In this example, $p(x) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ is the orthogonal projection matrix from $\mathbb{R}^2$ to $N_x M_z$. To verify (IN), we note that $qc = 0$ and

$$q\tilde{c} = (e^{x_1}, 0)I = (\sigma(x), 0).$$

We also assume:
Condition (C): Both c and $\tilde{c}$ are continuously differentiable.

Throughout the rest of this section, the assumptions (R, S, BC, X0, IN, ND, C) will be in force. The following theorem gives us the conditional distribution of X_0 given Z_0. Since, in this case, $\mathscr{F}^Y_{0+}$ coincides with the σ-field generated by Z_0, it gives, in effect, π_{0+} in the case of $d > \tilde{m}$.

We introduce λ_u to be the surface measure on the level set $M_{a^{-1}(u)}$ for $u \in \mathbb{R}^{\tilde{m}}$ and μ_z to be the conditional probability distribution of X_0 given $Z_0 = z$, i.e.,

$$\mu_z(dx) = P\left(X_0 \in dx | Z_0 = z\right).$$

The following theorem shows that μ_z is absolutely continuous with respect to λ_u.

Theorem 4.3. *Suppose that the density $\tilde{\pi}_0$ is not identically zero on $M_{a^{-1}(u)}$ and satisfies the following integrability condition:*

$$\int_{M_{a^{-1}(u)}} \frac{\tilde{\pi}_0(x)}{J(x)} \lambda_u(dx) < \infty.$$

Then

$$\mu_z(dx) = p(x)\lambda_{a(z)}(dx),$$

where

$$p(x) = \frac{\tilde{\pi}_0(x)/J(x)}{\int_{M_z} \tilde{\pi}_0(y)/J(y)\lambda_u(dy)}.$$

Proof: For any test function ϕ defined on $\mathbb{R}^d$, and any Borel set D in $\mathbb{R}^{m\times m}$, define

$$g(x) = \phi(x) 1_{\sigma(x)\in D} \frac{\tilde{\pi}_0(x)}{J(x)}.$$

Then, by the co-area formula (cf. Evans and Gariepy [2], chapter 3), we have

$$\begin{aligned}
\mathbb{E}\left[\phi(X_0)1_{\sigma(X_0)\in D}\right] &= \int_{\mathbb{R}^d} g(x)J(x)dx \\
&= \int_{\mathbb{R}^{\tilde{m}}}\int_{M_{a^{-1}(u)}} g(x)\mathscr{H}^{d-\tilde{m}}(dx)du \\
&= \int_D\int_{M_{a^{-1}(u)}} \phi(x)\frac{\tilde{\pi}_0(x)}{J(x)}\mathscr{H}^{d-\tilde{m}}(dx)du \\
&= \int_D\int_{M_{a^{-1}(u)}} \phi(x)\frac{\tilde{\pi}_0(x)}{J(x)}\lambda_u(dx)du.
\end{aligned}$$

By taking $\phi(x)\equiv 1$, we get

$$P(\sigma(X_0)\in D) = \int_D\int_{M_{a^{-1}(u)}} \frac{\tilde{\pi}_0(x)}{J(x)}\lambda_u(dx)du.$$

The result follows from the definition of the conditional expectation. ∎

We now decompose the vector fields in the SDE satisfied by the signal according to their components in the spaces T_xM_z and N_xM_z. It is more convenient to use Stratonovich form for the signal process. That is, the signal X satisfies the following SDE in Stratonovich form:

$$dX_t = \tilde{b}(X_t)dt + c(X_t)\circ dW_t + \tilde{c}(X_t)\circ dB_t, \tag{4.2}$$

where for $i = 1,2,\cdots,d$, the ith component of $\tilde{b}$ is

$$\tilde{b}_i = b_i - \frac{1}{2}\sum_{k=1}^{d}\sum_{j=1}^{m}(\partial_k c_{ij})c_{kj} - \frac{1}{2}\sum_{k,j=1}^{k}(\partial_k\tilde{c}_{ij})\tilde{c}_{kj}.$$

