

## KALMAN–BUCY FILTERING AND MINIMUM MEAN SQUARE ESTIMATOR UNDER UNCERTAINTY\*

SHAOLIN JI<sup>†</sup>, CHUILLIU KONG<sup>‡</sup>, CHUANFENG SUN<sup>§</sup>, AND JI-FENG ZHANG<sup>¶</sup>

**Abstract.** In this paper, we study a generalized Kalman–Bucy filtering problem under uncertainty. The drift uncertainty for both signal process and observation process is considered, and the attitude to uncertainty is characterized by a convex operator (convex risk measure). The optimal filter or the minimum mean square estimator (MMSE) is calculated by solving the minimum mean square estimation problem under a convex operator. In the first part of this paper, this estimation problem is studied under  $g$ -expectation which is a special convex operator. For this case, we prove that there exists a worst-case prior  $P^{\theta^*}$ . Based on this  $P^{\theta^*}$  we obtained the Kalman–Bucy filtering equation under  $g$ -expectation. In the second part of this paper, we study the minimum mean square estimation problem under general convex operators. The existence and uniqueness results of the MMSE are deduced.

**Key words.** Kalman–Bucy filtering, minimum mean square estimator, drift uncertainty, convex operator, minimax theorem, backward stochastic differential equation

**AMS subject classifications.** 93E11, 62M20, 60G35, 62F86

**DOI.** 10.1137/20M137954X

**1. Introduction.** It is well-known that Kalman–Bucy filtering theory was originally derived from Kalman and Bucy [27] and is the foundation of modern filtering theory; see these classic monographs: Åström [4], Davis [13], Jazwinski [23], and Liptser and Shiryaev [31]. It lays the groundwork for further study of optimization problems under partial information in various fields. For example, Anderson and Moore [2] which introduced the fundamental result to linear-Quadratic (LQ) optimal control problem under partial information and the separation theorem; Bensoussan [6] and Huang, Wang, and Zhang [22] also studied the optimal control for partially observed stochastic systems; Lakner [30] and Xiong and Zhou [40] considered, respectively, the utility maximization problem and mean-variance portfolio selection under partial information in mathematical finance field and so on. Further, the nonlinear filtering theory can be referred to Xiong [39].

Let's first recall the classic Kalman–Bucy filtering theory. The model is described as follows: under the probability measure  $\mathbb{P}$ ,

---

\*Received by the editors November 10, 2020; accepted for publication (in revised form) April 23, 2021; published electronically July 19, 2021.

<https://doi.org/10.1137/20M137954X>

**Funding:** This work was supported by the National Key R&D Program of China (2018YFA0703900), the National Natural Science Foundation of China (11971263, 11871458, 11701214), and the Natural Science Foundation of Shandong Province (ZR2017BA032).

<sup>†</sup>Zhongtai Securities Institute for Financial Studies, Shandong University, Jinan, Shandong 250100, People's Republic of China (jsl@sdu.edu.cn).

<sup>‡</sup>Corresponding author. Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and School of Mathematics Sciences, University of Chinese Academy of Sciences, Beijing 100149, China (chuliukong@amss.ac.cn).

<sup>§</sup>School of Mathematical Sciences, University of Jinan, Jinan, Shandong 250022, People's Republic of China (sms\_sunfc@ujn.edu.cn).

<sup>¶</sup>Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and School of Mathematics Sciences, University of Chinese Academy of Sciences, Beijing 100149, China (jif@iss.ac.cn).

$$(1.1) \quad \begin{cases} dx(t) = (B(t)x(t) + b(t))dt + dw(t), \\ x(0) = x_0, \\ dm(t) = (H(t)x(t) + h(t))dt + dv(t), \\ m(0) = 0, \end{cases}$$

where  $x(\cdot)$  is the signal process,  $m(\cdot)$  is the observation process,  $w(\cdot)$  and  $v(\cdot)$  are two independent Brownian motions. The coefficients  $B(t)$ ,  $H(t)$ ,  $b(t)$ , and  $h(t)$  are deterministic uniformly bounded functions in  $t \in [0, T]$ ,  $x_0$  is a given constant vector. Set  $\mathcal{Z}_t = \sigma\{m(s); 0 \leq s \leq t\}$  which represents all the observable information up to time  $t$ . The Kalman filter  $\bar{x}(t)$  of  $x(t)$  is

$$\bar{x}(t) = \mathbb{E}[x(t)|\mathcal{Z}_t],$$

where  $\mathbb{E}[\cdot]$  denotes the expectation with respect to the probability measure  $\mathbb{P}$ . It is well known that the optimal estimator  $\bar{x}(t)$  of the signal  $x(t)$  solves the following minimum mean square estimation problem:

$$\min_{\zeta \in L^2_{\mathcal{Z}_t}(\Omega, \mathbb{P})} \mathbb{E}[\|x(t) - \zeta\|^2].$$

So  $\bar{x}(t)$  is also called the minimum mean square estimator (MMSE).

In this paper, we suppose that there exists model uncertainty for the system (1.1). One of the uncertainty systems obtained by introducing uncertainty parameters into a linear stochastic system is called uncertain-stochastic linear dynamic systems (USS), and the relevant references are Barton and Poor [5], Borisov and Pankov [7], Nagpal and Khargonekar [32], Pankov and Borisov [33] and [34], and so on. To analyze USS one may mainly uses robust ([5]),  $H_\infty$  ([32]), and minimax ([7], [33], and [34]) approaches.

Different from the foregoing system model and the causes of uncertainty, the meaning of uncertainty in our paper is that we don't know the true probability  $\mathbb{P}$  and only know that it falls in a set of probability measures  $\mathcal{P}$  which is called the prior set. The uncertainty in USS is caused by the uncertainty of input and initial state, and these uncertainties do not affect the system probability measure; please see [33] and [34] for details.

For continuous-time models, Chen and Epstein [10] first proposed one kind of model uncertainty which is usually called drift uncertainty. Later Epstein and Ji proposed more general uncertainty models (see [16] and [17] for details), and Guo [21] introduced some basic scientific problems concerning the estimation, control, and games of dynamical systems with uncertainty and shared some related theoretical progress. In this paper, we introduce the following drift uncertainty model: for every  $P^\theta \in \mathcal{P}$ , consider

$$(1.2) \quad \begin{cases} dx(t) = (B(t)x(t) + b(t) - \theta_1(t))dt + dw^{\theta_1}(t), \\ x(0) = x_0, \\ dm(t) = (H(t)x(t) + h(t) - \theta_2(t))dt + dv^{\theta_2}(t), \\ m(0) = 0, \end{cases}$$

where  $w^{\theta_1}$  and  $v^{\theta_2}$  are Brownian motions under  $P^\theta$  and  $\theta = (\theta_1, \theta_2) \in \Theta$  is called the uncertainty parameter. When  $\theta$  changes, the distributions of the solutions  $x(\cdot)$  and  $m(\cdot)$  of the above equations also change. The question now is how to calculate the Kalman filter in such an uncertain environment. A natural idea is to calculate the worst-case minimum mean square estimation problem

$$(1.3) \quad \min_{\zeta} \sup_{P^\theta \in \mathcal{P}} E_{P^\theta}[\|x(t) - \zeta\|^2],$$

which is to minimize the maximum expected loss over a range of possible models. Under the minimax criteria, Borisov and Pankov [7] considered the minimax filtering in USS, and Borisov [8] and [9] studied this type of estimator for finite state Markov processes with uncertainty of the transition intensity and the observation matrices. Allan and Cohen [1] investigated the Kalman–Bucy filtering with an uncertainty parameter by a control approach. Moreover, in the past decade, much research has been discussed depending on the technique of  $H_\infty$  filter; see [11], [12], [32], and so on. Different from this paper, the design goal of  $H_\infty$  filter is to guarantee that the filtering error system is asymptotically stable, while achieving a prescribed  $H_\infty$  performance level. From another perspective, (1.3) can be rewritten as a minimum mean square estimation problem under a sublinear operator:

$$\min_{\zeta} \mathcal{E}(\|x(t) - \zeta\|^2),$$

where  $\mathcal{E}(\cdot) := \sup_{P^\theta \in \mathcal{P}} E_{P^\theta}[\cdot]$  is a sublinear operator. Recently, Ji, Kong, and Sun [24] and [25] studied Kalman–Bucy filtering under sublinear operators when the drift uncertainty appears in the signal process and the observation process, respectively. The related literatures about the minimum mean square estimation problems under sublinear operators include Sun and Ji [38] and Ji, Kong, and Sun [26] in which they considered these problems on  $L^\infty(\Omega, \mathbb{P})$  and  $L^p(\Omega, \mathbb{P})$ , respectively.

However, when we study some problems, especially financial and risk management problems, we need to use a more general nonlinear operator: the convex operator or convex risk measure. For example, in the last decade, the concept of convex risk measure (a special convex operator) has been extensively studied in various fields (see Föllmer and Schied [19], Arai and Fukasawa [3]). So it is an interesting problem to solve the minimum mean square estimation problem under the convex operator. Unlike sublinear operators, the lack of positive homogeneity results in an extra penalty term in the expression of convex operators. For the convex operator  $\rho(\cdot)$ , that is to say,  $\rho(\cdot)$  can be represented as

$$\rho(\cdot) = \sup_{P^\theta \in \mathcal{P}} [E_{P^\theta}[\cdot] - \alpha(P^\theta)],$$

where  $\alpha(P^\theta)$  is a penalty function defined on a probability measure set. If  $\rho(\cdot)$  is sublinear, the  $\alpha(P^\theta)$  takes values in  $\{0, \infty\}$ . The main difference between this paper and the previous ones is how to deal with the penalty term.

In this paper, we first generalize the Kalman–Bucy filtering to accommodate drift uncertainty in both signal process and observation process, and the attitude to uncertainty is characterized by a convex operator (convex risk measure). In more details, we consider system (1.2) and calculate the MMSE by solving

$$\min_{\zeta} \sup_{P^\theta} [E_{P^\theta}[\|x(t) - \zeta\|^2] + \alpha_{0,t}(P^\theta)] = \min_{\zeta} \mathcal{E}_g[\|x(t) - \zeta\|^2],$$

where

$$(1.4) \quad \mathcal{E}_g[\cdot] := \sup_{P^\theta} [E_{P^\theta}[\|\cdot\|^2] + \alpha_{0,t}(P^\theta)]$$

is called  $g$ -expectation introduced by Peng [35]. Different from [5], [7], [33], and [34], on the one hand, in essence, this paper considers the optimal estimation of signal process in the sense of the weak framework, and the optimal estimator is defined by

a saddle point. This means that there exists a set of filtering systems and filtering problems under all the probability measure  $P^\theta \in \mathcal{P}$ . For the explanation of the weak framework, please refer to Chapter 2 of Bensoussan [6]. On the other hand, the filtering problem itself (2.5) has an additional penalty term  $\alpha_{0,t}(P^\theta)$ . It is because the presence of the penalty term brings great difficulty in getting the solution to the filtering problem. In our context,  $\mathcal{E}_g[\cdot]$  is a special convex operator, and (1.4) is its dual representation obtained in El Karoui, Peng, and Quenez [15]. Under some mild conditions, we prove that there exists a worst-case prior  $P^{\theta^*}$ . Based on this  $P^{\theta^*}$  we obtained the filtering equation by which the MMSE  $\hat{x}$  is governed.

