Chapter 2
Filtering of BSDE and FBSDE

In this chapter, we develop some filtering results for the solutions to BSDEs and FBSDEs, which play an important role in studying the optimal control with incomplete information. We first state a theorem on the stochastic filtering of a general stochastic process. The proof of that result can be found in Liptser and Shiyayev [49], so we omit it here. Then, we apply this result to the stochastic filtering for the solutions to BSDEs in Section 3.2 and to those for FBSDEs in Section 3.3.

2.1 Stochastic Filtering of Stochastic Processes

Consider a stochastic process

$$x(t) = x(0) + \int_0^t b(s)ds + m(t), \tag{2.1}$$

where $m(\cdot)$ is an $\mathscr{F}_t$-martingale, and $b(\cdot)$ is a stochastic process with

$$\mathbb{P}\left(\int_0^T |b(s)|ds < \infty\right) = 1.$$

Assume that $x(\cdot)$ is observed via an Itô process

$$Y(t) = Y(0) + \int_0^t h(s)ds + \int_0^t f(s,Y)dW(s).$$

Here $W(\cdot)$ is a 1-dimensional standard Brownian motion defined on $(\Omega, \mathscr{F}, (\mathscr{F}_t), P)$; $h:\ \Omega \times [0,T] \to \mathbb{R}$ and $f:\ [0,T] \times \mathbb{R} \to \mathbb{R}$ satisfy

$$\mathbb{P}\left(\int_0^T |h(s)|ds < \infty\right) = 1, \quad \mathbb{P}\left(\int_0^T |f(s,Y)|^2 ds < \infty\right) = 1$$

G. Wang et al., *An Introduction to Optimal Control of FBSDE with Incomplete Information*, SpringerBriefs in Mathematics,
https://doi.org/10.1007/978-3-319-79039-8_2

with $f(t,Y)$, $Y \in C([0,T];\mathbb{R})$, being $\mathscr{B}_t$-measurable for each $0 \le t \le T$. Furthermore, we assume that for any Y, Y_1, $Y_2 \in C([0,T];\mathbb{R})$, $0 \le t \le T$, there are three constants C, C_1, C_2 and a nondecreasing right continuous function $0 \le K(t) \le 1$ such that

$$|f(t,Y)|^2 \le C_1 \int_0^t |1+Y(s)|^2 dK(s) + C_2|1+Y(t)|^2$$

and

$$|f(t,Y_1) - f(t,Y_2)|^2 \le C_1 \int_0^t |Y_1(s) - Y_2(s)|^2 dK(s) + C_2|Y_1(t) - Y_2(t)|^2.$$

For a stochastic process $X(t)$, we call

$$\hat{X}(t) \equiv \mathbb{E}\left[X(t)|\mathscr{F}_t^Y\right]$$

the optimal filtering of $X(t)$ based on $Y(\cdot)$ up to time t, where $\mathscr{F}_t^Y = \sigma\{Y(s); 0 \le s \le t\}$. We now state the filtering equation of $x(t)$ given in (2.1), whose proof can be found in Liptser and Shiyayev (Theorem 8.1 of [49]).

Theorem 2.1. *Let*

$$\sup_{0 \le t \le T} \mathbb{E}x^2(t) < \infty, \quad \mathbb{E}\int_0^T [b^2(t) + h^2(t)]dt < \infty, \quad f^2(t,Y) \ge C > 0.$$

Then the optimal filtering of $x(t)$ satisfies

$$\hat{x}(t) = \hat{x}(0) + \int_0^t \hat{b}(s)ds + \int_0^t \left[\hat{D}(s) + \frac{\widehat{(xh)}(s) - \hat{x}(s)\hat{h}(s)}{f(s,Y)}\right] d\hat{W}(s),$$

where $(xh)(t) = x(t)h(t)$,

$$\hat{W}(t) = \int_0^t \frac{dY(s) - \hat{h}(s)ds}{f(s,Y)}$$

is a Brownian motion, and $D(t)$ is the stochastic process given by

$$D(t) = \frac{d\langle m, W\rangle_t}{dt}.$$

2.2 Stochastic Filtering for BSDE

Suppose that the stochastic process $(y(\cdot), z_1(\cdot), z_2(\cdot))$ is governed by a BSDE

$$\begin{aligned} y(t) = \xi &+ \int_t^T g(s, y(s), z_1(s), z_2(s))ds \\ &- \int_t^T z_1(s)dW_1(s) - \int_t^T z_2(s)dW_2(s), \end{aligned} \tag{2.2}$$

where $(W_1(\cdot), W_2(\cdot))$ is a 2-dimensional standard Brownian motion, $\xi \in \mathscr{L}^2_{\mathbb{F}}(\Omega;\mathbb{R})$, and $g : [0,T] \times \Omega \times \mathbb{R}^3 \to \mathbb{R}$ is a given function. Equation (2.2) admits a unique solution $(y(\cdot), z_1(\cdot), z_2(\cdot)) \in \mathscr{L}^2_{\mathbb{F}}(0,T;\mathbb{R}^3)$ under Conditions (Ha.1–Ha.2). Note that the solution involves $(z_1(\cdot), z_2(\cdot))$, which can be regarded as a control term to the equation such that the adapted solutions exist. Next, we assume that the observable process $Y(\cdot)$ is the Itô process given by

