

Linear filtering with Ornstein–Uhlenbeck process as noise

ABHAY G BHATT

Indian Statistical Institute, 7, S J S Sansanwal Marg, New Delhi 110 016, India
e-mail: abhay@isid.ac.in

Abstract. We consider a linear filtering model (with feedback) when the observation noise is an Ornstein–Uhlenbeck (OU) process with parameter β . The coefficients appearing in the model are all assumed to be bounded. In addition, the coefficients appearing in the observation equation are also assumed to be differentiable. We consider the general case when the OU noise is also correlated with the signal. Under these conditions, we derive the filtering equations for the optimal filter.

Keywords. Linear filtering; Ornstein–Uhlenbeck process; optimal filter.

1. Introduction

The theory of filtering can be explained as follows. The process of interest X , called the Signal, is not directly observable. However, it is assumed that at time t a (known) function of $\{X(s) : s \leq t\}$ is observable in the presence of an additive noise.

The question of interest is to estimate the signal $X(t)$ given the observations $\{Y(s) : 0 \leq s \leq t\}$. This is known as filtering the noise (to recover the signal). The best estimate, called the *optimal filter* is the conditional distribution of $X(t)$ given the observation σ -field $\mathcal{F}_t^Y = \sigma\{Y(s) : 0 \leq s \leq t\}$. We will denote this conditional distribution by $\pi(t)$.

In the classical theory of filtering, the observation noise is modelled to be a Brownian motion. Then, under fairly general conditions on the signal and on the observation function, the filter π solves an infinite dimensional stochastic differential equation called the Fujisaki–Kallianpur–Kunita (FKK) equation (Fujisaki *et al* 1972). Uniqueness of solution to the FKK equation is also known (under some integrability conditions) when the signal-observation pair (X, Y) is uniquely determined either via a martingale problem (Bhatt *et al* 1995) or as a solution of a (non-anticipative) stochastic differential equation (Bhatt & Karandikar 1999).

In analogy with discrete time models, we can assume that the observation noise has stationary independent increments. But this implies that the noise is a Brownian motion. Hence, the classical model of filtering is a natural model for filtering. At the same time, it has given rise to a very rich theory as mentioned in the previous paragraph. (See also Elliott 1982 and Kallianpur 1980.)

However, this model has been objected to by engineers from a practical point of view. The actual observed paths of the accumulative observation process Y are smooth. But the classical

model gives zero probability to all such smooth paths. (See Kallianpur & Karandikar 1988 for a detailed discussion on this.)

Recently, several authors have considered the filtering model when the observation noise is a process other than a Brownian motion. In Kunita (1993), the author initiated the study of filtering theory with general Gaussian noise processes. This was also followed up in Gawarecki & Mandrekar (2000) and Mandrekar & Mandal (2000).

A special case of the above is when the noise is an Ornstein–Uhlenbeck process (OUP) with parameter $\beta > 0$. Such a filtering model answers the criticism by engineers of the classical model. Moreover, when β is large, the OUP approximates Brownian motion. Thus a filtering model with OUP noise approximates the classical model.

The optimal filter π^β for the model with OUP noise has been studied. This is done by looking at the Bayes' formula for the filter in Mandrekar & Mandal (2000), Bhatt & Karandikar (2003) and Gawarecki & Mandrekar (2003). The filtering equation has been studied by Gawarecki & Mandrekar (2000), Bhatt *et al* (2004) and Kouritzin & Xiong (2002). However, each of these articles uses different assumptions on the model, in particular, on the function h . Gawarecki & Mandrekar (2000) and Mandrekar & Mandal (2000) require that $h(X(t))$ is almost surely differentiable in t , a very stringent condition. A perturbed model (with a small time-lag in recording the observations) is used by Bhatt & Karandikar (2003) and Gawarecki & Mandrekar (2003). Bhatt *et al* (2004) and Kouritzin & Xiong (2002) require the observation function h to be bounded and smooth. Also, a point to note is that all these authors assume that the observation noise and the signal are independent.

In this article, we consider a very important special case – linear filtering. In the classical set-up this was first considered by Kalman and Bucy (Kalman 1960; Kalman & Bucy 1961). The Kalman–Bucy filter is also the most widely used filter because of its simplicity. In this case the optimal filter is *finite dimensional* in the sense that π is characterized as the unique solution of a system of *two* SDE's (as opposed to infinitely many in the general nonlinear filtering problem).

The linear filtering model with observation noise being an Ornstein–Uhlenbeck process, which is also independent of the signal, was considered by Gawarecki & Mandrekar (2003). However, as mentioned earlier, they look at a *perturbed* model where there is a time-lag in recording the observations. They show that in this case the filter approximates the classical Kalman filter.