The convex  $g$ -expectation is just a special convex operator. It is worth studying the minimum mean square estimation problem under the general convex operators. In the second part of this paper, we solve the following problem (for the convenience of readers, we misused some notations in the introduction and section 4):

$$\min_{\zeta} \rho(\|x(t) - \zeta\|^2),$$

where  $\rho(\cdot)$  is a general convex operator (convex risk measure). The existence and uniqueness results of the MMSE under the general convex operators are deduced.

The paper is organized as follows. In section 2, we give some preliminaries and formulate our filtering problem under  $g$ -expectations. In section 3, the worst-case prior  $P^{\theta^*}$  is obtained, and the corresponding Kalman–Bucy filtering equation (3.8) is deduced. We study the minimum mean square estimation problem under general convex operators on  $L^p_{\mathcal{F}}(\mathbb{P})$  and obtain the existence and uniqueness results of the MMSE in section 4.

**2. Preliminaries and problem formulation.** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a complete probability space on which two standard, independent  $n$ -dimensional and  $m$ -dimensional Brownian motions  $w(\cdot)$  and  $v(\cdot)$  are defined. For a fixed time  $T > 0$ , denote by  $\mathbb{F} = \{\mathcal{F}_t, 0 \leq t \leq T\}$  the natural filtration of  $w(\cdot)$  and  $v(\cdot)$  satisfying the usual conditions. We assume  $\mathcal{F} = \mathcal{F}_T$ . For any given Euclidean space  $\mathbb{H}$ , denote by  $\langle \cdot, \cdot \rangle$  (resp.,  $\|\cdot\|$ ) the scalar product (resp., norm) of  $\mathbb{H}$ . Let  $A^\top$  denote the transpose of a matrix  $A$ . For an  $\mathbb{R}^n$ -valued vector  $x = (x_1, \dots, x_n)^\top$ ,  $|x| := (|x_1|, \dots, |x_n|)^\top$ ; for two  $\mathbb{R}^n$ -valued vectors  $x$  and  $y$ ,  $x \leq y$  means that  $x_i \leq y_i$  for  $i = 1, \dots, n$ . Throughout this paper,  $0$  denotes the matrix/vector with appropriate dimension whose all entries are zero. For  $1 < p < \infty$ , denote by  $L^p_{\mathbb{F}}(0, T; \mathbb{H})$  the space of all the  $\mathbb{H}$ -valued,  $\mathbb{F}$ -adapted, and  $p$ th power integrable stochastic processes  $f(\cdot)$  on  $[0, T]$  such that

$$\mathbb{E} \left[ \int_0^T \|f(r)\|^p dr \right] < \infty.$$

The Kalman–Bucy filtering theory is based on a reference probability measure  $\mathbb{P}$  for the system (1.1). However, if we don't know the true probability measure  $\mathbb{P}$  and only know that it falls in the set  $\mathcal{P}$  which is a suitably chosen space of equivalent probability measures, then it is naturally to study the worst-case MMSEs.

**2.1. Prior set and  $g$ -expectation.** In order to characterize uncertainty, we introduce the prior set  $\mathcal{P}$  and  $g$ -expectation which is a special convex operator.

Let  $\theta(\cdot) = (\theta_1(\cdot), \theta_2(\cdot))^\top$  be an  $\mathbb{R}^{n+m}$ -valued progressively measurable process on  $[0, T]$ . For a given constant  $\mu$ , let  $\Theta$  be the set of all  $\mathbb{R}^{n+m}$ -valued progressively measurable processes  $\theta$  with  $\|\theta_i(t)\| \leq \mu$ ,  $0 \leq t \leq T$ . Moreover, the set  $\Theta$  is also requested to satisfy stochastically convex and compact under  $L^2_{\mathbb{F}}([0, T], \mathbb{P})$ -norm.

*Remark 2.1.* Say that the set  $\Theta$  is stochastically convex if for any real-process  $\{\lambda(t)\}_{t \geq 0}$  with  $0 \leq \lambda(t) \leq 1 \forall 0 \leq t \leq T$ ,  $\theta$  and  $\theta'$  in  $\Theta$  implies that  $(\lambda(t)\theta(t) + (1 - \lambda(t))\theta'(t)) \in \Theta$ .

Define

$$(2.1) \quad \mathcal{P} = \left\{ P^\theta \mid \frac{dP^\theta}{d\mathbb{P}} = f^\theta(T) \text{ for } \theta \in \Theta \right\},$$

where

$$f^\theta(T) := \exp \left( - \int_0^T \theta_1^\top(t) dw(t) - \frac{1}{2} \int_0^T \|\theta_1(t)\|^2 dt - \int_0^T \theta_2^\top(t) dv(t) - \frac{1}{2} \int_0^T \|\theta_2(t)\|^2 dt \right).$$

Due to the boundedness of  $\theta$ , the Novikov condition holds (see Karatzas and Shreve [29]). Therefore,  $P^\theta$  defined by (2.1) is a probability measure which is equivalent to probability measure  $\mathbb{P}$ , and the processes  $w^{\theta_1}(t) = w(t) + \int_0^t \theta_1(s) ds$  and  $v^{\theta_2}(t) = v(t) + \int_0^t \theta_2(s) ds$  are Brownian motions under this probability measure  $P^\theta$  by Girsanov's theorem. The set  $\Theta$  characterizes the ambiguity, and  $\mathcal{P}$  is usually called the prior set.

Then, we introduce  $g$ -expectation and its dual representation (see [35] and [15]). In the following we will see that  $g$ -expectation is a powerful tool for studying uncertainty.

**DEFINITION 2.2.** We call a function  $g : \Omega \times [0, T] \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  a standard generator if it satisfies the following conditions:

- $(g(\omega, t, z_1, z_2))_{t \in [0, T]}$  is an adapted process with

$$\mathbb{E} \int_0^T |g(\omega, t, z_1, z_2)|^2 dt < \infty$$

for all  $z_1 \in \mathbb{R}^n$  and  $z_2 \in \mathbb{R}^m$ ;

- $g(\omega, t, z_1, z_2)$  is Lipschitz continuous in  $z_1$  and  $z_2$ , uniformly in  $t$  and  $\omega$ : there exists constant  $\mu > 0$  such that for all  $z_1, \tilde{z}_1 \in \mathbb{R}^n$  and  $z_2, \tilde{z}_2 \in \mathbb{R}^m$  we have

$$|g(\omega, t, z_1, z_2) - g(\omega, t, \tilde{z}_1, \tilde{z}_2)| \leq \mu(\|z_1 - \tilde{z}_1\| + \|z_2 - \tilde{z}_2\|);$$

- $g(\omega, t, 0, 0) = 0$  for all  $t \geq 0$  and  $\omega \in \Omega$ .

For a standard generator  $g$ , the following backward stochastic differential equation (BSDE)

$$\begin{cases} -dY(t) = g(t, Z_1(t), Z_2(t))dt - Z_1^\top(t)dw(t) - Z_2^\top(t)dv(t), & t \in [0, T], \\ Y(T) = \xi, \end{cases}$$

with terminal condition  $\xi \in L^2_{\mathcal{F}_T}(\Omega, \mathbb{P})$  has a unique square integrable solution  $(Y(t), Z_1(t), Z_2(t))_{t \in (0, T]}$  (see [35]). Peng [35] calls  $Y(t) := \mathcal{E}_g(\xi | \mathcal{F}_t)$  the (condition)  $g$ -expectation of  $\xi$  at time  $t$ .

**DEFINITION 2.3.** A standard generator  $g$  is called a convex generator if  $g(\omega, t, z_1, z_2)$  is convex in  $z_1$  and  $z_2$  for  $z_1 \in \mathbb{R}^n$  and  $z_2 \in \mathbb{R}^m$ . The  $g$ -expectation with a convex generator is called the convex  $g$ -expectation.

Now we give the dual representation of the convex  $g$ -expectation through the prior set and the concave dual function of  $g$ .

Let

$$G(\omega, t, \theta_1, \theta_2) = \inf_{z_1 \in \mathbb{R}^n, z_2 \in \mathbb{R}^m} [g(\omega, t, z_1, z_2) + \langle z_1, \theta_1 \rangle + \langle z_2, \theta_2 \rangle]$$

$$(\omega \in \Omega, t \in [0, T], \theta_1 \in \mathbb{R}^n, \theta_2 \in \mathbb{R}^m)$$

be the concave dual function of  $g(\omega, t, z_1, z_2)$ .

El Karoui, Peng, and Quenez [15] (also see Delbaen, Peng, and Gianin [14]) established the following dual representation for  $g$ -expectation: for an  $\mathcal{F}_s$ -measurable random variable  $\xi$ , the  $g$ -expectation at time  $t$  can be represented as

$$(2.2) \quad \mathcal{E}_g(\xi|\mathcal{F}_t) = \sup_{P^\theta \in \mathcal{P}} [E_{P^\theta}[\xi|\mathcal{F}_t] + \alpha_{t,s}(P^\theta)],$$

where

$$(2.3) \quad \alpha_{t,s}(P^\theta) := E_{P^\theta} \left[ \int_t^s G(r, \theta_1(r), \theta_2(r)) dr | \mathcal{F}_t \right], \quad 0 \leq t \leq s \leq T.$$

*Remark 2.4.* It is easy to check that  $\mathcal{E}_g(\cdot|\mathcal{F}_t)$  is a special convex operator (see (4.1)). Moreover, if we let the standard generator  $g(t, z_1, z_2) = \mu(|z_1| + |z_2|)$ , then the corresponding dual function of  $g(t, z_1, z_2)$  and penalty term  $\alpha_{t,s}(P^\theta)$  are simultaneously equal to 0. Then the above convex operator  $\mathcal{E}_g(\cdot|\mathcal{F}_t)$  degenerates to a sublinear operator.