$$Y(t) = \int_0^t h(s)ds + \int_0^t f(s)dW_1(s), \tag{2.3}$$

where $f : [0,T] \to \mathbb{R}$ and $h : \Omega \times [0,T] \to \mathbb{R}$ are measurable mappings. The optimal nonlinear filtering is to compute $\hat{X}(t) = \mathbb{E}[X(t)|\mathscr{F}_t^Y]$, where $X = y$, z_1 and z_2. Since $z_i(\cdot)$ can be calculated by the Malliavin derivatives of $y(\cdot)$ with respect to $W_i(\cdot)$ $(i = 1,2)$ (see, e.g., El Karoui et al. [19]), we focus on the filtering of $y(\cdot)$. For any $t \in [0,T]$, we adopt the following notations for simplification of the presentation

$$\begin{aligned}
\hat{h}(t) &= \mathbb{E}[h(t)|\mathscr{F}_t^Y],\\
\hat{g}(t) &= \mathbb{E}[g(t, y(t), z_1(t), z_2(t))|\mathscr{F}_t^Y],\\
\widehat{(yh)}(t) &= \mathbb{E}[y(t)h(t)|\mathscr{F}_t^Y].
\end{aligned}$$

Also, we need the following assumption on the coefficients of (2.2) and (2.3).

(H2.1) *The function* $g(\cdot,0,0,0) \in \mathscr{L}^2_{\mathscr{F}}(0,T;\mathbb{R})$*, and* g *is Lipschitz in* (y, z_1, z_2) *uniformly for* $(\omega, t) \in \Omega \times [0,T]$*.* f *is bounded and deterministic, and* f^{-1} *is also bounded.* h *is in* $\mathscr{L}^2_{\mathbb{F}}(0,T;\mathbb{R})$*.*

We now state the main result of this section, which plays an important role in the study of incomplete information stochastic control for BSDE.

Theorem 2.2. *Under (H2.1), the optimal filtering* $\hat{y}(\cdot)$ *is governed by*

$$\begin{aligned}
\hat{y}(t) &= \mathbb{E}[\xi|\mathscr{F}_T^Y] + \int_t^T \hat{g}(s)ds\\
&\quad - \int_t^T \left\{ \hat{z}_1(s) + \frac{1}{f(s)} \left[\widehat{(yh)}(s) - \hat{y}(s)\hat{h}(s) \right] \right\} d\hat{W}(s),
\end{aligned} \tag{2.4}$$

where

$$\hat{W}(t) = \int_0^t \frac{1}{f(s)} \left(dY(s) - \hat{h}(s)ds \right) \tag{2.5}$$

is a 1-dimensional standard Brownian motion defined on $(\Omega, \mathscr{F}^Y, (\mathscr{F}_t^Y)_{0\le t\le T}, \mathbb{P})$*.*

Proof. Equation (2.2) admits a unique solution $(y(\cdot), z_1(\cdot), z_2(\cdot)) \in \mathscr{L}^2_{\mathbb{F}}(0,T;\mathbb{R}^3)$, so $y(\cdot)$ can be rewritten as an Itô process as follows:

$$\begin{aligned}
y(t) &= y(0) - \int_0^t g(s, y(s), z_1(s), z_2(s))ds\\
&\quad + \int_0^t z_1(s)dW_1(s) + \int_0^t z_2(s)dW_2(s).
\end{aligned} \tag{2.6}$$

Now (2.6) and (2.3) can be regarded as the state equation and the observation equation, respectively. Using Theorem 2.1, we have

$$\hat{y}(t)=\hat{y}(0)-\int_0^t \hat{g}(s)ds+\int_0^t \left\{\hat{z}_1(s)+\frac{1}{f(s)}\left[\widehat{(yh)}(s)-\hat{y}(s)\hat{h}(s)\right]\right\}d\hat{W}(s), \quad (2.7)$$

where $\hat{W}(\cdot)$ is given by (2.5). Similarly, we have

$$\hat{y}(T)=\hat{y}(0)-\int_0^T \hat{g}(s)ds+\int_0^T \left\{\hat{z}_1(s)+\frac{1}{f(s)}\left[\widehat{(yh)}(s)-\hat{y}(s)\hat{h}(s)\right]\right\}d\hat{W}(s). \quad (2.8)$$

Subtracting (2.8) from (2.7), we obtain that

$$\hat{y}(t)=\hat{y}(T)+\int_t^T \hat{g}(s)ds-\int_t^T \left\{\hat{z}_1(s)+\frac{1}{f(s)}\left[\widehat{(yh)}(s)-\hat{y}(s)\hat{h}(s)\right]\right\}d\hat{W}(s).$$

The verification of the terminal condition $\hat{y}(T)=\mathbb{E}[\xi|\mathscr{F}_T^Y]$ is trivial. Thus the proof is completed. □

The following result is an immediate consequence of Theorem 2.2.