We also consider the linear filtering model with OUP as noise. Our model (given by (16)–(17) below) is very general in the sense that it allows feedback (from the observations into the signal). Also, we consider the case when the observation noise is correlated with the noise driving the signal. We however assume that the (non-random) coefficients that appear in the model are all bounded and the ones appearing in (17) are also differentiable. Under these conditions the optimal filter continues to be finite-dimensional and we derive the corresponding filtering equations.

The paper is organised as follows. In the next section we consider the corresponding classical linear filtering model and write down the filtering equations (theorem 1). Also, for later use, we note down a result on FKK equations from the general theory of nonlinear filtering (theorem 2).

In § 3, we introduce the linear filtering model for the signal-observation process with OUP as the observation noise. We derive the equations of filtering for the optimal filter (theorem 4). This is done by effecting a series of transformations on the model to recast it as a classical model of filtering (i.e., one where the observation noise is a certain Brownian motion). However, in doing so, the terms appearing in the SDE for the *transformed* signal-observation

pair process depend on the entire past of the process. This necessitates the use of results from the nonlinear filtering theory. We also show that the filtering equations admit a unique solution.

For notational simplicity, we consider the one-dimensional case.

2. The classical model

We start with a general linear filtering model (allowing feedback) in the classical set-up. The signal process X and the observation process Y are given by the system of stochastic differential equations,

$$dX(t) = [a_0(t) + a_1(t)X(t) + a_2(t)Y(t)] dt + b(t)dW^S(t), \tag{1}$$

$$dY(t) = [h_0(t) + h_1(t)X(t) + h_2(t)Y(t)] dt + dW^N(t), \tag{2}$$

for $0 \leq t \leq T$ where $X(0)$ is a Gaussian random variable independent of (W^S, W^N) and $Y(0) = 0$. The coefficients $a_0, a_1, a_2, h_1, h_2, h_3$ and b are all assumed to be non-random, bounded and measurable. The observation noise W^N and the noise W^S driving the SDE for the signal X are assumed to be standard Brownian motions with

$$\langle W^S, W^N \rangle_t = \int_0^t \rho(u) du.$$

The following facts are well-known. Such a system of SDE's admits a unique solution (X, Y) . By virtue of the linear nature of the coefficients, it follows that the pair (X, Y) is jointly Gaussian. Then the optimal filter π , which is the conditional distribution of the signal given the observations, is defined by

$$\int f d\pi(t) = E[f(X(t)) | \mathcal{F}_t^Y], \quad \text{for all bounded, continuous functions } f. \tag{3}$$

Further, $\pi(t)$ is also a Gaussian measure (see for example Elliott 1982, Kallianpur 1980). Hence $\pi(t)$ is completely determined by its first two moments $\hat{X}(t)$ and $P(t)$ which are defined by

$$\hat{X}(t) = E[X(t) | \mathcal{F}_t^Y], \tag{4}$$

$$P(t) = E[(X(t) - \hat{X}(t))^2 | \mathcal{F}_t^Y]. \tag{5}$$

Moreover, the conditional variance $P(t)$ is non-random. (See for example p. 522 of Rao 1985.)

The additional information got by observing Y at the instant t is given by the *innovations process* $I(t)$, which is defined by

$$I(t) = Y(t) - \int_0^t [h_0(u) + h_1(u)\hat{X}(u) + h_2(u)Y(u)] du.$$

It is well-known that I is a standard Brownian motion with respect to the observation σ -field \mathcal{F}_t^Y .

We have the following theorem (see theorem 10.5.1. of Kallianpur 1980) regarding the optimal filter.

Theorem 1. The Kalman filter (\hat{X}, P) for the signal-observation model (1)–(2) is the unique solution of the following system of equations.

$$\begin{aligned} \hat{X}(t) = E[X_0] + \int_0^t [a_0(u) + a_1(u)\hat{X}(u) + a_2(u)Y(u)] du \\ + \int_0^t [b(u)\rho(u) + P(u)h_1(u)] dI(u) \end{aligned} \quad (6)$$

$$P(t) = V[X_0] + \int_0^t [2a_1(u)P(u) + (b(u))^2 - (b(u)\rho(u) + h_1(u)P(u))^2] du. \quad (7)$$

Remark 1. The above result is also true for unbounded coefficients under some appropriate integrability conditions, see theorem 10.5.1. of Kallianpur (1980).