Recall from (3.1) that the $\mathscr{F}_t^Y$-measurable process $a(Z_t)$ satisfies

$$da(Z_t) = La_\sigma(X_t)dt + dV_t, \tag{4.3}$$

where V_t is the $\tilde{m}$-dimensional continuous martingale

$$V_t = \int_0^t qc(X_s)dW_s + q\tilde{c}(X_s)dB_s, \quad t\geq 0$$

with quadratic covariation process

$$\langle V\rangle_t = \int_0^t H(a(Z_s))ds,$$

and $H = H_1H_1^T + H_2H_2^T$, where $H_1 = qc$ and $H_2 = q\tilde{c}$.

Lemma 4.4. *The second observation process of (2.2) $a(Z_t)$ satisfies*

$$da(Z_t) = qc(X_t) \circ dW_t + q\tilde{c}(X_t) \circ dB_t + q\tilde{b}(X_t)dt. \tag{4.4}$$

Proof: We have

$$\begin{aligned} dV_t &= qc(X_t) \circ dW_t + q\tilde{c}(X_t) \circ dB_t - \frac{1}{2}\sum_{k,j=1}^{d}\sum_{\ell=1}^{m} \partial_k(q_{\cdot j}c_{j\ell})c_{k\ell}(X_t)dt \\ &\quad -\frac{1}{2}\sum_{k,j,\ell=1}^{d} \partial_k(q_{\cdot j}\tilde{c}_{j\ell})\tilde{c}_{k\ell}(X_t)dt \\ &= qc(X_t) \circ dW_t + q\tilde{c}(X_t) \circ dB_t - \frac{1}{2}\sum_{k,j=1}^{d}(cc^T)_{jk}\partial^2_{kj}a_\sigma dt - \frac{1}{2}\sum_{k=1}^{d} q\partial_k cc^T_{\cdot k}dt \\ &\quad -\frac{1}{2}\sum_{k,j=1}^{d}(\tilde{c}\tilde{c}^T)_{jk}\partial^2_{kj}a_\sigma dt - \frac{1}{2}\sum_{k=1}^{d} q\partial_k \tilde{c}\tilde{c}^T_{\cdot k}dt \\ &= qc(X_t) \circ dW_t + q\tilde{c}(X_t) \circ dB_t - La_\sigma(X_t)dt + q\tilde{b}(X_t)dt. \end{aligned}$$

Combining with (4.3), we see that (4.4) holds. ∎

Finally, we arrive at the main decomposition result.

Theorem 4.5. *Under Condition (BC, S, R, X_0, C), the filtering model (2.1) is rewritten as*

$$dX_t = (I-p)\left(\tilde{b}(X_t)dt + c(X_t) \circ dW_t + \tilde{c}(X_t) \circ dB_t\right) + \rho(X_t) \circ da(Z_t), \tag{4.5}$$

with observations

$$da(Z_t) = La_\sigma(X_t)dt + dV_t, \tag{4.6}$$

and

$$dY_t = h(X_t)dt + Z_t dW_t. \tag{4.7}$$

Proof: From equation (4.2) we have

$$\begin{aligned} dX_t &= \tilde{b}(X_t)dt + c(X_t) \circ dW_t + \tilde{c}(X_t) \circ dB_t \\ &= (I-p)\tilde{b}(X_t)dt + (I-p)c(X_t) \circ dW_t + (I-p)\tilde{c}(X_t) \circ dB_t \\ &\quad + p\tilde{b}(X_t)dt + pc(X_t) \circ dW_t + p\tilde{c}(X_t) \circ dB_t. \end{aligned} \tag{4.8}$$

By (4.4), we get

$$\rho(X_t) \circ da(Z_t) = p\tilde{b}(X_t)dt + pc(X_t) \circ dW_t + p\tilde{c}(X_t) \circ dB_t.$$