Corollary 2.1. *Under (H2.1), if* $g_1:[0,T]\times\mathbb{R}^2\to\mathbb{R}$ *and* $h:\Omega\times[0,T]\to\mathbb{R}$ *satisfy*

$$\hat{g}(t)=g_1(t,\hat{y}(t),\hat{z}_1(t)) \text{ and } \widehat{(yh)}(t)=\hat{y}(t)\hat{h}(t),$$

respectively, then the optimal filtering $(\hat{y}(\cdot),\hat{z}_1(\cdot))$ *is a solution of the following backward filtering equation:*

$$\hat{y}(t)=\mathbb{E}[\xi|\mathscr{F}_T^Y]+\int_t^T g_1(s,\hat{y}(s),\hat{z}_1(s))ds-\int_t^T \hat{z}_1(s)d\hat{W}(s),$$

where

$$\hat{W}(t)=\int_0^t \frac{1}{f(s)}\left(dY(s)-\hat{h}(s)ds\right)$$

is a 1-dimensional standard Brownian motion defined on $(\Omega,\mathscr{F}^Y,(\mathscr{F}_t^Y)_{0\le t\le T},\mathbb{P})$.

In what follows, suppose that the conditional probability distribution of $y(t)$ based on $\mathscr{F}_t^Y$ has the density

$$\psi(t,x)=\frac{d\mathbb{P}(y(t)\le x|\mathscr{F}_t^Y)}{dx}, \quad (t,x)\in[0,T]\times\mathbb{R},$$

which is measurable in (t,x,ω). We proceed to deriving the equation satisfied by this conditional density. Note that the nonlinear filtering $\hat{y}(\cdot)$ can be represented by

$$\hat{y}(t)=\mathbb{E}[y(t)|\mathscr{F}_t^Y]=\int_{-\infty}^{\infty} x\psi(t,x)dx.$$

Next, we assume that the observation function $h(s)$ in (2.3) depends on the signal in a deterministic way, namely, we abuse notation a bit, there is a function $h: [0,T] \times \mathbb{R} \to \mathbb{R}$ such that observation equation (2.3) is replaced by

$$Y(t) = \int_0^t h(s, y(s))ds + \int_0^t f(s)dW_1(s). \tag{2.9}$$

We introduce the following assumptions:

(H2.2) *The function (to be used in Theorem 2.3 below as test function for filtering)* $\varphi(\cdot) : \mathbb{R} \to \mathbb{R}$ *and its derivatives up to order 2 are uniformly bounded.*

(H2.3) *The function* $\varphi(\cdot) : \mathbb{R} \to \mathbb{R}$ *has compact support.*

(H2.4) *The functions* $g : [0,T] \times \mathbb{R}^3 \to \mathbb{R}$ *and* $h : [0,T] \times \mathbb{R} \to \mathbb{R}$ *are Borel measurable, and h is Lipschitz in* $x \in \mathbb{R}$ *uniformly for* $t \in [0,T]$.

(H2.5) *The solution of (2.2) satisfies* $z_i(t) = g_i(t, y(t))$, *where* $g_i : [0,T] \times \mathbb{R} \to \mathbb{R}$ $(i = 1,2)$ *are also Borel measurable.*

(H2.6) *The partial derivatives*

$$[\psi(t,x)g(t,x,y,z)]_x, \qquad [yh(t,x)]_x,$$

$$[\psi(t,x)(h(t,x) - \hat{h}(t,x))g(t,x,y,z)]_x,$$

$$[(y^2+z^2)\psi(t,x)]_{xx} \text{ and } [(y^2+z^2)(h(t,x) - \hat{h}(t,x))\psi(t,x)]_{xx}$$

exist.

(H2.7)

$$\int_0^T \int_{-\infty}^{\infty} |\varphi(x)\mathscr{L}^*\psi(t,x)|dxdt < \infty,$$

$$\mathbb{E}\int_0^T \int_{-\infty}^{\infty} \varphi^2(x) \left[\psi(t,x)\left(h(t,x) - \hat{h}(t,x)\right) + \mathscr{N}^*\psi(t,x)\right]^2 dxdt < \infty$$

with the notations

$$\mathscr{L}^*\psi(t,x) = -[\psi(t,x)g(t,x,y,z)]_x - \frac{1}{2}[\psi(t,x)(y^2+z^2)]_{xx}$$

and

$$\mathscr{N}^*\psi(t,x) = -[\psi(t,x)y]_x.$$

Note that (H2.2), (H2.3), (H2.4), (H2.6), and (H2.7) are standard in the theory of nonlinear filtering, while (H2.5) is reasonable under some constraints on ξ and g. We give an example below for which (H2.5) holds.

Example 2.1. We consider the BSDE

$$\begin{aligned} y(t) = \xi + \int_t^T &(a(s)y(s) + b_1(s)z_1(s) + b_2(s)z_2(s))\, ds \\ &- \int_t^T z_1(s)dW_1(s) - \int_t^T z_2(s)dW_2(s), \end{aligned}$$

where $a(\cdot)$, $b_i(\cdot)$ $(i=1,2)$ are bounded and deterministic, and

$$\xi = \exp\left(\sum_{i=1}^{2}\int_0^T f_i(t)dW_i(t)\right).$$

Then, (H2.5) holds.