To end this section we give the general FKK equation (theorem 8.4.4 from Kallianpur 1980, also theorem 18.11 from Elliott 1982) which is used in the next section. Suppose that the observation model is given by

$$Z(t) = \int_0^t H(u, X, Z) du + \int_0^t \alpha(u) dW(u),$$

where W is standard Brownian motion, the observation function H a non-anticipating functional of (X, Z) and α a deterministic function bounded away from zero. We assume the energy condition

$$\int_0^T |H(u, X, Z)|^2 du < \infty, \quad \text{a.s.} \quad (8)$$

Suppose that the signal process X is such that

$$M^f(t) = f(X(t)) - \int_0^t \tilde{A}f(u) du$$

is a martingale, with

$$\langle M^f, W \rangle(t) = \int_0^t \tilde{D}f(u) du,$$

where $\tilde{A}f$ and $\tilde{D}f$ are some (non-anticipative) functionals of X and Z . Let $\pi(t, f)$ denote the conditional distribution of $f(X(t))$ given the observation σ -field $\mathcal{F}_t^Z = \sigma\{Z(u) : 0 \leq u \leq t\}$. Let

$$I(t) = \int_0^t (\alpha(u))^{-1} dY(u) - \int_0^t (\alpha(u))^{-1} \pi(u, H) du$$

be the innovations process. Then we have the following theorem.

Theorem 2. Under the above setup, $\pi(t, f)$ satisfies

$$\begin{aligned} \pi(t, f) = & \pi(0, f) + \int_0^t \pi(u, \tilde{A}f) du \\ & + \int_0^t [\pi(u, \tilde{D}f) + (\alpha(u))^{-1} (\pi(u, Hf) - \pi(u, H)\pi(u, f))] dI(u). \end{aligned} \tag{9}$$

3. Linear signal–observation model with OUP noise

Let W^N and W^S be standard Brownian motions with

$$\langle W^S, W^N \rangle(t) = \int_0^t \rho(u) du. \tag{10}$$

We assume that ρ is a bounded measurable function on $[0, T]$ with

$$\rho(u) > -1, \quad 0 \leq u \leq T. \tag{11}$$

Fix a $\beta > 0$. Consider the Ornstein–Uhlenbeck *velocity* process

$$\begin{aligned} dV^\beta(t) = & -\beta V^\beta(t) dt + dW^N(t), \quad 0 \leq t \leq T, \\ V^\beta(0) = & 0. \end{aligned} \tag{12}$$

Then V^β is given by

$$V^\beta(t) = \beta \int_0^t e^{-\beta(t-u)} dW^N(u), \quad 0 \leq t \leq T. \tag{13}$$

Let O^β denote the Ornstein–Uhlenbeck *displacement* process defined by

$$O^\beta(t) = \int_0^t V^\beta(u) du, \quad 0 \leq t \leq T. \tag{14}$$

Then O^β has the form

$$O^\beta(t) = \int_0^t (1 - e^{-\beta(t-u)}) dW^N(u), \quad 0 \leq t \leq T. \tag{15}$$

It is well-known that O^β converges to W^N in L^2 as $\beta \rightarrow \infty$ (see Nelson 1967).

We will consider the following linear signal–observation model. Let X_0 be a Gaussian random variable independent of (W^S, W^N) . Throughout the article, the coefficients $a_0, a_1, a_2, h_0, h_1, h_2$ and b will be assumed to be non-random, bounded and measurable functions on $[0, T]$. Further the coefficients h_0, h_1, h_2 will be assumed to be continuously differentiable on $[0, T]$. Let $C < \infty$ denote a common bound for all the coefficients and their derivatives. The model is then given by

$$dX^\beta(t) = [a_0(t) + a_1(t)X^\beta(t) + a_2(t)Y^\beta(t)] dt + b(t)dW^S(t) \tag{16}$$

$$dY^\beta(t) = [h_0(t) + h_1(t)X^\beta(t) + h_2(t)Y^\beta(t)] dt + dO^\beta(t) \tag{17}$$

for $0 \leq t \leq T$ with $X^\beta(0) = X_0$ and $Y^\beta(0) = 0$. This is a SDE with linear coefficients driven by the Gaussian semi-martingale (W^S, O^β) . Existence and uniqueness of solution to such an equation is quite well-known (see for example Karandikar 1989, Métivier 1982). We also have the following observation which is crucially used in the rest of the article.