**2.2. Problem formulation .** We formulate the Kalman–Bucy filtering problem under uncertainty. For every  $\theta \in \Theta$ , under the probability measure  $P^\theta \in \mathcal{P}$ ,

$$(2.4) \quad \begin{cases} dx(t) = (B(t)x(t) + b(t) - \theta_1(t))dt + dw^{\theta_1}(t), \\ x(0) = x_0, \\ dm(t) = (H(t)x(t) + h(t) - \theta_2(t))dt + dv^{\theta_2}(t), \\ m(0) = 0, \end{cases}$$

where  $x(\cdot) \in L^2_{\mathbb{R}}(0, T; \mathbb{R}^n)$  is the signal process and  $m(\cdot) \in L^2_{\mathbb{R}}(0, T; \mathbb{R}^m)$  is the observation process. The coefficients  $B(t) \in \mathbb{R}^{n \times n}$ ,  $H(t) \in \mathbb{R}^{m \times n}$ ,  $b(t) \in \mathbb{R}^n$ , and  $h(t) \in \mathbb{R}^m$  are deterministic uniformly bounded functions in  $t \in [0, T]$ , and  $x_0 \in \mathbb{R}^n$  is a given constant vector. Set

$$\mathcal{Z}_t = \sigma\{m(s); 0 \leq s \leq t\},$$

which represents all the observable information up to time  $t$ . We want to calculate the MMSE of the signal  $x(t)$  by solving the following worst-case minimum mean square estimation problem:

$$(2.5) \quad \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R}^n)} \mathcal{E}_g(\|x(t) - \zeta(t)\|^2)$$

$$= \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R}^n)} \sup_{P^\theta \in \mathcal{P}} [E_{P^\theta}(\|x(t) - \zeta(t)\|^2), +\alpha_{0,t}(P^\theta)],$$

where  $L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R}^n)$  is the set of all the  $\mathbb{R}^n$ -valued  $(2 + \epsilon)$  integrable  $\mathcal{Z}_t$ -measurable random variables and  $0 < \epsilon < 1$ .

DEFINITION 2.5. If  $\hat{x}(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R}^n)$  satisfies

$$\mathcal{E}_g(\|x(t) - \hat{x}(t)\|^2) = \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R}^n)} \mathcal{E}_g(\|x(t) - \zeta(t)\|^2),$$

then we call  $\hat{x}(t)$  the MMSE of  $x(t)$ .

**3. Kalman-Bucy filtering under  $g$ -expectation.** In this section, we calculate the MMSE  $\hat{x}(t)$  of (2.5) for  $t \in [0, T]$ . Without loss of generality, all the statements in this section are only proved in the one-dimensional case, i.e.,  $n = m = 1$ .

LEMMA 3.1. *The set  $\{\frac{dP^\theta}{d\mathbb{P}} : P^\theta \in \mathcal{P}\} \subset L^{1+\frac{2}{\epsilon}}(\Omega, \mathcal{F}, \mathbb{P})$  is  $\sigma(L^{1+\frac{2}{\epsilon}}(\Omega, \mathcal{F}, \mathbb{P}), L^{1+\frac{\epsilon}{2}}(\Omega, \mathcal{F}, \mathbb{P}))$ -compact, and  $\mathcal{P}$  is convex.*

*Proof.* Since  $\theta$  is bounded, by Theorem 5.3 in section 5, the set  $\{\frac{dP^\theta}{d\mathbb{P}} : P^\theta \in \mathcal{P}\}$  is bounded in norm  $\|\cdot\|_{1+\frac{2}{\epsilon}}$ . Next, we need to prove that this set is  $\sigma(L^{1+\frac{2}{\epsilon}}(\Omega, \mathcal{F}, \mathbb{P}), L^{1+\frac{\epsilon}{2}}(\Omega, \mathcal{F}, \mathbb{P}))$ -closed.

From Theorem 4.1 of Chapter 1 in Simons [37], we take a sequence  $\{f^{\theta_n}(T)\}_{n \geq 1} \subset \{\frac{dP^\theta}{d\mathbb{P}} : P^\theta \in \mathcal{P}\}$  such that  $\{f^{\theta_n}(T)\}_{n \geq 1}$  is weakly convergence to  $\tilde{f}(T)$ , where  $\{\theta_n\}_{n \geq 1}$  is generator of  $\{f^{\theta_n}(T)\}_{n \geq 1}$ . Based on the compactness of set  $\Theta$ , we can take a subsequence  $\{\theta_{n_k}\}_{k \geq 1}$  of  $\{\theta_n\}_{n \geq 1}$  such that  $\theta_{n_k} \rightarrow_{L^2_{\mathbb{F}}([0, T], \mathbb{P})} \bar{\theta}$ , where  $\theta_{n_k} = (\theta_{n_k, 1}, \theta_{n_k, 2}) \in \Theta, \bar{\theta} = (\bar{\theta}_1, \bar{\theta}_2) \in \Theta$ .

Let

$$\begin{cases} \tilde{f}(T) = \exp\left(-\int_0^T \bar{\theta}_1(t)dw(t) - \frac{1}{2}\int_0^T |\bar{\theta}_1(t)|^2 dt - \int_0^T \bar{\theta}_2(t)dv(t) \right. \\ \quad \left. - \frac{1}{2}\int_0^T |\bar{\theta}_2(t)|^2 dt\right), \\ H(\theta_{n_k}) = -\int_0^T \theta_{n_k, 1}(t)dw(t) - \frac{1}{2}\int_0^T |\theta_{n_k, 1}(t)|^2 dt - \int_0^T \theta_{n_k, 2}(t)dv(t) \\ \quad - \frac{1}{2}\int_0^T |\theta_{n_k, 2}(t)|^2 dt, \\ H(\bar{\theta}) = -\int_0^T \bar{\theta}_1(t)dw(t) - \frac{1}{2}\int_0^T |\bar{\theta}_1(t)|^2 dt - \int_0^T \bar{\theta}_2(t)dv(t) - \frac{1}{2}\int_0^T |\bar{\theta}_2(t)|^2 dt. \end{cases}$$

After a short calculation, we have  $\lim_{k \rightarrow \infty} \mathbb{E}[(H(\theta_{n_k}) - H(\bar{\theta}))^2] = 0$ . Based on Theorem 5.6 in section 5, it results in that  $f^{\theta_{n_k}}(T) \xrightarrow{\mathbb{P}} \tilde{f}(T)$ . By Theorem 5.3 in section 5, for any given constant  $r > 1$ , we have  $\mathbb{E}[f^{\theta_{n_k}}(T)]^K \leq M$ , where  $K = (1 + \frac{2}{\epsilon})r$  and  $M = \exp((K^2 - K)\mu^2 T)$ . Then,  $\{|f^{\theta_{n_k}}(T)|^{1+\frac{2}{\epsilon}}\}_{k \geq 1}$  is uniformly integrable. Therefore,  $f^{\theta_{n_k}}(T) \rightarrow_{L^{1+\frac{2}{\epsilon}}(\mathbb{F})} \tilde{f}(T)$ . Further, based on the uniqueness of the limit, it reduces that  $\tilde{f}(T) = \tilde{f}(T) \in \{\frac{dP^\theta}{d\mathbb{P}} : P^\theta \in \mathcal{P}\}$ . Then, the set  $\{\frac{dP^\theta}{d\mathbb{P}} : P^\theta \in \mathcal{P}\}$  is  $\sigma(L^{1+\frac{2}{\epsilon}}(\Omega, \mathcal{F}, \mathbb{P}), L^{1+\frac{\epsilon}{2}}(\Omega, \mathcal{F}, \mathbb{P}))$ -compact.

Let  $\theta^1 = (\theta_1^1, \theta_2^1)^\top$  and  $\theta^2 = (\theta_1^2, \theta_2^2)^\top$  belong to  $\Theta$ .  $f^{\theta^1}$  and  $f^{\theta^2}$  denote the corresponding exponential martingales: for  $t \in [0, T]$ ,

$$f^{\theta^i}(t) = \exp\left(\int_0^t \theta_1^i(s)dw(s) - \frac{1}{2}\int_0^t (\theta_1^i(s))^2 ds + \int_0^t \theta_2^i(s)dv(s) - \frac{1}{2}\int_0^t (\theta_2^i(s))^2 ds\right)$$

which satisfies

$$df^{\theta^i}(t) = f^{\theta^i}(t)(\theta_1^i(t)dw(t) + \theta_2^i(t)dv(t)), \quad i = 1, 2.$$

Let  $\lambda_1$  and  $\lambda_2$  be nonnegative constants which belong to  $(0, 1)$  with  $\lambda_1 + \lambda_2 = 1$ . Define

$$\begin{cases} \theta_1^\lambda(t) = \frac{\lambda_1 \theta_1^1(t) f^{\theta^1}(t) + \lambda_2 \theta_1^2(t) f^{\theta^2}(t)}{\lambda_1 f^{\theta^1}(t) + \lambda_2 f^{\theta^2}(t)}, \\ \theta_2^\lambda(t) = \frac{\lambda_1 \theta_2^1(t) f^{\theta^1}(t) + \lambda_2 \theta_2^2(t) f^{\theta^2}(t)}{\lambda_1 f^{\theta^1}(t) + \lambda_2 f^{\theta^2}(t)}. \end{cases}$$

It is easy to verify that

$$d(\lambda_1 f^{\theta^1}(t) + \lambda_2 f^{\theta^2}(t)) = (\lambda_1 f^{\theta^1}(t) + \lambda_2 f^{\theta^2}(t))(\theta_1^\lambda(t)dw(t) + \theta_2^\lambda(t)dv(t)).$$

Since  $f^{\theta^i}(t) > 0, i = 1, 2$ , the process  $\theta^\lambda = (\theta_1^\lambda, \theta_2^\lambda)^\top$  belongs to  $\Theta$ . Therefore, it results in that the set  $\mathcal{P}$  is convex. This completes the proof.  $\square$

LEMMA 3.2. *The penalty term  $\alpha_{0,T}(P^\theta)$  is a concave functional on  $\mathcal{P}$ .*

*Proof.* Let  $\theta^1 = (\theta_1^1, \theta_2^1)^\top$  and  $\theta^2 = (\theta_1^2, \theta_2^2)^\top$  belong to  $\Theta$ .  $f^{\theta^1}$  and  $f^{\theta^2}$  denote the exponential martingales, respectively, as in Lemma 3.1. By Lemma 3.1, the exponential martingale  $(\lambda_1 \frac{dP^{\theta^1}}{d\mathbb{P}} + \lambda_2 \frac{dP^{\theta^2}}{d\mathbb{P}})$  is generated by  $\theta^\lambda = (\theta_1^\lambda, \theta_2^\lambda)$ . It yields that

$$\alpha_{0,T}(\lambda_1 P^{\theta^1} + \lambda_2 P^{\theta^2}) = \mathbb{E} \left[ (\lambda_1 f^{\theta^1}(T) + \lambda_2 f^{\theta^2}(T)) \int_0^T G(t, \theta_1^\lambda(t), \theta_2^\lambda(t)) dt \right].$$