Plugging back into (4.8), we see that (4.5) holds. The equality (4.6) is a rewriting of (4.3). The equation (4.7) is just the original observation model (2.1). ∎

Let $\{\xi_{t,s} : 0 \le s \le t\}$ be the stochastic flow associated with the SDE:

$$d\xi_t = \rho(\xi_t) \circ da(Z_t). \tag{4.9}$$

Lemma 4.6. *The flow $\xi_{t,s}$ maps M_{Z_s} to M_{Z_t}. Further, the process ξ_t is $\mathscr{F}_t^Z$-adapted.*

Proof: Applying the Stratonovich form of Itô's formula, we get

$$da_\sigma(\xi_t) = \nabla^T a_\sigma(\xi_t)\rho(\xi_t) \circ da(Z_t) = da(Z_t).$$

Thus for $\sigma(\xi_s) = Z_s$, we get $\sigma(\xi_t) = Z_t$. The second conclusion follows from the uniqueness of the solution to the SDE (4.9). ∎

Denote the column vectors of $(I-p)c$ and $(I-p)\tilde{c}$ by $g_1, \cdots, g_m$ and $\tilde{g}_1, \cdots, \tilde{g}_d$ respectively. Let $b_0 = (I-p)\tilde{b}$. Then for each $x \in M_z$, the vectors $g_1(x), \cdots, g_m(x);\ \tilde{g}_1(x), \cdots, \tilde{g}_d(x);\ b_0(x)$ are all in $T_x M_z$. The signal process X_t satisfies

$$dX_t = \rho(X_t) \circ da(Z_t) + \sum_{i=1}^m g_i(X_t) \circ dW_t^i + \sum_{j=1}^d \tilde{g}_j(X_t) \circ dB_t^j + b_0(X_t)dt. \tag{4.10}$$

It is well-known that the Jacobian matrix $\xi'_{t,s}$ of the stochastic flow $\xi_{t,s}$ is invertible (cf. Ikeda and Watanabe [5]). The operator $(\xi_{t,s}^{-1})_*$ defined below pulls a vector at $\xi_{t,s}(x)$ back to a vector at x.

Definition 4.7. *Let g be a vector field in $\mathbb{R}^d$. The random vector field $(\xi_{t,s}^{-1})_* g$ is defined as*

$$(\xi_{t,s}^{-1})_* g(x) \equiv (\xi'_{t,s})^{-1} g(\xi_{t,s}(x))$$

for any a_σ regular point $x \in \mathbb{R}^d$.

We consider the following SDE on $\mathbb{R}^d$:

$$d\kappa_t = (\xi_{t,0}^{-1})_* b_0(\kappa_t)dt + \sum_{i=1}^m (\xi_{t,0}^{-1})_* g_i(\kappa_t) \circ dW_t^i + \sum_{j=1}^d (\xi_{t,0}^{-1})_* \tilde{g}_j(\kappa_t) \circ dB_t^j. \tag{4.11}$$

Lemma 4.8. *The SDE (4.11) has a unique strong solution. Further, if $Z_0 = z$, then $\kappa_t \in M_z$ for all $t \geq 0$ a.s.*

Proof: Note that $(\xi'_{t,0})^{-1}$ satisfies a diffusion SDE with bounded coefficients (cf. Ikeda and Watanabe [5]). By Gronwall's inequality, it is easy to show that

$$\mathbb{E} \sup_{0 \leq t \leq T} \|(\xi'_{t,0})^{-1}\|^p < \infty, \qquad \forall p > 1.$$

Therefore, we can prove that there is a constant K such that

$$\mathbb{E}\left|(\xi_{t,0}^{-1})_* b_0(\kappa_1) - (\xi_{t,0}^{-1})_* b_0(\kappa_2)\right|^2 \leq K|\kappa_1 - \kappa_2|^2, \qquad \forall\ \kappa_1,\ \kappa_2 \in \mathbb{R}^d.$$

Similarly, the above inequality holds with b_0 replaced by g_i, $1 \leq i \leq m$ or $\tilde{g}_j$, $1 \leq j \leq d$. By standard arguments, we see that (4.11) has a unique strong solution.