Proof. It is easy to see that the BSDE has a unique solution $(y(\cdot), z_1(\cdot), z_2(\cdot)) \in \mathscr{L}^2_{\mathbb{F}}(0,T;\mathbb{R}^3)$. In fact, the solution can be represented as

$$y(t) = \mathbb{E}[\xi x(T)|\mathscr{F}_t]$$

with

$$\begin{aligned} x(T) = \ &\exp\left\{\int_0^T \left[a(t) - \frac{1}{2}\left(b_1^2(t)+b_2^2(t)\right)\right]dt \right. \\ &\left. + \int_0^T b_1(t)dW_1(t) + \int_0^T b_2(t)dW_2(t)\right\}. \end{aligned}$$

According to Proposition A.1, $z_i(\cdot)$ is expressed by

$$\begin{aligned} z_i(t) &= D_t^{(W_i)}\mathbb{E}[\xi x(T)|\mathscr{F}_t] \\ &= \mathbb{E}\left[x(T)D_t^{(W_i)}\xi|\mathscr{F}_t\right] + b_i(t)y(t), \end{aligned}$$

where $D_t^{(W_i)}\eta$ stands for the Malliavin derivative of η with respect to $W_i(\cdot)$ $(i=1,2)$. Note that $D_t^{(W_i)}\xi = \xi f_i(t)$. Thus, $z_i(t) = (f_i(t)+b_i(t))y(t)$, and hence (H2.5) holds. □

Theorem 2.3. *(i) If (H2.1), (H2.2), and (H2.5) hold, then the optimal filtering $\hat{\varphi}(y(t))$ satisfies*

$$\begin{aligned} \hat{\varphi}(y(t)) = \ &\mathbb{E}[\varphi(\xi)|\mathscr{F}_T^Y] + \int_t^T \widehat{\mathscr{L}\varphi}(y(s))ds \\ &- \int_t^T \left\{\widehat{\mathscr{N}\varphi}(y(s)) + \frac{1}{f(s)}\left[\widehat{\varphi h}(s,y(s)) - \hat{\varphi}(y(s))\hat{h}(s,y(s))\right]\right\}d\hat{W}(s), \end{aligned} \tag{2.10}$$

where

$$\begin{aligned} \widehat{\mathscr{L}\varphi}(y(s)) &= \mathbb{E}[\mathscr{L}\varphi(y(s))|\mathscr{F}_s^Y], \\ \widehat{\mathscr{N}\varphi}(y(s)) &= \mathbb{E}[\mathscr{N}\varphi(y(s))|\mathscr{F}_s^Y], \\ \widehat{\varphi h}(s,y(s)) &= \mathbb{E}[\varphi(y(s))h(s,y(s))|\mathscr{F}_s^Y], \end{aligned}$$

with

$$\begin{aligned} \mathscr{L}\varphi(y(s)) &= \varphi_x(y(s))g(s,y(s),z_1(s),z_2(s)) - \frac{1}{2}\varphi_{xx}(y(s))(z_1^2(s)+z_2^2(s)), \\ \mathscr{N}\varphi(y(s)) &= \varphi_x(y(s))z_1(s), \end{aligned}$$

and

$$\hat{W}(s) = \int_0^s \frac{1}{f(t)} \left(dY(t) - \hat{h}(t, y(t))\right) dt.$$

(ii) If (H2.1), (H2.3), (H2.4), (H2.6), (H2.7), (H2.8), (H2.9), and (H2.10) hold, then the conditional density $\psi(t,x)$ *satisfies*

$$\begin{aligned}\psi(t,x) = \psi(T,x) + \int_t^T \mathscr{L}^* \psi(s,x) ds - \int_t^T \Big[\mathscr{N}^* \psi(s,x) \\ + \frac{1}{f(s)} \psi(s,x) \left(h(s,x) - \int_{-\infty}^{\infty} h(s,x)\psi(s,x) dx\right)\Big] d\hat{W}(s).\end{aligned} \tag{2.11}$$

Proof. (i) Applying Itô's formula to $\varphi(y(t))$, we have

$$\begin{aligned}\varphi(y(t)) =& \varphi(\xi) + \int_t^T \mathscr{L}\varphi(y(s)) ds \\ & - \int_t^T \varphi_x(y(s)) z_1(s) dW_1(s) - \int_t^T \varphi_x(y(s)) z_2(s) dW_2(s).\end{aligned}$$

By the uniqueness of $(y(\cdot), z_1(\cdot), z_2(\cdot))$, Theorems 2.1 and 2.2, the nonlinear filtering equation (2.10) is obtained directly.

(ii) Due to (H2.5), (2.10) can be rewritten as

$$\int_{-\infty}^{\infty} \varphi(x)\psi(t,x) dx = \int_{-\infty}^{\infty} \varphi(x)\psi(T,x) dx + I - II, \tag{2.12}$$

where

$$\begin{aligned}I &= \int_t^T \int_{-\infty}^{\infty} \mathscr{L}\varphi(x)\psi(s,x) dx ds, \\ II &= \int_t^T \int_{-\infty}^{\infty} \left[\mathscr{N}\varphi(x) + \frac{1}{f(s)} \varphi(x) \left(h(s,x) - \hat{h}(s,x)\right)\right] \psi(s,x) dx d\hat{W}(s).\end{aligned}$$

It follows from integration by parts and Fubini's theorem that

$$I = \int_{-\infty}^{\infty} \int_t^T \varphi(x) \mathscr{L}^* \psi(s,x) ds dx.$$

Similarly,

$$II = \int_{-\infty}^{\infty} \int_t^T \varphi(x) \left[\mathscr{N}^* \psi(s,x) + \frac{1}{f(s)} \psi(s,x) \left(h(s,x) - \hat{h}(s,x)\right)\right] d\hat{W}(s) dx.$$