Theorem 3. *Let the processes W^S, W^N, O^β , the random variable X_0 and the coefficients $a_0, a_1, a_2, h_0, h_1, h_2$ and b be as above. Then the solution of the SDE (16)–(17) is a Gaussian process.*

Proof. We note that (W^S, O^β, X_0) is jointly Gaussian. We now construct a Gaussian solution for (16)–(17). Let

$$X^{\beta,0}(t) = X_0, \quad Y^{\beta,0}(t) = 0, \quad 0 \leq t \leq T,$$

and define successively for $m \geq 1$,

$$\begin{aligned} X^{\beta,m}(t) &= \int_0^t [a_0(u) + a_1(u)X^{\beta,m-1}(u) + a_2(u)Y^{\beta,m-1}(u)] du \\ &\quad + \int_0^t b(u)dW^S(u) \\ Y^{\beta,m}(t) &= \int_0^t [h_0(u) + h_1(u)X^{\beta,m-1}(u) + h_2(u)Y^{\beta,m-1}(u)] du + O^\beta(t). \end{aligned}$$

Note that $(X^{\beta,m}, Y^{\beta,m})$ is a Gaussian process for every $m \geq 1$. Also

$$\begin{aligned} &|(X^{\beta,m} - X^{\beta,m-1})(t)|^2 \\ &= \left| \int_0^t [a_1(u)(X^{\beta,m-1} - X^{\beta,m-2})(u) + a_2(u)(Y^{\beta,m-1} - Y^{\beta,m-2})(u)] du \right|^2 \\ &\leq 2C^2T \int_0^t [|(X^{\beta,m-1} - X^{\beta,m-2})(u)|^2 + |(Y^{\beta,m-1} - Y^{\beta,m-2})(u)|^2] du. \end{aligned}$$

Similarly, we get

$$\begin{aligned} &|(Y^{\beta,m} - Y^{\beta,m-1})(t)|^2 \\ &\leq 2C^2T \int_0^t [|(X^{\beta,m-1} - X^{\beta,m-2})(u)|^2 + |(Y^{\beta,m-1} - Y^{\beta,m-2})(u)|^2] du. \end{aligned}$$

Define $\eta^m(t) = \sup_{0 \leq u \leq t} \{E[|(X^{\beta,m} - X^{\beta,m-1})(t)|^2 + |(Y^{\beta,m} - Y^{\beta,m-1})(t)|^2]\}$. Then using Fubini's theorem it follows that

$$\eta^m(t) \leq 4C^2T \int_0^t \eta^{m-1}(u)du \leq \dots \leq \frac{(4C^2T)^{m-1}}{(m-1)!} \eta^1(t), \quad (18)$$

where

$$\begin{aligned} \eta^1(t) &\leq \sup_{0 \leq u \leq t} \{ E [|X^{\beta,1}(u) - X_0|^2 + |Y^{\beta,1}(u)|^2] \} \\ &\leq 2 \sup_{0 \leq u \leq t} \left\{ E \left| \int_0^u (a_0(u) + a_1(u)X_0 - X_0) du \right|^2 + E \left| \int_0^u b(u) dW^S(u) \right|^2 \right. \\ &\quad \left. + E \left| \int_0^u (h_0(u) + h_1(u)X_0) du \right|^2 + E(O^\beta(t))^2 \right\} \\ &< \infty. \end{aligned} \tag{19}$$

It now follows from standard arguments that $(X^{\beta,m}, Y^{\beta,m})$ converge a.s. to a process (X^β, Y^β) , which is necessarily Gaussian. It is also clear from the construction that the limiting process is a solution of (16)–(17). This completes the proof. \square

The optimal filter $\pi^\beta(t)$ for the model (16)–(17) is the conditional distribution of $X^\beta(t)$ given the observation σ -field $\mathcal{F}_t^{Y^\beta} := \sigma\{Y^\beta(s); 0 \leq s \leq t\}$. Theorem 3 implies that $\pi^\beta(t)$ is Gaussian and hence, as in the classical case, it is completely determined by the first two conditional moments $\hat{X}^\beta(t)$ and $P^\beta(t)$ where

$$\hat{X}^\beta(t) = E[X^\beta(t) | \mathcal{F}_t^{Y^\beta}] \tag{20}$$

and

$$P^\beta(t) = E[(X^\beta(t) - \hat{X}^\beta(t))^2 | \mathcal{F}_t^{Y^\beta}] = E[(X^\beta(t))^2 | \mathcal{F}_t^{Y^\beta}] - (\hat{X}^\beta(t))^2. \tag{21}$$

Further it is well-known that the conditional variance $P^\beta(t)$ is in fact deterministic and is equal to $E[(X^\beta(t) - \hat{X}^\beta(t))^2]$ (see for example p. 522 of Rao (1985)).

Remark 2. In Bhatt & Karandikar (2003) the Ornstein–Uhlenbeck noise V^β was assumed to be stationary which implies that $V^\beta(0) \sim N(0, \beta/2)$. However this is not a necessary assumption and the analysis there works for a general $V^\beta(0)$. In this article, as seen from the above comments, we use the Gaussian nature of the process (X^β, Y^β) . To ensure this we would need $V^\beta(0)$ to be normally distributed. Thus the assumption that $V^\beta(0) = 0$ is only for convenience and is not a restriction.