Since  $G(t, \cdot, \cdot)$  is a concave function, we have

$$\begin{aligned} & \alpha_{0,T}(\lambda_1 P^{\theta^1} + \lambda_2 P^{\theta^2}) \\ & \geq \mathbb{E} \left[ (\lambda_1 f^{\theta^1}(T) + \lambda_2 f^{\theta^2}(T)) \left( \int_0^T \frac{\lambda_1 f^{\theta^1}(t)}{\lambda_1 f^{\theta^1}(t) + \lambda_2 f^{\theta^2}(t)} G(t, \theta_1^1(t), \theta_2^1(t)) \right. \right. \\ & \quad \left. \left. + \int_0^T \frac{\lambda_2 f^{\theta^2}(t)}{\lambda_1 f^{\theta^1}(t) + \lambda_2 f^{\theta^2}(t)} G(t, \theta_1^2(t), \theta_2^2(t)) dt \right) \right] \\ & = \mathbb{E} \left[ \left( \int_0^T \lambda_1 f^{\theta^1}(t) G(t, \theta_1^1(t), \theta_2^1(t)) + \int_0^T \lambda_2 f^{\theta^2}(t) G(t, \theta_1^2(t), \theta_2^2(t)) \right) dt \right] \\ & = \mathbb{E} \left[ \lambda_1 f^{\theta^1}(T) \int_0^T G(t, \theta_1^1(t), \theta_2^1(t)) dt \right] + \mathbb{E} \left[ \lambda_2 f^{\theta^2}(T) \int_0^T G(t, \theta_1^2(t), \theta_2^2(t)) dt \right] \\ & = \lambda_1 \alpha_{0,T}(P^{\theta^1}) + \lambda_2 \alpha_{0,T}(P^{\theta^2}). \end{aligned}$$

Therefore, the penalty term  $\alpha_{0,T}(P^\theta)$  is a concave functional on  $\mathcal{P}$ . This completes the proof.  $\square$

Remark 3.3. It is easy to check that for any  $t \in [0, T]$ ,  $\alpha_{0,t}(P^\theta)$  is a concave functional on  $\mathcal{P}$  and

$$\alpha_{0,t}(P^\theta) = \mathbb{E} \left[ f^\theta(T) \cdot \int_0^t G(s, \theta_1(s), \theta_2(s)) ds \right] = \mathbb{E} \left[ f^\theta(t) \cdot \int_0^t G(s, \theta_1(s), \theta_2(s)) ds \right].$$

By abuse of notation, we sometimes write  $\alpha_{0,t}(f^\theta)$  instead of  $\alpha_{0,t}(P^\theta)$ .

Denote the generators of  $(g_m(t))_{t \in [0, T]}$ ,  $m = 1, 2, \dots$  and  $(f^*(t))_{t \in [0, T]}$  by  $\theta^m(t) = (\theta_1^m(t), \theta_2^m(t)) \in \Theta$  and  $\theta^*(t) = (\theta_1^*(t), \theta_2^*(t)) \in \Theta$ , respectively; i.e., for  $0 \leq t \leq T$ ,

$$\begin{aligned} g_m(t) &= \exp \left( \int_0^t \theta_1^m(s) dw(s) - \frac{1}{2} \int_0^t (\theta_1^m(s))^2 ds + \int_0^t \theta_2^m(s) dv(s) - \frac{1}{2} \int_0^t (\theta_2^m(s))^2 ds \right), \\ f^*(t) &= \exp \left( \int_0^t \theta_1^*(s) dw(s) - \frac{1}{2} \int_0^t (\theta_1^*(s))^2 ds + \int_0^t \theta_2^*(s) dv(s) - \frac{1}{2} \int_0^t (\theta_2^*(s))^2 ds \right). \end{aligned}$$

LEMMA 3.4. *Suppose that the stochastic processes  $(g_m(t))_{t \in [0, T]}$ ,  $m = 1, 2, \dots$  and  $(f^*(t))_{t \in [0, T]}$  are exponential martingales with respect to the filtration  $\mathbb{F}$  and  $(g_m(T) - f^*(T)) \rightarrow_{L^2_{\mathcal{F}}(\mathbb{P})} 0$ . Then for any  $0 \leq t \leq T$ , we have*

$$(\theta_i^m(t) - \theta_i^*(t)) \xrightarrow{L^2_{\mathcal{F}}(\mathbb{P})} 0, \quad i = 1, 2.$$



*Proof.* We want to prove that  $(\theta^m)$  converges to  $\theta^*$ . Since  $g_m(\cdot)$  and  $f^*(\cdot)$  are martingales and  $g_m(T) \rightarrow_{L^2_{\mathcal{F}}(\mathbb{P})} f^*(T)$ , it is easy to verify that  $g_m(t) \rightarrow_{L^2_{\mathcal{F}}(\mathbb{P})} f^*(t)$  for any  $t \in [0, T]$ . Applying Itô's formula to  $(g_m(t) - f^*(t))^2$ , we have

$$\begin{aligned} & d(g_m(t) - f^*(t))^2 \\ &= 2(g_m(t) - f^*(t))[(g_m(t)\theta_1^m(t) - f^*(t)\theta_1^*(t))dw(t) + (g_m(t)\theta_2^m(t) \\ &\quad - f^*(t)\theta_2^*(t))dv(t)] + (g_m(t)\theta_1^m(t) - f^*(t)\theta_1^*(t))^2 dt + (g_m(t)\theta_2^m(t) \\ &\quad - f^*(t)\theta_2^*(t))^2 dt. \end{aligned}$$

Taking expectation on both sides,

$$(3.1) \quad \mathbb{E}[(g_m(T) - f^*(T))^2] = \mathbb{E} \left[ \int_0^T (g_m(t)\theta_1^m(t) - f^*(t)\theta_1^*(t))^2 dt \right] \\ + \mathbb{E} \left[ \int_0^T (g_m(t)\theta_2^m(t) - f^*(t)\theta_2^*(t))^2 dt \right].$$

Since  $\lim_{m \rightarrow \infty} \mathbb{E}[(g_m(T) - f^*(T))^2] = 0$ , it yields that

$$(3.2) \quad \lim_{m \rightarrow \infty} \mathbb{E} \left[ \int_0^T (g_m(t)\theta_i^m(t) - f^*(t)\theta_i^*(t))^2 dt \right] = 0, \quad i = 1, 2.$$

Note that

$$\begin{aligned} & \mathbb{E} \left[ \int_0^T (g_m(t)\theta_1^m(t) - f^*(t)\theta_1^*(t))^2 dt \right] \\ &= \mathbb{E} \int_0^T [(f^*(t) - g_m(t))^2(\theta_1^*(t))^2 + (g_m(t))^2(\theta_1^*(t) - \theta_1^m(t))^2 \\ &\quad + 2(f^*(t) - g_m(t))g_m(t)\theta_1^*(t)(\theta_1^*(t) - \theta_1^m(t))] dt. \end{aligned}$$

Because  $g_m(t) \rightarrow_{L^2_{\mathcal{F}}(\mathbb{P})} f^*(t)$  and  $\theta$  is bounded, we have

$$\begin{aligned} & \lim_{m \rightarrow \infty} \mathbb{E}[(f^*(t) - g_m(t))^2(\theta_1^*(t))^2] = 0, \\ & \lim_{m \rightarrow \infty} \mathbb{E}[(f^*(t) - g_m(t))g_m(t)\theta_1^*(t)(\theta_1^*(t) - \theta_1^m(t))] = 0. \end{aligned}$$

Therefore,  $\lim_{m \rightarrow \infty} \mathbb{E}[(g_m(t))^2(\theta_1^*(t) - \theta_1^m(t))^2] = 0$ . It results in that  $(g_m(t))^2(\theta_1^*(t) - \theta_1^m(t))^2 \rightarrow_{\mathbb{P}} 0$ . Since  $g_m(t) \rightarrow_{\mathbb{P}} f^*(t)$ , we have  $(\theta_1^*(t) - \theta_1^m(t))^2 \rightarrow_{\mathbb{P}} 0$ . Due to the boundedness of  $\theta$ , we obtain  $(\theta_1^*(t) - \theta_1^m(t)) \rightarrow_{L^2_{\mathcal{F}}(\mathbb{P})} 0$ . Similarly, we can obtain  $(\theta_2^*(t) - \theta_2^m(t)) \rightarrow_{L^2_{\mathcal{F}}(\mathbb{P})} 0$ . This completes the proof.  $\square$

LEMMA 3.5.  $\alpha_{0,t}(f^\theta)$  is an upper semicontinuous function on  $\{\frac{dP^\theta}{d\mathbb{P}} : P^\theta \in \mathcal{P}\}$ .

*Proof.* Take a sequence  $\{f^{\theta_n}(T)\}_{n \geq 1} \in \{\frac{dP^\theta}{d\mathbb{P}} : P^\theta \in \mathcal{P}\}$  such that  $f^{\theta_n}(T) \rightarrow_{L^2_{\mathcal{F}}(\mathbb{P})} f^*(T)$ . Then, we can take a subsequence  $\{f^{\theta_{n_k}}(T)\}_{k \geq 1} \subset \{f^{\theta_n}(T)\}_{n \geq 1}$  such that  $f^{\theta_{n_k}}(T) \rightarrow_{\mathbb{P}-a.s.} f^*(T)$ , where  $\theta_{n_k} \in \Theta$  is the generator of  $f^{\theta_{n_k}}(T)$  for any  $k \geq 1$  and  $\theta_{n_k} \rightarrow_{\mathbb{P}} \theta^*$  by Lemma 3.4. Next, take a subsequence  $\{\theta^{n_{k_l}}\}_{l \geq 1} \subset \{\theta_{n_k}\}_{k \geq 1}$  such that  $\theta^{n_{k_l}} \rightarrow_{\mathbb{P}-a.s.} \theta^*$  and denote  $\{f^{\theta^{n_{k_l}}}(T)\}_{l \geq 1}$  by the exponential martingale sequence

with respect to generator  $\{\theta^{n_{k_l}}\}_{l \geq 1}$ , then  $f^{\theta^{n_{k_l}}}(T) \xrightarrow{\mathbb{P}\text{-a.s.}} f^*(T)$ . Therefore, the following relations hold:

$$\begin{aligned} \limsup_{n \rightarrow \infty} \alpha_{0,t}(f^{\theta^n}) &= \limsup_{l \rightarrow \infty} \alpha_{0,t}(f^{\theta^{n_{k_l}}}) \\ &= \limsup_{l \rightarrow \infty} \mathbb{E} \left[ f^{\theta^{n_{k_l}}}(T) \cdot \int_0^t G(s, \theta_1^{n_{k_l}}(s), \theta_2^{n_{k_l}}(s)) ds \right] \\ &= \limsup_{l \rightarrow \infty} \mathbb{E} \left[ f^{\theta^{n_{k_l}}}(T) \cdot \int_0^t \inf_{z_1 \in \mathbb{R}, z_2 \in \mathbb{R}} [g(s, z_1, z_2) + \langle z_1, \theta_1^{n_{k_l}}(s) \rangle, \langle z_2, \theta_2^{n_{k_l}}(s) \rangle] ds \right] \\ &\leq \mathbb{E} \left[ \int_0^t \limsup_{l \rightarrow \infty} \inf_{z_1 \in \mathbb{R}, z_2 \in \mathbb{R}} f^{\theta^{n_{k_l}}}(T) \cdot [g(s, z_1, z_2) + \langle z_1, \theta_1^{n_{k_l}}(s) \rangle, \langle z_2, \theta_2^{n_{k_l}}(s) \rangle] ds \right] \\ &\leq \mathbb{E} \left[ \int_0^t \inf_{z_1 \in \mathbb{R}, z_2 \in \mathbb{R}} \limsup_{l \rightarrow \infty} f^{\theta^{n_{k_l}}}(T) \cdot [g(s, z_1, z_2) + \langle z_1, \theta_1^{n_{k_l}}(s) \rangle, \langle z_2, \theta_2^{n_{k_l}}(s) \rangle] ds \right] \\ &= \mathbb{E} \left[ \int_0^t \inf_{z_1 \in \mathbb{R}, z_2 \in \mathbb{R}} f^*(T) \cdot [g(s, z_1, z_2) + \langle z_1, \theta_1^*(s) \rangle, \langle z_2, \theta_2^*(s) \rangle] ds \right] \\ &= \alpha_{0,t}(f^*). \end{aligned}$$

This completes the proof. □

In the following, we prove that the worst-case prior  $P^{\theta^*}$  exists.