Substituting the above two identities into (2.12), we obtain

$$
\begin{aligned}
&\int_{-\infty}^{\infty} \varphi(x)\psi(t,x)dx \\
&= \int_{-\infty}^{\infty} \varphi(x)\left\{\psi(T,x) + \int_t^T \mathscr{L}^*\psi(s,x)ds \right.\\
&\quad \left. - \int_t^T \left[\mathscr{N}^*\psi(s,x) + \frac{1}{f(s)}\psi(s,x)\left(h(s,x)-\hat{h}(s,x)\right)\right] d\hat{W}(s)\right\} dx.
\end{aligned}
$$

The arbitrariness of $\varphi(\cdot)$ implies (2.11). The proof is then completed. □

We emphasize that (2.11) is a new kind of backward stochastic partial differential equation (BSPDE). Since the noise term is very complicated, it is not easy to prove the existence and uniqueness of solution to the equation. We pose it here as an open problem.

2.3 Stochastic Filtering for FBSDE

In this section, we first introduce the four-step scheme in solving FBSDEs. As an example to this scheme, we then consider an optimization problem. To obtain an explicit solution, we will apply Girsanov's transformation to convert it to an LQ control problem which can be solved in terms of an FBSDE without control variable. Finally, we study the stochastic filtering problem for this FBSDE based on a linear observation equation.

Consider a fully coupled FBSDE

$$
\begin{cases}
dx(t) = b(t,x(t),y(t),z(t))dt + \sigma(t,x(t),y(t),z(t))dW(t), \\
-dy(t) = g(t,x(t),y(t),z(t))dt - z(t)dW(t), \\
x(0) = x_0, \quad y(T) = f(x(T)),
\end{cases}
\tag{2.13}
$$

where $b, g : [0,T]\times\mathbb{R}^{n+n+n\times m} \to \mathbb{R}^n$, $\sigma : [0,T]\times\mathbb{R}^{n+n+n\times m} \to \mathbb{R}^{n\times m}$, $f : \Omega\times\mathbb{R}^n \to \mathbb{R}^n$ are continuous functions; $W(\cdot)$ is an m-dimensional standard Brownian motion defined on the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0\le t\le T}, \mathbb{P})$; $\mathscr{F}_t$ is the natural filtration generated by $W(\cdot)$, and $x_0 \in \mathbb{R}^n$. Under Conditions (Ha.3-Ha.4), there is a unique solution $(x(\cdot), y(\cdot), z(\cdot))$ to (2.13). Furthermore, using the four-step scheme (see, e.g., Yong and Zhou [109]), $y(\cdot)$ and $z(\cdot)$ can be expressed as a functional of $x(\cdot)$, respectively. Indeed,

$$
\begin{aligned}
y(t) &= U(t,x(t)), \\
z(t) &= Z(t,x(t),U(t,x(t)),U_x(t,x(t))),
\end{aligned}
$$

where U, Z, and $x(\cdot)$ satisfy

$$\begin{cases} U_t^\ell + \frac{1}{2} tr\left[U_{xx}^\ell(\sigma\sigma^\top)(t,X,U,Z(t,X,U,U_x))\right] + \langle b(t,X,U,U_x), U_X^\ell\rangle \\ \quad + g^\ell(t,X,U,U_x) = 0, \quad (t,X) \in (0,T)\times\mathbb{R}^n, \quad \ell = 1,\cdots,n, \\ U(T,X) = f(X), \quad X \in \mathbb{R}^n, \end{cases}$$

$$Z(t,X,\bar{X},\tilde{X}) = \tilde{X}\sigma(t,X,\bar{X},Z(t,X,\bar{X},\tilde{X})), \quad (t,X,\bar{X},\tilde{X}) \in [0,T]\times\mathbb{R}^{n+n+n\times n},$$

and

$$\begin{cases} dx(t) = \tilde{b}(t,x(t))dt + \tilde{\sigma}(t,x(t))dW(t), \\ x(0) = x_0 \end{cases}$$

with

$$\tilde{b}(t,X) = b(t,X,U(t,X),Z(t,X,U(t,X),U_x(t,X))),$$
$$\tilde{\sigma}(t,X) = \sigma(t,X,U(t,X),Z(t,X,U(t,X),U_x(t,X))).$$

According to the relationship between $(y(\cdot),z(\cdot))$, and $x(\cdot)$, it suffices to compute the optimal filtering of $x(\cdot)$. The detailed arguments are omitted due to the page limit.