Our aim now is to characterize $(\hat{X}^\beta(t), P^\beta(t))$ via filtering equations. Towards this end we proceed by making some *transformations* of the model (16)–(17). First, let us define

$$y^\beta(t) = \frac{d}{dt} Y^\beta(t); \quad \xi^\beta(t) = h_0(t) + h_1(t)X^\beta(t) + h_2(t)Y^\beta(t). \tag{22}$$

Then in view of (14) the observation model (17) can be rewritten in an equivalent form as

$$y^\beta(t) = \xi^\beta(t) + V^\beta(t), \quad 0 \leq t \leq T. \tag{23}$$

Remark 3. Model (23) can be thought of as the instantaneous observation model as opposed to (17) which is an accumulative observation model.

Now let

$$\tilde{y}^\beta(t) = e^{\beta t} y^\beta(t); \quad \tilde{\xi}^\beta(t) = e^{\beta t} \xi^\beta(t); \quad \tilde{V}^\beta(t) = e^{\beta t} V^\beta(t) \quad (24)$$

Then it follows that

$$\tilde{y}^\beta(t) = \tilde{\xi}^\beta(t) + \tilde{V}^\beta(t), \quad 0 \leq t \leq T. \quad (25)$$

Using (13) we have

$$\tilde{V}^\beta(t) = \beta \int_0^t e^{\beta u} dW^N(u).$$

Thus we get that \tilde{V}^β is a semimartingale and

$$\int_0^t \frac{e^{-\beta u}}{\beta} d\tilde{V}^\beta(u) = W^N(t). \quad (26)$$

The relation (25) implies that $\tilde{y}^\beta(t)$ is also a semimartingale. Now define

$$\tilde{Y}^\beta(t) = \int_0^t \frac{e^{-\beta u}}{\beta} d\tilde{y}^\beta(u); \quad \tilde{\xi}^\beta(t) = \int_0^t \frac{e^{-\beta u}}{\beta} d\tilde{\xi}^\beta(u). \quad (27)$$

It follows from (25)–(27) that

$$\tilde{Y}^\beta(t) = \tilde{\xi}^\beta(t) + W^N(t), \quad 0 \leq t \leq T. \quad (28)$$

Now, using (16), (17), (22), (24), (27) and applying integration by parts, we get

$$\begin{aligned} \tilde{\xi}^\beta(t) &= \int_0^t \frac{e^{-\beta u}}{\beta} d\tilde{\xi}^\beta(u) \\ &= \int_0^t \frac{e^{-\beta u}}{\beta} d[e^{\beta u} \xi^\beta(u)] \\ &= \int_0^t \xi^\beta(u) du + \int_0^t \frac{1}{\beta} d\xi^\beta(u) \\ &= \int_0^t \xi^\beta(u) du + \frac{1}{\beta} \int_0^t [h'_0(u) + h'_1(u)X^\beta(u) + h'_2(u)Y^\beta(u)] du \\ &\quad + \frac{1}{\beta} \int_0^t h_1(u) dX^\beta(u) + \frac{1}{\beta} \int_0^t h_2(u) dY^\beta(u) \\ &= \int_0^t [H_0^\beta(u) + H_1^\beta(u)X^\beta(u) + H_2^\beta(u)Y^\beta(u)] du \\ &\quad + \int_0^t B_1^\beta(u) dW^S(u) + \int_0^t B_2^\beta(u) dW^N(u), \end{aligned} \quad (29)$$

where

$$H_0^\beta(u) = h_0(u) + \frac{1}{\beta} [h_0'(u) + a_0(u)h_1(u) + h_0(u)h_2(u)], \quad (30a)$$

$$H_1^\beta(u) = h_1(u) + \frac{1}{\beta} [h_1'(u) + a_1(u)h_1(u) + h_1(u)h_2(u)], \quad (30b)$$

$$H_2^\beta(u) = h_2(u) + \frac{1}{\beta} [h_2'(u) + a_2(u)h_1(u) + (h_2(u))^2], \quad (30c)$$

$$B_1^\beta(u) = \frac{1}{\beta} h_1(u)b(u) \quad (30d)$$

and

$$B_2^\beta(u) = \frac{1}{\beta} h_2(u). \quad (30e)$$

Let a deterministic function α^β be defined by

$$(\alpha^\beta(t))^2 = (B_1^\beta(t))^2 + (B_2^\beta(t) + 1)^2 + 2\rho(t)B_1^\beta(t)(B_2^\beta(t) + 1), \quad 0 \leq t \leq T, \quad (31)$$

where ρ is as in (10). It follows from (11) that $(\alpha^\beta(t))^2 > 0$ for all t . Let

$$W(t) = \int_0^t \frac{B_1^\beta(u)}{\alpha^\beta(u)} dW^S(u) + \int_0^t \frac{B_2^\beta(u) + 1}{\alpha^\beta(u)} dW^N(u) \quad (32)$$