**THEOREM 3.6.** *For a given  $t \in [0, T]$ , there exists a  $\theta^* \in \Theta$  such that*

$$\begin{aligned} (3.3) \quad \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} \mathcal{E}_g[(x(t) - \zeta(t))^2] &= \sup_{P^\theta \in \mathcal{P}} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^\theta}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^\theta)] \\ &= \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^{\theta^*}}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^{\theta^*})]. \end{aligned}$$

*Proof.* Firstly, we prove the first equality. According to Lemmas 3.1, 3.2, and 3.5, the function  $\mathbb{E}[f^\theta(t)(x(t) - \zeta(t))^2] + \alpha_{0,t}(f^\theta)$  is upper semicontinuous with respect to  $f^\theta(t)$ ; then the original robust estimation problem (2.5) satisfies all the conditions in Theorem 5.1. Therefore, the first equality is verified.

Secondly, we prove the second equality. Choose a sequence  $\{\theta^n\}$ ,  $n = 1, 2, \dots$  such that

$$\begin{aligned} (3.4) \quad \lim_{n \rightarrow \infty} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^{\theta^n}}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^{\theta^n})] \\ = \sup_{P^\theta \in \mathcal{P}} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [\alpha_{0,t}(P^\theta) + E_{P^\theta}[(x(t) - \zeta(t))^2]]. \end{aligned}$$

Set  $f^{\theta^n}(T) = \frac{dP^{\theta^n}}{d\mathbb{P}}$ . By the Komlós theorem A.3.4 in [36], there exist a subsequence  $\{f^{\theta^{n_k}}(T)\}_{k \geq 1}$  of  $\{f^{\theta^n}(T)\}_{n \geq 1}$  and an  $f^*(T) \in L^1(\Omega, \mathcal{F}, \mathbb{P})$  such that

$$(3.5) \quad \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m f^{\theta^{n_k}}(T) = f^*(T), \quad \mathbb{P}\text{-a.s.}$$

Let  $g_m(T) = \frac{1}{m} \sum_{k=1}^m f^{\theta^{n_k}}(T)$ . We have  $g_m(T) \xrightarrow{\mathbb{P}\text{-a.s.}} f^*(T)$ . By Theorem 5.3 in section 5, for any given constant  $p > 1$  and  $m \geq 1$ , we have  $\mathbb{E}(g_m(T))^K \leq M$ , where  $K = (1 + \frac{2}{\epsilon})p$  and  $M = \exp((K^2 - K)\mu^2 T)$ . Then,  $\{|g_m(T)|^{1+\frac{2}{\epsilon}} : m = 1, 2, \dots\}$  is uniformly integrable. Therefore, it results in that  $g_m(T) \xrightarrow{L^{1+\frac{2}{\epsilon}}(\mathbb{P})} f^*(T)$  and

$f^*(T) \in L^{1+\frac{2}{\epsilon}}(\Omega, \mathcal{F}, \mathbb{P})$ . According to the convexity and weak compactness of the set  $\{\frac{dP^\theta}{d\mathbb{P}} : P^\theta \in \mathcal{P}\}$ , there exists a  $\theta^*$  such that  $\frac{dP^{\theta^*}}{d\mathbb{P}} = f^*(T)$ .

Then we prove that the probability measure  $P^{\theta^*}$  with respect to obtained generator  $\theta^*$  satisfies (3.3). Based on (3.4) and (3.5), we have

$$\begin{aligned}
 (3.6) \quad & \sup_{P^\theta \in \mathcal{P}} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^\theta}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^\theta)] \\
 &= \lim_{n \rightarrow \infty} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [\mathbb{E}[f^{P^{\theta_n}}(T)(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^{\theta_n})] \\
 &= \lim_{k \rightarrow \infty} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [\mathbb{E}[f^{P^{\theta_{n_k}}}(T)(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^{\theta_{n_k}})] \\
 &= \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{k=1}^m \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [\mathbb{E}[f^{P^{\theta_{n_k}}}(T)(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^{\theta_{n_k}})] \\
 &\leq \liminf_{m \rightarrow \infty} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} \frac{1}{m} \sum_{k=1}^m [\mathbb{E}[f^{P^{\theta_{n_k}}}(T)(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^{\theta_{n_k}})] \\
 &\leq \liminf_{m \rightarrow \infty} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [\mathbb{E}[g_m(T)(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^{\theta^m})],
 \end{aligned}$$

where the last inequality is due to the concavity of  $\alpha_{0,t}(\cdot)$ .

According to (3.6) and Lemmas 3.4 and 3.5, it results in that

$$\begin{aligned}
 (3.7) \quad & \sup_{P^\theta \in \mathcal{P}} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^\theta}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^\theta)] \\
 &\geq \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^{\theta^*}}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^{\theta^*})] \\
 &\geq \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} \left[ \mathbb{E} \left[ \lim_{m \rightarrow \infty} g_m(T)(x(t) - \zeta(t))^2 \right] + \limsup_{m \rightarrow \infty} \alpha_{0,t}(f^{\theta^m}) \right] \\
 &= \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} \left[ \mathbb{E} \left[ \limsup_{m \rightarrow \infty} g_m(T)(x(t) - \zeta(t))^2 \right] + \limsup_{m \rightarrow \infty} \alpha_{0,t}(f^{\theta^m}) \right] \\
 &\geq \limsup_{m \rightarrow \infty} \left\{ \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [\mathbb{E}[g_m(T)(x(t) - \zeta(t))^2] + \alpha_{0,t}(f^{\theta^m})] \right\} \\
 &\geq \sup_{P^\theta \in \mathcal{P}} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^\theta}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^\theta)].
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 & \sup_{P^\theta \in \mathcal{P}} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^\theta}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^\theta)] \\
 &= \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^{\theta^*}}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^{\theta^*})].
 \end{aligned}$$

By minimax theorem (Theorem 5.1 in section 5), it implies that

$$\begin{aligned}
 \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} \mathcal{E}_g[(x(t) - \zeta(t))^2] &= \sup_{P^\theta \in \mathcal{P}} \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^\theta}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^\theta)] \\
 &= \inf_{\zeta(t) \in L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^{\theta^*}}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^{\theta^*})].
 \end{aligned}$$

This completes the proof. □

For the obtained  $\theta^*(t) = (\theta_1^*(t), \theta_2^*(t))$  in Theorem 3.6, set  $\widehat{\theta_i^*}(t) = E_{P^{\theta^*}}[\theta_i^*(t)|\mathcal{Z}_t]$ ,  $i = 1, 2$ .

**THEOREM 3.7.** *Set  $P(t) = E_{P^{\theta^*}}[(x(t) - \hat{x}(t))^2]$ . The MMSE  $\hat{x}(t)$  of (2.5) equals  $E_{P^{\theta^*}}[x(t)|\mathcal{Z}_t]$  and satisfies the following equation:*

$$(3.8) \quad \begin{cases} d\hat{x}(t) = (B(t)\hat{x}(t) + b(t) - \widehat{\theta_1^*}(t))dt + (P(t)H(t) - x(t)\widehat{\theta_2^*}(t) + \hat{x}(t)\widehat{\theta_2^*}(t))d\hat{I}(t), \\ \hat{x}(0) = x_0, \end{cases}$$

where  $\theta^*$  is obtained in Theorem 3.6,  $x(t)\widehat{\theta_2^*}(t) := E_{P^{\theta^*}}[x(t)\theta_2^*(t)|\mathcal{Z}_t]$  and the so called innovation process  $\hat{I}(t) := m(t) - \int_0^t (H(s)\hat{x}(s) + h(s) - \widehat{\theta_2^*}(s))ds$ ,  $0 \leq t \leq T$  is a  $\mathcal{Z}_t$ -measurable Brownian motion. The variance of the estimation error  $P(t)$  satisfies the following equation:

$$(3.9) \quad \begin{cases} \frac{dP(t)}{dt} = -E_{P^{\theta^*}}[(P(t)H(t) - x(t)\widehat{\theta_2^*}(t) + \hat{x}(t)\widehat{\theta_2^*}(t))(H(t)P(t) - \widehat{\theta_2^*}(t)x(t) \\ \quad + \widehat{\theta_2^*}(t)\hat{x}(t))] + 2E_{P^{\theta^*}}[-x(t)\widehat{\theta_1^*}(t) + \hat{x}(t)\widehat{\theta_1^*}(t)] + 2B(t)P(t) + 1, \\ P(0) = 0. \end{cases}$$

*Proof.* For the obtained optimal  $\theta^*(t) = (\theta_1^*(t), \theta_2^*(t))$  in Theorem 3.6, the system (2.4) and (2.5) can be reformulated correspondingly under  $P^{\theta^*}$ . In more detail, on the filtered probability space  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leq t \leq T}, P^{\theta^*})$ , the processes  $x(\cdot)$  and  $m(\cdot)$  satisfy the following equations:

$$(3.10) \quad \begin{cases} dx(t) = (B(t)x(t) + b(t) - \theta_1^*(t))dt + dw^{\theta_1^*}(t), \\ x(0) = x_0, \\ dm(t) = (H(t)x(t) + h(t) - \theta_2^*(t))dt + dv^{\theta_2^*}(t), \\ m(0) = 0. \end{cases}$$

We solve the minimum mean square estimation problem

$$(3.11) \quad \inf_{\zeta(t) \in L_{\mathcal{Z}_t}^{2+\epsilon}(\Omega, \mathbb{P}, \mathbb{R})} [E_{P^{\theta^*}}[(x(t) - \zeta(t))^2] + \alpha_{0,t}(P^{\theta^*})].$$

Since  $\alpha_{0,t}(P^{\theta^*})$  is a constant, we only need to consider the following optimization problem:

$$(3.12) \quad \inf_{\zeta(t) \in L_{\mathcal{Z}_t}^{2+\epsilon}(\Omega, \mathbb{P}, \mathbb{R})} E_{P^{\theta^*}}[(x(t) - \zeta(t))^2].$$

In [31], Liptser and Shiryaev studied the optimal estimator of the following problem:

$$(3.13) \quad \inf_{\zeta(t) \in L_{\mathcal{Z}_t}^{2+\epsilon}(\Omega, P^{\theta^*}, \mathbb{R})} E_{P^{\theta^*}}[(x(t) - \zeta(t))^2].$$