To elaborate the above analysis, we present a simple example on filtering of stochastic Hamiltonian system arising from a stochastic control problem. Specifically, let us consider a 1-dimensional control system, whose evolution is described by

$$\begin{cases} dx(t) = (A(t)x(t) + B(t)v(t))dt + C_1(t)d\tilde{W}_1(t) + C_2(t)d\tilde{W}_2(t), \\ x(0) = x_0, \end{cases} \tag{2.14}$$

where $v(\cdot)$ is an element of the set

$$\mathscr{U}_{ad} = \Big\{ v(\cdot) \Big| v(t) \text{ is an } \mathscr{F}_t\text{-adapted process valued in } \mathbb{R} \text{ and satisfies } \mathbb{E}\int_0^T v^4(t)dt < \infty \Big\}.$$

Suppose that the cost functional is given by

$$J(v(\cdot)) = \bar{y}(0),$$

where $\bar{y}(\cdot)$ is a solution to the BSDE

$$\begin{cases} -d\bar{y}(t) = (a(t)x^2(t) + b(t)\bar{y}(t) + f_1(t)\bar{z}_1(t) + f_2(t)\bar{z}_2(t) + c(t)v^2(t))dt \\ \qquad\qquad - \bar{z}_1(t)d\tilde{W}_1(t) - \bar{z}_2(t)d\tilde{W}_2(t), \\ \bar{y}(T) = x^2(T). \end{cases} \tag{2.15}$$

Here $a(\cdot) \geq 0$, $c(\cdot) \geq \varepsilon > 0$, $A(\cdot)$, $B(\cdot)$, $C_1(\cdot)$, $C_2(\cdot)$, $f_1(\cdot)$, and $f_2(\cdot)$ are uniformly bounded, deterministic functions. For any $v(\cdot) \in \mathscr{U}_{ad}$, it is easy to see that

$$\mathbb{E}\bar{y}^2(T) < \infty,$$

and thus, there exists a unique solution to (2.14) and (2.15), respectively. Since the drift term in (2.15) contains $(\bar{z}_1(\cdot), \bar{z}_2(\cdot))$, it causes us some trouble to express the cost functional $J(v(\cdot))$. To simplify it, we define a probability measure Q on the space $(\Omega, \mathscr{F})$ by

$$\frac{dQ}{dP} = \exp\left\{ \int_0^T f_1(t) d\tilde{W}_1(t) + \int_0^T f_2(t) d\tilde{W}_2(t) - \frac{1}{2}\int_0^T (f_1^2(t) + f_2^2(t)) dt \right\}.$$

It follows from Girsanov's theorem that $(W_1(\cdot), W_2(\cdot))$ defined by

$$W_1(t) = \tilde{W}_1(t) - \int_0^t f_1(s) ds \text{ and } W_2(t) = \tilde{W}_2(t) - \int_0^t f_2(s) ds$$

is a 2-dimensional Brownian motion defined on the stochastic basis $(\Omega, \mathscr{F}, (\mathscr{F}_t), Q)$. Then we can rewrite (2.14) and (2.15) as

$$\begin{cases} dx(t) = (A(t)x(t) + B(t)v(t) + C_1(t)f_1(t) + C_2(t)f_2(t))dt \\ \qquad\qquad + C_1(t)dW_1(t) + C_2(t)dW_2(t), \\ x(0) = x_0, \end{cases} \tag{2.16}$$

$$\begin{cases} -d\bar{y}(t) = (a(t)x^2(t) + b(t)\bar{y}(t) + c(t)v^2(t))dt \\ \qquad\qquad - \bar{z}_1(t)dW_1(t) - \bar{z}_2(t)dW_2(t), \\ \bar{y}(T) = x^2(T). \end{cases} \tag{2.17}$$

Integrating on both sides of (2.17), we get

$$J(v(\cdot)) = \bar{y}(0) = \mathbb{E}_Q\left[\int_0^T e^{\int_0^t b(s)ds}(a(t)x^2(t) + c(t)v^2(t))dt + e^{\int_0^T b(t)dt} x^2(T) \right].$$

Then minimizing the cost functional subject to $v(\cdot) \in \mathscr{U}_{ad}$ and (2.16) formulates a complete information LQ optimal control problem. Since the drift term in (2.16) contains the deterministic function $C_1(\cdot)f_1(\cdot) + C_2(\cdot)f_2(\cdot)$, the classical technique of completing squares cannot be used directly to solve the control problem. However, stochastic maximum principle (see, e.g., Chapters 3–5) provides an alternative tool. According to the maximum principle, we derive the desired optimal control

$$u(t) = -\frac{1}{2}B(t)c^{-1}(t)e^{-\int_0^t b(s)ds}y(t),$$

where the adjoint process $y(\cdot)$ satisfies a Hamiltonian system

$$\begin{cases} dx(t) = \Big(A(t)x(t) - \frac{1}{2}B^2(t)c^{-1}(t)e^{-\int_0^t b(s)ds}y(t) + C_1(t)f_1(t) + C_2(t)f_2(t)\Big)dt \\ \qquad\qquad + C_1(t)dW_1(t) + C_2(t)dW_2(t), \\ -dy(t) = \Big(2a(t)e^{\int_0^t b(s)ds}x(t) + A(t)y(t)\Big)dt - z_1(t)dW_1(t) - z_2(t)dW_2(t), \\ x(0) = x_0, \quad y(T) = 2e^{\int_0^T b(s)ds}x(T). \end{cases}$$

It follows from Theorem A.3 that there is a unique solution to the above equation. Suppose that $(x(\cdot), y(\cdot), z_1(\cdot), z_2(\cdot))$ cannot be observed directly; however, we can observe a noisy process $Y(\cdot)$ related to $x(\cdot)$, whose dynamic is described by

$$\begin{cases} dY(t) = (D(t)x(t) + F(t)Y(t) + f_2(t)H(t))dt + H(t)dW_2(t), \\ Y(0) = 0, \end{cases} \tag{2.18}$$

where $D(\cdot)$, $F(\cdot)$, $H(\cdot)$, and $H^{-1}(\cdot)$ are uniformly bounded, deterministic functions. Obviously, there exists a unique solution for (2.18).