Then W is Standard Brownian motion. Moreover, using (29)–(32) we can now write the observation model (28) as

$$\begin{aligned} \tilde{Y}^\beta(t) &= \int_0^t [H_0^\beta(u) + H_1^\beta(u)X^\beta(u) + H_2^\beta(u)Y^\beta(u)] du \\ &\quad + \int_0^t \alpha^\beta(u) dW(u), \quad 0 \leq t \leq T. \end{aligned} \quad (33)$$

Also, it follows from (22), (24), and (27) that

$$\begin{aligned} \sigma(Y^\beta(u) : u \leq t) &= \sigma(y^\beta(u) : u \leq t) = \sigma(\tilde{y}^\beta(u) : u \leq t) \\ &= \sigma(\tilde{Y}^\beta(u) : u \leq t). \end{aligned} \quad (34)$$

Let us denote the σ -field in (34) by \mathcal{F}_t^β . Thus the optimal filter $(\hat{X}^\beta(t), P^\beta(t))$ is the conditional mean and variance of $X^\beta(t)$ given \mathcal{F}_t^β .

Remark 4. The new signal-observation model is now given by (16) and (33). The advantage in making the transformations is that both these equations are now driven by (correlated) Brownian motions W^S and W . We also note here, using (10) and (32), that

$$\begin{aligned} \langle W^S, W \rangle(t) &= \int_0^t [(\alpha^\beta(u))^{-1} B_1^\beta(u) + (\alpha^\beta(u))^{-1} \rho(u)(B_2^\beta(u) + 1)] du \\ &:= \int_0^t \lambda^\beta(u) du. \end{aligned} \quad (35)$$

Remark 5. It follows from (33) that the *transformed* observation process \tilde{Y}^β is still a Gaussian process. However the SDE's (16) and (33) are no longer linear in $(X^\beta, \tilde{Y}^\beta)$. In fact, it is easy to see that

$$Y^\beta(t) = \int_0^t e^{-\beta u} \left(\int_0^u \beta e^{\beta r} d\tilde{Y}^\beta(r) \right) du.$$

Thus the coefficients in (16) and (33) depend on the entire past of \tilde{Y}^β (in a non-anticipative way). Note that the dependence on X^β is still Markovian and linear.

Let us denote the non-anticipative functional by H^β . i.e.,

$$H^\beta(t, X^\beta(t), \tilde{Y}^\beta) = H_0^\beta(t) + H_1^\beta(t)X^\beta(t) + H_2^\beta(t)Y^\beta(t), \quad 0 \leq t \leq T. \quad (36)$$

Let $\hat{H}^\beta(t)$ denote its conditional expectation $E[H^\beta(t, X^\beta(t), \tilde{Y}^\beta) | \mathcal{F}_t^\beta]$. Define the *innovations process* I^β as follows.

$$I^\beta(t) = \int_0^t (\alpha^\beta(u))^{-1} d\tilde{Y}^\beta(u) - \int_0^t (\alpha^\beta(u))^{-1} \hat{H}^\beta(u) du, \quad 0 \leq t \leq T. \quad (37)$$

Then it is well-known (and also easy to see) that I^β is a standard Brownian motion adapted to the filtration \mathcal{F}_t^β . Now we are ready to prove the main theorem of this section.

Theorem 4. Let the processes W^S, W^N, O^β , the random variable X_0 and the coefficients $a_0, a_1, a_2, h_0, h_1, h_2$ and b be as above. Let the signal-observation model be given by (16)–(17). Let H_1^β be defined by (30b). Then the optimal filter (\hat{X}^β, P^β) (given by (20), (21)) is the unique solution of the equations

$$\begin{aligned} \hat{X}^\beta(t) &= E[X_0] + \int_0^t (a_0(s) + a_1(s)\hat{X}^\beta(s) + a_2(s)Y^\beta(s)) ds \\ &\quad + \int_0^t (b(s)\lambda^\beta(s) + (\alpha^\beta(s))^{-1} H_1^\beta(s)P^\beta(s)) dI^\beta(s) \end{aligned} \quad (38)$$

$$\begin{aligned} P^\beta(t) &= V[X_0] + \int_0^t \left\{ 2a_1(s)P^\beta(s) + (b(s))^2 \right. \\ &\quad \left. - (b(s)\lambda^\beta(s) + (\alpha^\beta(s))^{-1} H_1^\beta(s)P^\beta(s))^2 \right\} ds. \end{aligned} \quad (39)$$

Proof. We have already noted that the observation model (33) is equivalent to the original model (17). Also, as remarked after (34), to find the optimal filter for the original model it suffices to work only with the observation model (33). However, remark 5 tells us that we cannot directly apply the classical Kalman–Bucy filter formula. Instead, we use theorem 2, which is a more general result from non-linear filtering theory.