By Theorem 8.1 in [31], the optimal estimator  $\hat{x}(t) = E_{P^{\theta^*}}[x(t)|\mathcal{Z}_t]$  of (3.13) satisfies (3.8). Since  $B(t)$ ,  $H(t)$ ,  $b(t)$ , and  $h(t)$  are uniformly bounded, deterministic functions, and  $\theta^*$  is bounded by Theorem 6.3 (see Chapter 1 in [41]), and the solution  $\hat{x}(t)$  to (3.8) also belongs to  $L_{\mathcal{Z}_t}^{2+\epsilon}(\Omega, \mathbb{P}, \mathbb{R})$ . It yields that  $\hat{x}(t)$  is the optimal solution of problem (3.12) at time  $t \in [0, T]$ . This completes the proof.  $\square$

COROLLARY 3.8. *If  $\theta^*(t)$  is adapted to  $\mathcal{Z}_t$ , then  $\hat{x}(t)$  satisfies the following equation:*

$$(3.14) \quad \begin{cases} d\hat{x}(t) = (B(t)\hat{x}(t) + b(t) - \theta_1^*(t))dt + P(t)H(t)d\hat{I}(t), \\ \hat{x}(0) = x_0, \end{cases}$$

where  $P(t)$  satisfies the following Riccati equation:

$$(3.15) \quad \begin{cases} \frac{dP(t)}{dt} = 2B(t)P(t) - P^2(t)H(t)^2 + 1, \\ P(0) = 0. \end{cases}$$

According to the classical Kalman-Bucy filter theory, the optimal filter  $\bar{x}(t) = \mathbb{E}[x(t)|\mathcal{Z}_t]$  is the solution of the following equation:

$$(3.16) \quad \begin{cases} d\bar{x}(t) = (B(t)\bar{x}(t) + b(t))dt + P(t)H(t)dI(t), \\ \bar{x}(0) = x_0, \end{cases}$$

where

$$I(t) = m(t) - \int_0^t (H(s)\bar{x}(s) + h(s))ds.$$

COROLLARY 3.9. *Let  $A(t, s) = \exp(\int_s^t (B(r) - P(r)(H(r))^2)dr) \forall 0 \leq s < t \leq T$ . If the optimal  $\theta^*(t)$  is adapted to subfiltration  $\mathcal{Z}_t$ , with (3.8) and (3.16), then the optimal estimator  $\hat{x}(t)$  for any time  $t \in [0, T]$  can be expressed as*

$$(3.17) \quad \hat{x}(t) = \bar{x}(t) + \int_0^t (P(s)H(s)\theta_2^*(s) - \theta_1^*(s))A(t, s)ds,$$

where  $\bar{x}(t)$  is defined by (3.16).

*Proof.* Firstly, we conjecture  $\hat{x}(t)$  has the following structure:

$$\hat{x}(t) = \bar{x}(t) + M(t),$$

where  $M(t)$  is the variable to be determined.

Secondly, based on (3.14) and (3.16), we can verify that  $M(t)$  satisfies the following ODE after a simple calculation:

$$(3.18) \quad \begin{cases} dM(t) = (B(t) - P(t)H(t)^2)M(t)dt + (P(t)H(t)\theta_2^*(t) - \theta_1^*(t))dt, \\ M(0) = 0. \end{cases}$$

It reduces that

$$M(t) = \int_0^t (P(s)H(s)\theta_2^*(s) - \theta_1^*(s))A(t, s)ds.$$

This completes the proof. □

*Remark 3.10.* In [34], the authors gave a remark to explain that they cannot explicitly evaluate the optimal uncertainty parameter and estimator about the USS in the Kalman filter sense. In our context, it is still difficult to give the construction of  $\theta^*$  explicitly. Based on this reason, we come down to using the numerical calculation methods to solve this problem in the future.

**4. MMSE under general convex operators on  $L_{\mathcal{F}}^p(\mathbb{P})$ .** In section 3, we boil down the calculation of the Kalman–Bucy filter under uncertainty to solving a minimum mean square estimation problem under the convex  $g$ -expectation. The worst-case prior  $P^{\theta^*}$  is obtained, and the corresponding filtering equation (3.8) is deduced.

It is an interesting question whether there are similar results for general convex operators. So in this section, we investigate the minimum mean square estimation problem under general convex operators on  $L_{\mathcal{F}}^p(\mathbb{P})$  and obtain the existence and uniqueness results of the MMSE.

**4.1. General convex operators on  $L_{\mathcal{F}}^p(\mathbb{P})$ .** For a given probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , we denote the set of all  $\mathcal{F}$ -measurable  $p$ th power integrable random variables by  $L^p(\Omega, \mathcal{F}, \mathbb{P})$ . Sometimes we use  $L_{\mathcal{F}}^p(\mathbb{P})$  for short. Let  $\mathcal{C}$  be a sub  $\sigma$ -algebra of  $\mathcal{F}$ .  $L_{\mathcal{C}}^p(\mathbb{P})$  denotes the set of all the  $p$ th power integrable  $\mathcal{C}$ -measurable random variables. In this paper, we only consider the case that  $1 < p \leq 2$ .

Let  $\mathcal{M}$  denote the set of probability measures absolutely continuous with respect to  $\mathbb{P}$ . For  $P \in \mathcal{M}$ , we will use  $f^P$  to denote the Radon–Nikodym derivative  $\frac{dP}{d\mathbb{P}}$  and  $E_P[\cdot]$  to denote the expectation under  $P$ . Especially, the expectation under  $\mathbb{P}$  is denoted as  $\mathbb{E}[\cdot]$ . For a sub  $\sigma$ -algebra  $\mathcal{C}$  of  $\mathcal{F}$  and  $P \in \mathcal{M}$ , define  $f_{\mathcal{C}}^P = \mathbb{E}[f^P | \mathcal{C}]$ .

**DEFINITION 4.1.** A convex operator is an operator  $\rho(\cdot) : L_{\mathcal{F}}^p(\mathbb{P}) \mapsto \mathbb{R}$  satisfying

- (i) *monotonicity:* for any  $\xi_1, \xi_2 \in L_{\mathcal{F}}^p(\mathbb{P})$ ,  $\rho(\xi_1) \geq \rho(\xi_2)$  if  $\xi_1 \geq \xi_2$ ;
- (ii) *constant invariance:*  $\rho(\xi + c) = \rho(\xi) + c$  for any  $\xi \in L_{\mathcal{F}}^p(\mathbb{P})$  and  $c \in \mathbb{R}$ ;
- (iii) *convexity:* for any  $\xi_1, \xi_2 \in L_{\mathcal{F}}^p(\mathbb{P})$  and  $\lambda \in [0, 1]$ ,  $\rho(\lambda\xi_1 + (1 - \lambda)\xi_2) \leq \lambda\rho(\xi_1) + (1 - \lambda)\rho(\xi_2)$ .

**DEFINITION 4.2.** A convex operator  $\rho(\cdot)$  is called normalized if  $\rho(0) = 0$ .

*Remark 4.3.* In this paper, we will always assume the convex operator is normalized. Moreover, if we define  $\rho'(\xi) = \rho(-\xi)$ , then  $\rho'(\cdot)$  is a convex risk measure on  $L_{\mathcal{F}}^p(\mathbb{P})$ .

If  $\rho(\cdot)$  is a convex operator, then by Proposition 2.10 and Theorem 2.11 in [28], for any random variable  $\xi \in L_{\mathcal{F}}^p(\mathbb{P})$ , there exists a set  $\mathcal{P}$  such that  $\rho(\cdot)$  can be represented as

$$\rho(\xi) = \sup_{P \in \mathcal{P}} [E_P[\xi] - \alpha(P)],$$

where  $\alpha(P) := \sup_{\zeta \in \mathcal{A}_\rho} E_P[\zeta]$ ,  $\mathcal{A}_\rho := \{\zeta \in L_{\mathcal{F}}^p(\mathbb{P}); \rho(\zeta) \leq 0\}$  called acceptance set,  $\mathcal{P} := \{P \in \mathcal{M}; f^P \in L_{\mathcal{F}}^q(\mathbb{P}), \alpha(P) < \infty\}$ . Moreover,  $\mathcal{D} := \{f^P; P \in \mathcal{P}\}$  is norm-bounded in  $L_{\mathcal{F}}^q(\mathbb{P})$  and  $\sigma(L_{\mathcal{F}}^q(\mathbb{P}), L_{\mathcal{F}}^p(\mathbb{P}))$ -compact, where  $\sigma(L_{\mathcal{F}}^q(\mathbb{P}), L_{\mathcal{F}}^p(\mathbb{P}))$  denotes the weak topology defined on  $L_{\mathcal{F}}^q(\mathbb{P})$  and  $\frac{1}{p} + \frac{1}{q} = 1$ . The set  $\mathcal{P}$  is called the representation set of  $\rho(\cdot)$ . Since  $\alpha(\cdot)$  is a convex function defined on  $\mathcal{M}$ ,  $\mathcal{P}$  is a convex set.

*Remark 4.4.* Note that  $\alpha(P) = \sup_{\zeta \in \mathcal{A}_\rho} E_P[\zeta] = \sup_{\zeta \in \mathcal{A}_\rho} \mathbb{E}[f^P \zeta]$ . By abuse of notation, we sometimes write  $\alpha(f^P)$  instead of  $\alpha(P)$ .

**DEFINITION 4.5.** The set  $\mathcal{P}$  is called stable if for any element  $P \in \mathcal{P}$  and any sub  $\sigma$ -algebra  $\mathcal{C}$  of  $\mathcal{F}$ ,  $\frac{f^P}{f_{\mathcal{C}}^P}$  still lies in the set  $\mathcal{D}$ .

**DEFINITION 4.6.** The convex operator  $\rho(\cdot)$  is called regular if all the elements in set  $\mathcal{P}$  are equivalent to  $\mathbb{P}$ .

**DEFINITION 4.7.** A convex operator  $\rho(\cdot)$  is called stable if its representation set  $\mathcal{P}$  is stable.

For a given  $\xi \in L^4_{\mathcal{F}}(\mathbb{P})$ , when we only know the information  $\mathcal{C}$ , we want to find the MMSE of  $\xi$  under the convex operator  $\rho(\cdot)$ . In more detail, we will solve the following optimization problem.

**Problem:** For a given  $\xi \in L^4_{\mathcal{F}}(\mathbb{P})$ , find a  $\hat{\eta} \in L^{2p}_{\mathcal{C}}(\mathbb{P})$  such that

$$(4.1) \quad \rho((\xi - \hat{\eta})^2) = \inf_{\eta \in L^{2p}_{\mathcal{C}}(\mathbb{P})} \rho((\xi - \eta)^2).$$

The optimal solution  $\hat{\eta}$  of (4.1) is called the MMSE, and we will denote it by  $\rho(\xi|\mathcal{C})$ .

*Remark 4.8.* If we set  $\mathcal{C} = \mathcal{Z}_t$  and  $p = 1 + \frac{\epsilon}{2}$  with  $\epsilon \in (0, 1)$ , then  $L^{2p}_{\mathcal{C}}(\mathbb{P})$  is just the space  $L^{2+\epsilon}_{\mathcal{Z}_t}(\Omega, \mathbb{P}, \mathbb{R}^n)$  in subsection 2.2.