We now study the filtering $(\hat{x}(t), \hat{y}(t), \hat{z}_1(t), \hat{z}_2(t))$ of $(x(t), y(t), z_1(t), z_2(t))$ with respect to the observation $Y(\cdot)$ up to time t, i.e., we want to derive the explicit expressions for

$$\begin{aligned} &\hat{x}(t) = \mathbb{E}_Q[x(t)|\mathscr{F}_t^Y], \quad \hat{y}(t) = \mathbb{E}_Q[y(t)|\mathscr{F}_t^Y], \\ &\hat{z}_1(t) = \mathbb{E}_Q[z_1(t)|\mathscr{F}_t^Y], \quad \hat{z}_2(t) = \mathbb{E}_Q[z_2(t)|\mathscr{F}_t^Y] \end{aligned} \tag{2.19}$$

and their square error estimates, where

$$\mathscr{F}_t^Y = \sigma\{Y(s); 0 \le s \le t\}.$$

The method used here is first to look for the relationship between $x(\cdot)$ and $(y(\cdot), z(\cdot))$ by the four-step scheme, then to compute $(\hat{x}(\cdot), \hat{y}(\cdot), \hat{z}_1(\cdot), \hat{z}_2(\cdot))$ by traditional filtering theory for SDE.

Set $y(t) = U(t, x(t))$. It follows from the four-step scheme that $z_1(t)$ and $z_2(t)$ can be written as

$$z_1(t) = C_1(t)U_x(t, x(t)), \quad z_2(t) = C_2(t)U_x(t, x(t)), \tag{2.20}$$

where $U(t,x)$ is a classical solution of the PDE

$$\begin{cases} U_t(t,x) + \mathscr{L}U(t,x) + 2a(t)e^{\int_0^t b(s)ds}x + A(t)U(t,x) = 0, \\ U(T,x) = 2e^{\int_0^T b(s)ds}x, \end{cases} \tag{2.21}$$

with

$$\begin{aligned}\mathscr{L}U(t,x) = & \frac{1}{2}(C_1^2(t)+C_2^2(t))U_{xx}(t,x) \\ & +\Big(A(t)x-\frac{1}{2}B^2(t)c^{-1}(t)e^{-\int_0^t b(s)ds}U(t,x) \\ & +C_1(t)f_1(t)+C_2(t)f_2(t)\Big)U_x(t,x).\end{aligned}$$

Noticing the terminal condition of (2.21), we set

$$U(t,x)=\Pi(t)x+\pi(t),$$

where $\Pi(\cdot)$ and $\pi(\cdot)$ satisfy

$$\begin{cases}\dot{\Pi}(t)+2A(t)\Pi(t)-\frac{1}{2}B^2(t)c^{-1}(t)e^{-\int_0^t b(s)ds}\Pi^2(t)+2a(t)e^{\int_0^t b(s)ds}=0, \\ \Pi(T)=2e^{\int_0^T b(s)ds},\end{cases} \tag{2.22}$$

and

$$\begin{cases}\dot{\pi}(t)+\Big(A(t)-\frac{1}{2}B^2(t)c^{-1}(t)e^{-\int_0^t b(s)ds}\Pi(t)\Big)\pi(t) \\ \quad +(C_1(t)f_1(t)+C_2(t)f_2(t))\Pi(t)=0, \\ \pi(T)=0,\end{cases} \tag{2.23}$$

respectively. From the classical ODE theory, we know that there exists a unique solution for (2.22) and (2.23), respectively. Combining (2.20) with (2.23), we get

$$y(t)=\Pi(t)x(t)+\pi(t), \quad z_1(t)=C_1(t)\Pi(t), \quad z_2(t)=C_2(t)\Pi(t), \tag{2.24}$$

where $x(\cdot)$ satisfies

$$\begin{cases}dx(t)=\Big[\Big(A(t)-\frac{1}{2}B^2(t)c^{-1}(t)\Pi(t)e^{-\int_0^t b(s)ds}\Big)x(t) \\ \qquad +C_1(t)f_1(t)+C_2(t)f_2(t)-\frac{1}{2}B^2(t)c^{-1}(t)\pi(t)e^{-\int_0^t b(s)ds}\Big]dt \\ \qquad +C_1(t)dW_1(t)+C_2(t)dW_2(t), \\ x(0)=x_0.\end{cases}$$

Obviously,

$$\hat{z}_1(t)=C_1(t)\Pi(t), \quad \hat{z}_2(t)=C_2(t)\Pi(t). \tag{2.25}$$

Then we only need to compute $\hat{x}(t)$ and $\hat{y}(t)$. Let $P(t)=\mathbb{E}_Q(x(t)-\hat{x}(t))^2$ be the square error of the estimate $\hat{x}(t)$. From the fact that $(x(t)-\hat{x}(t))\perp\mathscr{F}_t^Y$ and $x(t)-\hat{x}(t)$ is Gaussian, we know that $x(t)-\hat{x}(t)$ is independent of $\mathscr{F}_t^Y$. So

$$
\begin{aligned}
P(t) &= \mathbb{E}_Q(x(t)-\hat{x}(t))^2 \\
&= \mathbb{E}_Q[(x(t)-\hat{x}(t))^2|\mathscr{F}_t^Y].
\end{aligned}
$$