Applying Itô's formula to (4.11), we get

$$\begin{aligned} da_\sigma(\kappa_t) &= q(\kappa_t)(\xi_{t,0}^{-1})_* b_0(\kappa_t)dt + \sum_{i=1}^m q(\kappa_t)(\xi_{t,0}^{-1})_* g_i(\kappa_t) \circ dW_t^i \\ &\quad + \sum_{j=1}^d q(\kappa_t)(\xi_{t,0}^{-1})_* \tilde{g}_j(\kappa_t) \circ dB_t^j. \end{aligned}$$

Since $b_0(\xi_{t,0}(\kappa_t)) \in T_{\xi_{t,0}(\kappa_t)} M_{Z_t}$, we have $(\xi_{t,0}^{-1})_* b_0(\kappa_t) \in T_{\kappa_t} M_{\sigma(\kappa_t)}$. As the row vectors of $q(\kappa_t)$ are in $N_{\kappa_t} M_{\sigma(\kappa_t)}$, we see that

$$q(\kappa_t)(\xi_{t,0}^{-1})_* b_0(\kappa_t) = 0.$$

Also the above equality holds with b_0 replaced by g_i, $1 \le i \le m$ or $\tilde{g}_j$, $1 \le j \le d$. Thus, $da_\sigma(\kappa_t) = 0$, and hence, $a_\sigma(\kappa_t) = a_\sigma(\kappa_0)$, $\forall\ t \ge 0$. This proves that $\sigma(\kappa_t) = z$, and hence $\kappa_t \in M_z$, $\forall\ t \ge 0$, a.s. ∎

The next theorem gives the decomposition of the signal process.

Theorem 4.9. *For almost all $\omega \in \Omega$, we have*

$$X_t(\omega) = \xi_{t,0}(\kappa_t(\omega), \omega), \qquad \forall t \ge 0. \tag{4.12}$$

Proof: Denote the right hand side of (4.12) by $\tilde{X}_t(\omega)$. Applying Itô's formula, we get

$$\begin{aligned} d\tilde{X}_t &= \sum_{j=1}^d \partial_j \xi_{t,0}(\kappa_t) \circ d\kappa_t^j + \rho_t(\xi_{t,0}(\kappa_t)) \circ da(Z_t) \\ &= \xi'_{t,0}(\kappa_t)(\xi'_{t,0}(\kappa_t))^{-1} b_0(\xi_{t,0}(\kappa_t))dt + \sum_{i=1}^m \xi'_{t,0}(\kappa_t)(\xi'_{t,0}(\kappa_t))^{-1} g_i(\xi_{t,0}(\kappa_t)) \circ dW_t^i \\ &\quad + \sum_{j=1}^d \xi'_{t,0}(\kappa_t)(\xi'_{t,0}(\kappa_t))^{-1} \tilde{g}_j(\xi_{t,0}(\kappa_t)) \circ dB_t^j + \rho_t(\tilde{X}_t) \circ da(Z_t) \\ &= \rho(\tilde{X}_t) \circ da(Z_t) + \sum_{i=1}^m g_i(\tilde{X}_t) \circ dW_t^i + \sum_{j=1}^d \tilde{g}_j(\tilde{X}_t) \circ dB_t^j + b_0(\tilde{X}_t)dt. \end{aligned}$$

By the uniqueness of the solution to (4.10), we see that the representation (4.12) holds. ∎

The optimal filter then satisfies

$$\pi_t f = \mathbb{E}(f(\xi_{t,0}(\kappa_t)) | \mathscr{F}_t^{\hat{Y}} \vee \mathscr{F}_t^Z), \qquad t > 0.$$