Our assumption of boundedness of all the coefficients, together with the fact that (X^β, Y^β) is Gaussian, implies that the observation function H^β (as in (36)) satisfies the energy condition (8). Thus theorem 2 is applicable. In fact, we need to apply the theorem for only two functions, viz., $f_1(x) = x$ and $f_2(x) = x^2$. This is easily done by identifying the various components appearing in (9).

It is immediate from (16) and (35) that

$$M^1(t) = X^\beta(t) - \int_0^t [a_0(s) + a_1(s)X^\beta(s) + a_2(s)Y^\beta(s)] ds,$$

is a martingale with $\langle M^1, W \rangle(t) = \int_0^t b(s)\lambda^\beta(s)ds$. Also, recalling (21) and using the fact that Y^β is (\mathcal{F}_t^β) adapted, the term $\pi(u, Hf_1) - \pi(u, H)\pi(u, f_1)$ appearing in the stochastic integral in (9) reduces to $H_1^\beta(u)P^\beta(u)$. Thus we get that $\widehat{X}^\beta(t)$ satisfies (38).

Similarly, to evaluate $\pi^\beta(t, f_2)$, we apply Ito's formula to (16) and use (35) to get that

$$M^2(t) = (X^\beta(t))^2 - \int_0^t [2a_0(s)X^\beta(s) + 2a_1(s)(X^\beta(s))^2 + 2a_2(s)X^\beta(s)Y^\beta(s) + (b(s))^2] ds$$

is a martingale with $\langle M^2, W \rangle(t) = \int_0^t 2b(s)\lambda^\beta(s)X^\beta(s)ds$. The cross quadratic variation gives the first term $(\pi^\beta(\cdot, \widehat{D}f_2))$ in the integrand of the stochastic integral in (9). Let $f_3(x) = x^3$. Then the second term can be easily seen to equal

$$H_1^\beta(u)\pi^\beta(u, f_3) - H_1^\beta(u)\widehat{X}^\beta(u)\pi^\beta(u, f_2).$$

Hence we get that $\pi^\beta(t, f_2)$ satisfies the SDE

$$\begin{aligned} \pi^\beta(t, f_2) &= E[X_0^2] \\ &+ \int_0^t [2a_0(s)\widehat{X}^\beta(s) + 2a_1(s)\pi^\beta(s, f_2) + 2a_2(s)\widehat{X}^\beta(s)Y^\beta(s) + (b(s))^2] ds \\ &+ \int_0^t [2b(s)\lambda^\beta(s)\widehat{X}^\beta(s) + (\alpha^\beta(s))^{-1} H_1^\beta(s)(\pi^\beta(s, f_3) - \widehat{X}^\beta(s)\pi^\beta(s, f_2))] dI^\beta(s). \end{aligned} \tag{40}$$

Now, another application of Ito's formula to (38) gives

$$\begin{aligned} (\widehat{X}^\beta(t))^2 &= [E(X_0)]^2 \\ &+ \int_0^t [2a_0(s)\widehat{X}^\beta(s) + 2a_1(s)(\widehat{X}^\beta(s))^2 + 2a_2(s)\widehat{X}^\beta(s)Y^\beta(s) \\ &\quad + (b(s)\lambda^\beta(s) + (\alpha^\beta(s))^{-1} H_1^\beta(s)P^\beta(s))^2] ds \\ &+ \int_0^t (2b(s)\lambda^\beta(s)\widehat{X}^\beta(s) + 2(\alpha^\beta(s))^{-1} H_1^\beta(s)P^\beta(s)\widehat{X}^\beta(s)) dI^\beta(s). \end{aligned} \tag{41}$$

Finally subtracting (41) from (40) we get

$$\begin{aligned}
 P^\beta(t) &= \pi^\beta(t, f_2) - (\widehat{X}^\beta(t))^2 \\
 &= V(X_0) + \int_0^t \left[2a_1(s)P^\beta(s) + (b(s))^2 \right. \\
 &\quad \left. - \left(b(s)\lambda^\beta(s) + (\alpha^\beta(s))^{-1} H_1^\beta(s)P^\beta(s) \right)^2 \right] ds \\
 &\quad + \int_0^t (\alpha^\beta(s))^{-1} H_1^\beta(s) \\
 &\quad \times (\pi^\beta(s, f_3) - \pi^\beta(s, f_2)\widehat{X}^\beta(s) - 2P^\beta(s)\widehat{X}^\beta(s)) dI^\beta(s). \quad (42)
 \end{aligned}$$