**4.2. Existence and uniqueness results.** In this section, we study the existence and uniqueness of the MMSE for (4.1). We first give the following assumption.

*Assumption 4.9.* The convex operator  $\rho(\cdot)$  is stable and regular.

**4.2.1. Existence.**

LEMMA 4.10. For any given real number  $\gamma \geq 2$ , if  $\xi \in L^{\gamma p}_{\mathcal{F}}(\mathbb{P})$ , then we have  $\sup_{P \in \mathcal{P}} E_P[|\xi|^{\frac{\gamma p}{2}}] < \infty$ .

*Proof.* Since  $\{f^P; P \in \mathcal{P}\}$  is norm-bounded in  $L^q_{\mathcal{F}}(\mathbb{P})$  and  $1 < p \leq 2$ , we have

$$\begin{aligned} \sup_{P \in \mathcal{P}} E_P[|\xi|^{\frac{\gamma p}{2}}] &= \sup_{P \in \mathcal{P}} \mathbb{E}[f^P |\xi|^{\frac{\gamma p}{2}}] \leq \sup_{P \in \mathcal{P}} \|f^P\|_{L^q} \|\xi^{\frac{\gamma p}{2}}\|_{L^p} \\ &\leq \sup_{P \in \mathcal{P}} \|f^P\|_{L^q} [\mathbb{E}[|\xi|^{\gamma p}]]^{\frac{1}{2}} < \infty. \end{aligned}$$

This completes the proof. □

LEMMA 4.11. Suppose that Assumption 4.9 holds. Then for any  $P \in \mathcal{P}$ ,  $\xi \in L^p_{\mathcal{F}}(\mathbb{P})$ , and sub  $\sigma$ -algebra  $\mathcal{C}$  of  $\mathcal{F}$ , there exists a  $\bar{P} \in \mathcal{P}$  such that  $E_{\bar{P}}[\xi] = \mathbb{E}[E_P[\xi|\mathcal{C}]]$ .

*Proof.* It is obvious that

$$\mathbb{E}[E_P[\xi|\mathcal{C}]] = \mathbb{E} \left[ \frac{\mathbb{E}[\xi f^P|\mathcal{C}]}{\mathbb{E}[f^P|\mathcal{C}]} \right] = \mathbb{E}[\mathbb{E} \left[ \xi \frac{f^P}{f^{\bar{C}}} \middle| \mathcal{C} \right]] = \mathbb{E} \left[ \xi \frac{f^P}{f^{\bar{C}}} \right].$$

By Definition 4.5, there exists a  $\bar{P} \in \mathcal{P}$  such that  $\frac{d\bar{P}}{dP} = \frac{f^P}{f^{\bar{C}}}$  which implies that  $E_{\bar{P}}[\xi] = \mathbb{E}[E_P[\xi|\mathcal{C}]]$ . This completes the proof. □

PROPOSITION 4.12. Suppose that Assumption 4.9 holds. If  $\xi \in L^{4p}_{\mathcal{F}}(\mathbb{P})$ , then there exists a constant  $M$  such that for any probability measure  $P \in \mathcal{P}$ ,

$$\inf_{\eta \in L^{2p}_{\mathcal{C}}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)] = \inf_{\eta \in L^{2p, M}_{\mathcal{C}}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)],$$

where  $L^{2p, M}_{\mathcal{C}}(\mathbb{P})$  denotes all the elements in  $L^{2p}_{\mathcal{C}}(\mathbb{P})$  which are norm-bounded by the constant  $M$ .

*Proof.* Set  $\mathbb{G} = \{E_P[\xi|\mathcal{C}]; P \in \mathcal{P}\}$ . For any  $P \in \mathcal{P}$ , we have  $\mathbb{E}[(E_P[\xi|\mathcal{C}])^{2p}] \leq \mathbb{E}[E_P[\xi^{2p}|\mathcal{C}]]$ . By Lemma 4.11, there exists a  $\bar{P} \in \mathcal{P}$  such that  $E_{\bar{P}}[\xi^{2p}] = \mathbb{E}[E_P[\xi^{2p}|\mathcal{C}]]$ . By Lemma 4.10, there exists a constant  $M_1$  such that  $\sup_{P \in \mathcal{P}} E_P[\xi^{2p}] \leq M_1$ . Then  $\mathbb{G} \subset L^{2p, M}_{\mathcal{C}}(\mathbb{P})$ , where  $M = M_1^{\frac{1}{2p}}$ . By the project property of conditional expectations, for any  $P \in \mathcal{P}$  and  $\eta \in L^{2p}_{\mathcal{C}}(\mathbb{P})$ , we have that

$$E_P[(\xi - E_P[\xi|\mathcal{C}])^2] \leq E_P[(\xi - \eta)^2],$$

which leads to

$$\inf_{\eta \in L_C^{2p}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)] \geq \inf_{\eta' \in \mathbb{G}} [E_P[(\xi - \eta')^2] - \alpha(P)].$$

On the other hand, the inverse inequality is obviously true. Then the following equality holds for any  $P \in \mathcal{P}$ :

$$\inf_{\eta \in L_C^{2p}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)] = \inf_{\eta \in \mathbb{G}} [E_P[(\xi - \eta)^2] - \alpha(P)].$$

Since  $\mathbb{G} \subset L_C^{2p,M}(\mathbb{P}) \subset L_C^{2p}(\mathbb{P})$ , it follows that

$$\inf_{\eta \in L_C^{2p}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)] = \inf_{\eta \in L_C^{2p,M}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)].$$

This completes the proof.  $\square$

By Proposition 4.12, it is easy to see that

$$\sup_{P \in \mathcal{P}} \inf_{\eta \in L_C^{2p}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)] = \sup_{P \in \mathcal{P}} \inf_{\eta \in L_C^{2p,M}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)].$$

LEMMA 4.13.  $\alpha(\cdot)$  is a lower semicontinuous (l.s.c.) function on the topology space  $(\mathcal{D}, \sigma(L_{\mathcal{F}}^q(\mathbb{P}), L_{\mathcal{F}}^p(\mathbb{P})))$ .

*Proof.* For any fixed random variable  $\zeta \in \mathcal{A}_\rho$ , define  $\varphi(\zeta, f^P) = \mathbb{E}[f^P \zeta]$ , where  $f^P$  belongs to  $\mathcal{D}$ . Then  $\varphi(\zeta, \cdot)$  is a continuous function on the topology space  $(\mathcal{D}, \sigma(L_{\mathcal{F}}^q(\mathbb{P}), L_{\mathcal{F}}^p(\mathbb{P})))$ . Take a sequence  $\{f^{P_n}\}_{n \geq 1} \rightarrow_{L_{\mathcal{F}}^q(\mathbb{P})} \bar{f}$ , where  $\{f^{P_n}\}_{n \geq 1}$  and  $\bar{f}$  all belong to the set  $\mathcal{D}$ ; then

$$\liminf_{n \rightarrow \infty} \alpha(f^{P_n}) = \liminf_{n \rightarrow \infty} \sup_{\zeta \in \mathcal{A}_\rho} \mathbb{E}[f^{P_n} \zeta] \geq \sup_{\zeta \in \mathcal{A}_\rho} \liminf_{n \rightarrow \infty} \mathbb{E}[f^{P_n} \zeta].$$

Next, take a subsequence  $\{f^{P_{n_i}}\}_{i \geq 1}$  of  $\{f^{P_n}\}_{n \geq 1}$  such that  $\{f^{P_{n_i}}\}_{i \geq 1} \rightarrow_{\mathbb{P}\text{-a.s.}} \bar{f}$ ; we then have

$$\liminf_{n \rightarrow \infty} \alpha(f^{P_n}) \geq \alpha(\bar{f}).$$

Therefore,  $\alpha(\cdot)$  is an l.s.c. function on the topology space  $(\mathcal{D}, \sigma(L_{\mathcal{F}}^q(\mathbb{P}), L_{\mathcal{F}}^p(\mathbb{P})))$ . This completes the proof.  $\square$

For  $\xi \in L_{\mathcal{F}}^{4p}(\mathbb{P})$ ,  $\eta \in L_C^{2p}(\mathbb{P})$ , and  $P \in \mathcal{P}$ , define

$$l(\xi, \eta, f^P) = \mathbb{E}[f^P(\xi - \eta)^2] - \alpha(f^P).$$

LEMMA 4.14. For any random variables  $\xi \in L_{\mathcal{F}}^{4p}(\mathbb{P})$  and  $\eta \in L_C^{2p}(\mathbb{P})$ ,  $l(\xi, \eta, \cdot)$  is an upper semicontinuous (u.s.c.) function on the topology space  $(\mathcal{D}, \sigma(L_{\mathcal{F}}^q(\mathbb{P}), L_{\mathcal{F}}^p(\mathbb{P})))$ .

*Proof.* Since  $\xi \in L_{\mathcal{F}}^{4p}(\mathbb{P})$  and  $\eta \in L_C^{2p}(\mathbb{P})$ , then  $(\xi - \eta)^2 \in L_{\mathcal{F}}^p(\mathbb{P})$  which implies that  $\mathbb{E}[f^P(\xi - \eta)^2]$  is a continuous function with respect to  $f^P$  on the topology space  $(\mathcal{D}, \sigma(L_{\mathcal{F}}^q(\mathbb{P}), L_{\mathcal{F}}^p(\mathbb{P})))$ . By Lemma 4.13,  $\alpha(\cdot)$  is an l.s.c. function on the topology space  $(\mathcal{D}, \sigma(L_{\mathcal{F}}^q(\mathbb{P}), L_{\mathcal{F}}^p(\mathbb{P})))$ . Thus,  $l(\xi, \eta, \cdot)$  is an u.s.c. function on the topology space  $(\mathcal{D}, \sigma(L_{\mathcal{F}}^q(\mathbb{P}), L_{\mathcal{F}}^p(\mathbb{P})))$ . This completes the proof.  $\square$

PROPOSITION 4.15. Suppose that Assumption 4.9 holds. Then for a given  $\xi \in L_{\mathcal{F}}^{4p}(\Omega, \mathbb{P})$ , there exists a  $\hat{P} \in \mathcal{P}$  such that

$$\inf_{\eta \in L_C^{2p,M}(\mathbb{P})} [E_{\hat{P}}[(\xi - \eta)^2] - \alpha(\hat{P})] = \sup_{P \in \mathcal{P}} \inf_{\eta \in L_C^{2p,M}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)],$$

where  $M$  is the constant given in Proposition 4.12.