Note that $\xi_{t,0}$ is $\mathscr{F}_t^Z$-measurable. Thus, we may regard $\xi_{t,0}$ as known and the singular filtering problem can be transformed to a classical one as follows: For $f \in C_b(M_z)$, let

$$U_t f \equiv \mathbb{E}[f(\kappa_t) | \mathscr{F}_t^{\hat{Y}} \vee \mathscr{F}_t^Z].$$

Then U_t is the optimal filter with the signal process κ_t being given by (4.11) and the observation $(\hat{Y}_t, a(Z_t))$ being given by

$$d\hat{Y}_t = Z_t^{-1} h(\xi_{t,0}(\kappa_t))dt + dW_t \tag{4.13}$$

and

$$da(Z_t) = La_\sigma(\xi_{t,0}(\kappa_t))dt + H_1(a(Z_t))dW_t + H_2(a(Z_t))dB_t. \tag{4.14}$$

Note that the filtering problem with signal (4.11) and observations (4.13) and (4.14) is classical. We sketch below the derivation of the Zakai equation for the unnormalized version $\tilde{U}_t$ of the process U_t, and leave the details to the reader. Once this is done, we note that

$$\pi_t f = U_t\left(f \circ \xi_{t,0}\right), \qquad t > 0.$$

Without loss of generality, we may and will assume that $H_2(a(Z_t))$ is of full rank a.s. Let $\tilde{H}_2(a(Z_t))$ be a $(d - \tilde{m}) \times d$-matrix-valued process such that $\begin{pmatrix} H_2(a(Z_t)) \\ \tilde{H}_2(a(Z_t)) \end{pmatrix}$ is invertible and $\tilde{H}_2(a(Z_t))H_2(a(Z_t))^T = 0$ a.s. Define independent Brownian motions $\tilde{B}_t$ and $\hat{B}_t$ of dimensions $\tilde{m}$ and $d - \tilde{m}$ respectively by

$$d\tilde{B}_t = \left(H_2(a(Z_t))H_2(a(Z_t))^T\right)^{-\frac{1}{2}} H_2(a(Z_t))dB_t$$

and

$$d\hat{B}_t = \left(\tilde{H}_2(a(Z_t))\tilde{H}_2(a(Z_t))^T\right)^{-\frac{1}{2}} \tilde{H}_2(a(Z_t))dB_t.$$

Denote

$$\begin{pmatrix} H_2(a(Z_t)) \\ \tilde{H}_2(a(Z_t)) \end{pmatrix}^{-1} = \left(h_2(a(Z_t)), \tilde{h}_2(a(Z_t))\right).$$

Then

$$\begin{aligned} dB_t &= h_2(a(Z_t))\left(H_2(a(Z_t))H_2(a(Z_t))^T\right)^{\frac{1}{2}} d\tilde{B}_t \\ &\quad + \tilde{h}_2(a(Z_t))\left(\tilde{H}_2(a(Z_t))\tilde{H}_2(a(Z_t))^T\right)^{\frac{1}{2}} d\hat{B}_t. \end{aligned}$$

Define process $\tilde{Z}_t$ by

$$d\tilde{Z}_t = \begin{pmatrix} I & 0 \\ H_1(a(Z_t)) & \left(H_2(a(Z_t))H_2(a(Z_t))^T\right)^{\frac{1}{2}} \end{pmatrix}^{-1} d\begin{pmatrix} \hat{Y}_t \\ a(Z_t) \end{pmatrix} \tag{4.15}$$

and function $\tilde{h}$ by

$$\tilde{h}(t, Y, \kappa_t) = \begin{pmatrix} I & 0 \\ H_1(a(Z_t)) & \left(H_2(a(Z_t))H_2(a(Z_t))^T\right)^{\frac{1}{2}} \end{pmatrix}^{-1} \begin{pmatrix} Z_t^{-1}h(\xi_{t,0}(\kappa_t)) \\ La_\sigma(\xi_{t,0}(\kappa_t)) \end{pmatrix}. \tag{4.16}$$