Thanks to Theorem 2.1, we obtain

$$
\begin{cases}
d\hat{x}(t) = \Big[\Big(A(t) - \dfrac{1}{2}B^2(t)c^{-1}(t)\Pi(t)e^{-\int_0^t b(s)ds}\Big)\hat{x}(t) + C_1(t)f_1(t) \\
\qquad\quad + C_2(t)f_2(t) - \dfrac{1}{2}B^2(t)c^{-1}(t)\pi(t)e^{-\int_0^t b(s)ds}\Big]dt \\
\qquad\quad + (C_2(t)+D(t)H^{-1}(t)P(t))d\bar{W}(t), \\
\hat{x}(0) = x_0,
\end{cases}
\tag{2.26}
$$

$$
\begin{cases}
\dot{P}(t) - 2\Big(A(t) - \dfrac{1}{2}B^2(t)c^{-1}(t)\Pi(t)e^{-\int_0^t b(s)ds}\Big)P(t) \\
+ (C_2(t)+D(t)H^{-1}(t)P(t))^2 - C_1^2(t) - C_2^2(t) = 0, \\
P(0) = 0,
\end{cases}
$$

where

$$
\begin{aligned}
\bar{W}(t) &= \int_0^t H^{-1}(s)(dY(s) - D(s)\hat{x}(s) - F(s)Y(s) - f_2(s)H(s))ds \\
&= W_2(t) + \int_0^t D(s)H^{-1}(s)(x(s)-\hat{x}(s))ds
\end{aligned}
$$

is an observable standard Brownian motion defined on $(\Omega, \mathscr{F}^Y, (\mathscr{F}_t^Y), \mathbb{P})$. Furthermore, taking conditional expectations on both sides of (2.24), we get

$$
\hat{y}(t) = \Pi(t)\hat{x}(t) + \pi(t), \tag{2.27}
$$

where $\hat{x}(\cdot)$ is the solution of (2.26). Then we have

Proposition 2.1. *The stochastic filtering process $(\hat{x}(\cdot), \hat{y}(\cdot), \hat{z}_1(\cdot), \hat{z}_2(\cdot))$ of the state process $(x(\cdot), y(\cdot), z_1(\cdot), z_2(\cdot))$ based on the observation process $Y(\cdot)$ is given by (2.26), (2.27), and (2.25).*

2.4 Notes

The filtering problem is to obtain the best linear estimate $\hat{x}(t)$ of an unobservable state $x(t)$ based on the noisy observation data $Y_0^t = \{Y(s); 0 \le s \le t\}$ related to the state. If $\mathbb{E}x^2(t) < \infty$, then the best estimate $\hat{x}(t)$ of $x(t)$ is equivalent to finding the conditional expectation

$$
\hat{x}(t) = \mathbb{E}[x(t)|\mathscr{F}_t^Y]
$$

with $\mathscr{F}_t^Y = \sigma\{Y(s); 0 \leq s \leq t\}$. When the estimate depends linearly on the observations, we call it the linear filtering. Otherwise, it is referred to as the nonlinear filtering. In the linear filtering theory, the most celebrated result is the linear quadratic estimation, also known as the Kalman–Bucy filtering. The filtering was discovered and was developed by Rudolf E. Kalman and Richard S. Bucy during the Cold War between North American Treaty Organization and Warsaw Treaty Organization. The Kalman–Bucy filtering works recursively and runs in real time, and thus, it has numerous applications in the fields of aerospace, telecommunication, economics, and so on. As far as we know, the most famous one among these applications is the Apollo Project, where the Kalman–Bucy filtering was used to estimate the trajectories of manned spaceship going to Moon and back.

The filtering equation provided in Theorem 2.1 is one fundamental equation of the nonlinear filtering theory. Lots of known filtering results can be deduced from the equation, say, the Kalman–Bucy filtering. The deduction of the fundamental equation follows the innovation process method, proposed originally by Bode and Shannon [11], whose modern form was presented first by Kailath [34] and Kailath and Frost [35]. Along this line, we get the equation of the conditional probability density $\psi(t,x)$ in Theorem 2.3, which is a new kind of nonlinear backward SPDE. Historically, a similar SPDE was derived early by Stratonovich [77] and Kushner [40] when they studied the condition probability density of a forward SDE. The innovation method achieved its culmination with the famous work of Fujisaki et al. [25]. The filtering equation is called the Kushner–Stratonovich equation or the Kushner–FKK equation. Almost at the same time, Duncan [18], Mortensen [56], and Zakai [112] studied the nonlinear filtering problem by virtue of the Kallianpur–Striebel formula and the unnormalized filtering. They obtained a linear SPDE of the unnormalized filtering, which is called the Duncan–Mortensen–Zakai equation, or, simply, Zakai's equation. See, e.g., Bensoussan [6] and Xiong [104] for a systematic account.

Most results of Section 3.2 are taken from Wang et al. [95]. If the diffusion coefficients of (2.14) in Section 3.3 contain the state or the control, then (2.14) is not Gaussian in general. Consequently, it is difficult to obtain an explicit filtering equation. The filtering example is taken from Wang and Wu [84], where some applications to optimal control with partially observed information are also studied.