The integrand in the stochastic integral above is zero as expected, since

$$\begin{aligned}
 &\pi^\beta(s, f_3) - \pi^\beta(s, f_2)\widehat{X}^\beta(s) - 2P^\beta(s)\widehat{X}^\beta(s) \\
 &= \pi^\beta(s, f_3) + 2(\widehat{X}^\beta(s))^3 - 3\widehat{X}^\beta(s)\pi^\beta(s, f_2) \\
 &= E \left[(X^\beta(s) - \widehat{X}^\beta(s))^3 \mid \mathcal{F}_s^\beta \right] = 0.
 \end{aligned}$$

Thus (42) is the same as (39).

Equations (38) and (39) have the same form as the filtering equations in the classical linear filtering case (i.e., (6) and (7)). Uniqueness of solution is also similarly proved. (See proof of theorem 10.5.1 in Kallianpur 1980). \square

Remark 6. In this article only the case of signal and noise being real valued is considered. This is done purely for the sake of notational simplicity. The results can be extended to the general case when both the signal and the noise take values in arbitrary Euclidean spaces.

References

- Bhatt A G, Karandikar R L 1999 Characterization of the optimal filter: The non-Markov case. *Stochastics Stoch. Rep.* 66: 177–204
- Bhatt A G, Karandikar R L 2003 On filtering with Ornstein–Uhlenbeck process as noise. *J. Indian Stat. Assoc.* 41: 205–220
- Bhatt A G, Kallianpur G, Karandikar R L 1995 Uniqueness and robustness of solution of measure valued equations of nonlinear filtering. *Ann. Probab.* 23: 1895–1938
- Bhatt A G, Rajput B, Xiong J 2004 Zakai equation of nonlinear filtering with Ornstein–Uhlenbeck noise: Existence and uniqueness. *Stochastic processes and functional analysis: A volume of recent advances in honor of professor M M Rao* (ed.) A C Krinik, R J Swift (Lecture notes in pure and applied mathematics) (New York: Marcel Dekker) 238: 67–79
- Elliott R J 1982 *Stochastic calculus and applications* (New York: Springer)
- Fujisaki M, Kallianpur G, Kunita H 1972 Stochastic differential equations for the nonlinear filtering problem. *Osaka J. Math.* 9: 19–40
- Gawarecki L, Mandrekar V 2000 On the Zakai equation of filtering with Gaussian noise. *Stochastics in finite and infinite dimensions, Volume in honour of Gopinath Kallianpur* (eds) T Hida, R L Karandikar, H Kunita, B S Rajput, S Watanabe, J Xiong (Boston: Birkhauser) 145–151

- Gawarecki L, Mandrekar V 2003 Remark on “instrumentation problem” of A V Balakrishnan. *J. Indian Stat. Assoc.* 41: 275–283
- Kallianpur G 1980 *Stochastic filtering theory* (New York, Heidelberg, Berlin: Springer-Verlag)
- Kallianpur G, Karandikar R L 1988 *White noise theory of prediction, filtering and smoothing* (New York: Gordon and Breach)
- Kalman R E 1960 New results in linear filtering and prediction theory. *J. Basic Eng. ASME* 82: 33–45
- Kalman R E, Bucy R S 1961 A new approach to linear filtering and prediction problems. *J. Basic Eng. ASME* 83: 95–108
- Karandikar R L 1989 On Métivier–Pellaumail inequality, Emery topology and pathwise formulae in stochastic calculus. *Sankhya A* 51: 121–143
- Kouritzin M, Xiong J 2002 Nonlinear filtering: from Ornstein–Uhlenbeck to white noise (preprint)
- Kunita H 1993 Representation and stability of nonlinear filters associated with Gaussian noises. *Stochastic processes: A festschrift in honour of Gopinath Kallianpur* (eds) S Cambanis, J K Ghosh, R L Karandikar, P K Sen (New York: Springer-Verlag) pp 201–210
- Mandrekar V, Mandal P K 2000 A Bayes formula for Gaussian processes and its applications. *SIAM J. Control Optim.* 39: 852–871
- Métivier M 1982 *Semimartingales: A course on stochastic processes* (Berlin: Walter de Gruyter)
- Nelson E 1967 *Dynamical theories of Brownian motion* (Princeton: University Press)
- Rao C R 1985 *Linear statistical inference and its applications* (New Delhi: Wiley Eastern)