Then, $\tilde{Z}_t$ is observable and

$$d\tilde{Z}_t = \tilde{h}(t, Y, \kappa_t)dt + d\begin{pmatrix} W_t \\ \tilde{B}_t \end{pmatrix}.$$

Note that (4.11) can be written in the Itô form

$$\begin{aligned} d\kappa_t &= \tilde{b}(t, Y, \kappa_t)dt + \tilde{\sigma}_1(t, Y, \kappa_t)dW_t + \tilde{\sigma}_2(t, Y, \kappa_t)dB_t \\ &= \tilde{b}(t, Y, \kappa_t)dt + \tilde{\sigma}_1(t, Y, \kappa_t)dW_t \\ &\quad + \tilde{\sigma}_2(t, Y, \kappa_t)h_2(a(Z_t))\left(H_2(a(Z_t))H_2(a(Z_t))^T\right)^{\frac{1}{2}} d\tilde{B}_t \\ &\quad + \tilde{\sigma}_2(t, Y, \kappa_t)\tilde{h}_2(a(Z_t))\left(\tilde{H}_2(a(Z_t))\tilde{H}_2(a(Z_t))^T\right)^{\frac{1}{2}} d\hat{B}_t \end{aligned}$$

for suitably defined coefficients $\tilde{b}$, $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$. Denote

$$\hat{\sigma}(t,Y,\kappa_t) = \left(\tilde{\sigma}_1(t,Y,\kappa_t), \tilde{\sigma}_2(t,Y,\kappa_t)h_2(a(Z_t))\left(H_2(a(Z_t))H_2(a(Z_t))^T\right)^{\frac{1}{2}}\right)$$

and

$$L_{t,Y}f(\kappa) = \frac{1}{2}\sum_{ij}\tilde{a}_{ij}(t,Y,\kappa)\partial_{ij}f(\kappa) + \tilde{b}(t,Y,\kappa)^T\nabla f(\kappa), \tag{4.17}$$

where $\tilde{a}(t,Y,\kappa) = \sum_{i=1}^2 \tilde{\sigma}_i(t,Y,\kappa)\tilde{\sigma}_i(t,Y,\kappa)^T$. Then, by Kallianpur-Striebel formula, we have

$$U_t = \left(\tilde{U}_t 1\right)^{-1}\tilde{U}_t, \tag{4.18}$$

while the unnormalized filter $\tilde{U}_t$ satisfies the following Zakai equation:

$$d\tilde{U}_t f = \tilde{U}_t(L_{t,Y}f)dt + \tilde{U}_t\left(\nabla^T f\hat{\sigma}(t,Y) + \tilde{h}(t,Y)^T f\right)d\tilde{Z}_t. \tag{4.19}$$

A filtering equation for U_t can be derived by applying Itô's formula to (4.18) and (4.19).

Now we summarize the discussion above to a theorem.

Theorem 4.10. *Suppose that the assumptions (R, S, BC, X0, IN, ND) are satisfied. Let $\tilde{Z}$ be the observation process given by (4.15). Let $\tilde{h}$ and $L_{t,Y}$ be given by (4.16) and (4.17) respectively. Let $\xi_{t,s}$ be the stochastic flow given by (4.9). Let $\tilde{U}_t$ be the unique $\mathscr{M}_F(M_{Z_0})$-valued solution to the Zakai equation (4.19) and $U_t = (\tilde{U}_t 1)^{-1}\tilde{U}_t$. Then, the optimal filter π_t is given by*

$$\pi_t(f) = U_t(f\circ\xi_{t,0}), \qquad \forall\, f \in C_b(\mathbb{R}^d).$$

Finally, we indicate that an analogue of the Kalman filter for the singular filtering setup is a part of a work in progress by Liu [10]. We indicate the model and the result anticipated. For simplicity, we take $m = 1$.

Let the signal be given by the linear equation:

$$dX_t = bX_t dt + dB_t,$$

and the observation is

$$dY_t = h^T X_t dt + |Q^T X_t| dW_t,$$

where (B,W) is the $d+1$-dimensional Brownian motion, b is a $d\times d$-matrix and h, $Q \in \mathbb{R}^d$. It is anticipated that π_t will be a linear combination of two normal measures on planes $\{x \in \mathbb{R}^d : Q^T x = \pm Z_t\}$, where $Z_t = |Q^T X_t|$ is observable. Equations for the condition means and variances, as well as the weight process in the linear combination, will be derived.

5 The filtering model with Ornstein-Uhlenbeck noise

In this section, we consider the filtering problem with OU-process as the observation noise. As we indicated in the introduction, the OU-process is an approximation of the white noise which exists in the sense of generalized function only. We transform this filtering problem with OU-process noise to a singular filtering problem of the form studied in the previous sections. To be more general, we assume that the OU process is given by

$$dO_t = a_1 O_t dt + b_1 dW_t$$

and y is as in (1.4). Let

$$Y_t = \int_0^t e^{-a_1 s} d(e^{a_1 s} y_s).$$

Then $\mathscr{F}_t^Y = \mathscr{F}_t^y$ and, by Itô's formula,

$$dY_t = (Lh(X_t) + a_1 h(X_t))dt + b_1 dW_t + \nabla hc(X_t)dB_t.$$

Define the pair process

$$\begin{aligned} dV_t &= (b_1^2 I + \nabla hc(\nabla hc)^T)^{-1/2}(X_t)\left(b_1 dW_t + \nabla hc(X_t)dB_t\right), \\ d\tilde{V}_t &= (b_1^2 I + (\nabla hc)^T \nabla hc)^{-1/2}(X_t)\left((\nabla hc)^T(X_t)dW_t - b_1 dB_t\right). \end{aligned}$$

Then the pair process $(V, \tilde{V})$ is an $(m+d)$-dimensional Brownian motion and

$$\begin{aligned} dB_t = &\ (b_1^2 I + \nabla hc(\nabla hc)^T)^{-1}(\nabla hc)^T(b_1^2 I + \nabla hc(\nabla hc)^T)^{-1/2}(X_t)dV_t \\ &-(b_1^2 I + (\nabla hc)^T \nabla hc)^{-1/2}(X_t)d\tilde{V}_t. \end{aligned}$$

Hence, if we define the following functions:

$$\begin{aligned} \sigma &= (b_1^2 I + \nabla hc(\nabla hc)^T)^{1/2} \\ c_1 &= c(b_1^2 I + (\nabla hc)^T \nabla hc)^{-1}(\nabla hc)^T(b_1^2 I + \nabla hc(\nabla hc)^T)^{1/2} \\ \tilde{c}_1 &= -c(b_1^2 I + (\nabla hc)^T \nabla hc)^{-1/2}, \end{aligned}$$

then the pair signal/observation process is written as

$$\begin{cases} dX_t = b(X_t)dt + c_1(X_t)dV_t + \tilde{c}_1\left(X_t\right)d\tilde{V}_t \\ dY_t = \left(Lh + a_1 h\right)(X_t)dt + \sigma(X_t)dV_t \end{cases}. \tag{5.1}$$

Hence the filtering model (5.1) is of the form (2.1).

Next, we give an example to demonstrate that for filtering problem with OU observation noise, both discrete and continuous singularity can occur. This example is a special case of those considered by Liu [10], so we omit the details here.

Example 5.1. *Let the signal X_t^i, $i = 1, 2$, be governed by SDEs*

$$dX_t^i = \sigma_i \sqrt{X_t^i} dB_t^i,$$

where σ_1 and σ_2 are two constants, and B_t^1 and B_t^2 are two independent one-dimensional Brownian motions. Suppose that the observation model is

$$y_t = X_t^1 + X_t^2 + O_t,$$

where O_t is a real-valued OU process with $a_1 = b_1 = 1$.

Applying the quadratic variation procedure indicated above, we see that $Z_t \equiv \sigma_1^2 X_t^1 + \sigma_2^2 X_t^2$ is observable. It can be proved that when $\sigma_1 = \sigma_2$, the singular filter is of continuous type. In this case, the additional condition (IN) in Section 4 is satisfied after transforming the filtering problem with OU noise to a singular filtering problem.

When $\sigma_1 \neq \sigma_2$, the condition (IN) is not satisfied. In this case,

$$Z_t^p \equiv \sigma_1^{2p} X_t^1 + \sigma_2^{2p} X_t^2, \qquad p = 1, 2, \cdots$$

are all observable and hence the level sets are countable and the conditional distribution is of discrete type.

Finally, we would like to point out that the linear filtering with OU process as the observation noise has been studied by Liu [10]. The model is given by

$$\begin{cases} dX_t = (b_1 X_t + b_0)dt + CdB_t, \\ y_t = HX_t + O_t, \end{cases}$$

where O_t is the m-dimensional OU-process independent of B_t, $H,\ b_1,\ b_0,\ C$ are matrices of dimensions $m \times d$, $d \times d$, $d \times 1$, $d \times d$, respectively. It is classified to three classes (classical, continuous singular, discrete singular) according to the ranks of some matrices determined by the coefficients $H,\ b_1,\ b_0,\ C$. We refer the interested reader to the working paper [10] for detail.

Acknowledgement. The authors would like to thank the referee for helpful suggestions.

References

[1] A. G. Bhatt and R. L. Karandikar (2003). On filtering with Ornstein-Uhlenbeck process as noise. *J. Indian Statist. Assoc.* **41**, no. 2, 205–220. MR2101994

[2] L.C. Evans and R.F. Gariepy (1992). *Measure Theory and Fine Properties of Functions.* CRC Press, Boca Raton, FL. MR1158660

[3] A. Friedman (1975). Stochastic Differential Equations and Applications, Volume 1, Academic Press. MR0494490

[4] L. Gawarecki and V. Mandrekar (2000). On the Zakai equation of filtering with Gaussian noise. *Stochastics in finite and infinite dimensions, Volume in honor of Gopinath Kallianpur, eds. T. Hida et al*, 145–151. MR1797085

[5] N. Ikeda and S. Watanabe (1989). *Stochastic Differential Equations and Diffusions*. North Holland Publishing Company. MR1011252

[6] M. Joannides and F. LeGland (1997). Nonlinear Filtering with Perfect Discrete Time Observations, *Proceedings of the 34th IEEE Conference on Decision and Control,* New Orleans, December 13-15, 1995, pp. 4012-4017.

[7] M. Joannides and F. LeGland (1997). Nonlinear Filtering with Continuous Time Perfect Observations and Noninformative Quadratic Variation. *Proceedings of the 36th IEEE Conference on Decision and Control*, San Diego, December 10-12, 1997, pp. 1645-1650.

[8] G. Kallianpur and R.L. Karandikar (1988), *White Noise Theory of Prediction, Filtering and Smoothing*. Gordon and Breach. MR0966884

[9] H. Kunita (1993). Representation and stability of nonlinear filters associated with Gaussian noises. *Stochastic Processes: A festschrift in honor of Gopinath Kallianpur, eds. Cambanis et al, Springer-Verlag, New York.* 201–210. MR1427316

[10] Z. Liu and J. Xiong (2009). Some computable examples for singular filtering. *Technical Report* at http://www.math.utk.edu/~jxiong/publist.html.

[11] V. Mandrekar and P. K. Mandal (2000). A Bayes formula for Gaussian processes and its applications. *SIAM J. Control Optim.* ***39***, 852–871. MR1786333

[12] W. Rudin, Principles of mathematical analysis, Third edition, International Series in Pure and Applied Mathematics, New York: McGraw-Hill Book Co., 1976. MR0385023