*Proof.* Define

$$\beta = \sup_{P \in \mathcal{P}} \inf_{\eta \in L_C^{2p,M}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)] = \sup_{f^P \in \mathcal{D}} \inf_{\eta \in L_C^{2p,M}(\mathbb{P})} [\mathbb{E}[f^P(\xi - \eta)^2] - \alpha(f^P)].$$

Take a sequence  $\{f^{P_n}; P_n \in \mathcal{P}\}_{n \geq 1}$  such that

$$\inf_{\eta \in L_C^{2p,M}(\mathbb{P})} [\mathbb{E}[f^{P_n}(\xi - \eta)^2] - \alpha(f^{P_n})] \geq \beta - \frac{1}{2^n}.$$

Since  $\mathcal{D}$  is a weakly compact set, we can take a subsequence  $\{f^{P_{n_i}}\}_{i \geq 1}$  which weakly converges to some  $f^{\hat{P}} \in L_{\mathcal{F}}^q(\mathbb{P})$ . Therefore,  $\hat{P} \in \mathcal{P}$ , and there exists a sequence  $\{f^{\tilde{P}_i} \in \mathbf{conv}(f^{P_{n_i}}, f^{P_{n_i+1}}, \dots)\}_{i \geq 1}$  such that  $f^{\tilde{P}_i}$  converges to  $f^{\hat{P}}$  in  $L_{\mathcal{F}}^q(\mathbb{P})$ -norm by Theorem 5.4 in section 5, where  $\mathbf{conv}(A) = \{\sum_{i=1}^n \alpha_i x_i | x_i \in A, \alpha_i > 0, \sum_{i=1}^n \alpha_i = 1, n \in \mathbb{N}\}$ .

For any  $\eta \in L_C^{2p,M}(\mathbb{P})$ ,

$$\lim_{i \rightarrow \infty} \mathbb{E}[|f^{\tilde{P}_i}(\xi - \eta)^2 - f^{\hat{P}}(\xi - \eta)^2|] \leq \lim_{i \rightarrow \infty} \|(f^{\tilde{P}_i} - f^{\hat{P}})\|_{L^q(\mathbb{P})} \|(\xi - \eta)^2\|_{L^p(\mathbb{P})} = 0,$$

which leads to

$$\lim_{i \rightarrow \infty} \mathbb{E}[f^{\tilde{P}_i}(\xi - \eta)^2] = \mathbb{E}[f^{\hat{P}}(\xi - \eta)^2].$$

On the other hand, according to Lemma 4.14,  $-\alpha(\cdot)$  is an u.s.c. function. It reduces that

$$-\alpha(f^{\hat{P}}) \geq \limsup_{i \rightarrow \infty} (-\alpha(f^{\tilde{P}_i})).$$

Then,

$$[\mathbb{E}[f^{\hat{P}}(\xi - \eta)^2] - \alpha(f^{\hat{P}})] \geq \limsup_{i \rightarrow \infty} [\mathbb{E}[f^{\tilde{P}_i}(\xi - \eta)^2] - \alpha(f^{\tilde{P}_i})].$$

Since

$$[\mathbb{E}[f^{\tilde{P}_i}(\xi - \eta)^2] - \alpha(f^{\tilde{P}_i})] \geq \inf_{\tilde{\eta} \in L_C^{2p,M}(\mathbb{P})} [\mathbb{E}[f^{\tilde{P}_i}(\xi - \tilde{\eta})^2] - \alpha(f^{\tilde{P}_i})]$$

for any  $\eta \in L_C^{2p,M}(\mathbb{P})$ , we have that

$$\limsup_{i \rightarrow \infty} [\mathbb{E}[f^{\tilde{P}_i}(\xi - \eta)^2] - \alpha(f^{\tilde{P}_i})] \geq \limsup_{i \rightarrow \infty} \inf_{\tilde{\eta} \in L_C^{2p,M}(\mathbb{P})} [\mathbb{E}[f^{\tilde{P}_i}(\xi - \tilde{\eta})^2] - \alpha(f^{\tilde{P}_i})].$$

It yields that

$$\begin{aligned} (4.2) \quad \inf_{\eta \in L_C^{2p,M}(\mathbb{P})} [\mathbb{E}[f^{\hat{P}}(\xi - \eta)^2] - \alpha(f^{\hat{P}})] &\geq \inf_{\eta \in L_C^{2p,M}(\mathbb{P})} \limsup_{i \rightarrow \infty} [\mathbb{E}[f^{\tilde{P}_i}(\xi - \eta)^2] - \alpha(f^{\tilde{P}_i})] \\ &\geq \limsup_{i \rightarrow \infty} \inf_{\tilde{\eta} \in L_C^{2p,M}(\mathbb{P})} [\mathbb{E}[f^{\tilde{P}_i}(\xi - \tilde{\eta})^2] - \alpha(f^{\tilde{P}_i})]. \end{aligned}$$

As  $\alpha(\cdot)$  is a convex function and  $f^{\tilde{P}_i} \in \mathbf{conv}(f^{P_{n_i}}, f^{P_{n_i+1}}, \dots)$ , we have

$$(4.3) \quad \limsup_{i \rightarrow \infty} \inf_{\tilde{\eta} \in L_C^{2p,M}(\mathbb{P})} [\mathbb{E}[f^{\tilde{P}_i}(\xi - \tilde{\eta})^2] - \alpha(f^{\tilde{P}_i})] \geq \beta.$$

Combining (4.2) and (4.3), we obtain the result. □

COROLLARY 4.16. *Suppose that Assumption 4.9 holds. Then for a given  $\xi \in L_{\mathcal{F}}^{4p}(\mathbb{P})$ , there exists a  $\hat{P} \in \mathcal{P}$  such that*

$$\inf_{\eta \in L_{\mathcal{C}}^{2p}(\mathbb{P})} [E_{\hat{P}}[(\xi - \eta)^2] - \alpha(\hat{P})] = \sup_{P \in \mathcal{P}} \inf_{\eta \in L_{\mathcal{C}}^{2p}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)].$$

*Proof.* Choose  $\hat{P}$  as in Proposition 4.15. By Propositions 4.12 and 4.15, the following relations hold

$$\begin{aligned} \sup_{P \in \mathcal{P}} \inf_{\eta \in L_{\mathcal{C}}^{2p}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)] &= \sup_{P \in \mathcal{P}} \inf_{\eta \in L_{\mathcal{C}}^{2p,M}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)] \\ &= \inf_{\eta \in L_{\mathcal{C}}^{2p,M}(\mathbb{P})} [E_{\hat{P}}[(\xi - \eta)^2] - \alpha(\hat{P})] = \inf_{\eta \in L_{\mathcal{C}}^{2p}(\mathbb{P})} [E_{\hat{P}}[(\xi - \eta)^2] - \alpha(\hat{P})]. \end{aligned}$$

This completes the proof. □

THEOREM 4.17 (existence theorem). *Suppose that Assumption 4.9 holds. Then there exists an  $\hat{\eta} \in L_{\mathcal{C}}^{2p}(\mathbb{P})$  which solves (4.1).*

*Proof.* For given  $\xi \in L_{\mathcal{F}}^{4p}(\mathbb{P})$ ,  $\eta \in L_{\mathcal{C}}^{2p}(\mathbb{P})$ , and  $P \in \mathcal{P}$ , it is easy to check that  $l(\xi, \cdot, f^P)$  is convex on  $L_{\mathcal{C}}^{2p}(\mathbb{P})$  and  $l(\xi, \eta, \cdot)$  is concave on  $L_{\mathcal{F}}^q(\mathbb{P})$ . As  $\mathcal{D}$  is  $\sigma(L_{\mathcal{F}}^q(\mathbb{P}), L_{\mathcal{F}}^p(\mathbb{P}))$ -compact and  $l(\xi, \eta, \cdot)$  is u.s.c. on topology space  $(L_{\mathcal{F}}^q(\mathbb{P}), \sigma(L_{\mathcal{F}}^q(\mathbb{P}), L_{\mathcal{F}}^p(\mathbb{P})))$  by Lemma 4.14, we have

$$\begin{aligned} \inf_{\eta \in L_{\mathcal{C}}^{2p}(\mathbb{P})} \max_{P \in \mathcal{P}} [E_P[(\xi - \eta)^2] - \alpha(P)] &= \max_{P \in \mathcal{P}} \inf_{\eta \in L_{\mathcal{C}}^{2p}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)]; \\ \inf_{\eta \in L_{\mathcal{C}}^{2p,M}(\mathbb{P})} \max_{P \in \mathcal{P}} [E_P[(\xi - \eta)^2] - \alpha(P)] &= \max_{P \in \mathcal{P}} \inf_{\eta \in L_{\mathcal{C}}^{2p,M}(\mathbb{P})} [E_P[(\xi - \eta)^2] - \alpha(P)] \end{aligned}$$

by Proposition 4.15, Corollary 4.16, and Theorem 5.1 in section 5. With the help of Proposition 4.12,

$$\inf_{\eta \in L_{\mathcal{C}}^{2p}(\mathbb{P})} \max_{P \in \mathcal{P}} [E_P[(\xi - \eta)^2] - \alpha(P)] = \inf_{\eta \in L_{\mathcal{C}}^{2p,M}(\mathbb{P})} \max_{P \in \mathcal{P}} [E_P[(\xi - \eta)^2] - \alpha(P)].$$

Therefore, we can take a sequence  $\{\eta_n; n \in \mathbb{N}\} \subset L_{\mathcal{C}}^{2p,M}(\mathbb{P})$  such that

$$\rho((\xi - \eta_n)^2) < \beta + \frac{1}{2^n},$$

where  $\beta := \inf_{\eta \in L_{\mathcal{C}}^{2p}(\mathbb{P})} \rho((\xi - \eta)^2)$ . Since  $L_{\mathcal{C}}^{2p,M}(\mathbb{P})$  is a weakly compact set, we can take a subsequence  $\{\eta_{n_i}\}_{i \in \mathbb{N}}$  of  $\{\eta_n\}_{n \in \mathbb{N}}$  which weakly converges to some  $\hat{\eta} \in L_{\mathcal{C}}^{2p,M}(\mathbb{P})$ . By Theorem 5.4 in section 5, there exists a sequence  $\{\tilde{\eta}_i \in \mathbf{conv}(\eta_{n_i}, \eta_{n_{i+1}}, \dots)\}_{i \in \mathbb{N}}$  such that  $\tilde{\eta}_i$  converges to  $\hat{\eta}$  in  $L_{\mathcal{C}}^{2p}(\mathbb{P})$ -norm. Then

$$\begin{aligned} (4.4) \quad \rho((\xi - \hat{\eta})^2) &= \rho((\xi - \tilde{\eta}_i + \tilde{\eta}_i - \hat{\eta})^2) \\ &= \sup_{P \in \mathcal{P}} [E_P[(\xi - \tilde{\eta}_i)^2 + (\tilde{\eta}_i - \hat{\eta})^2 + 2(\xi - \tilde{\eta}_i)(\tilde{\eta}_i - \hat{\eta})] - \alpha(P)] \\ &\leq \sup_{P \in \mathcal{P}} [E_P[(\xi - \tilde{\eta}_i)^2] - \alpha(P)] + \sup_{P \in \mathcal{P}} E_P[(\tilde{\eta}_i - \hat{\eta})^2] \\ &\quad + 2(\xi - \tilde{\eta}_i)(\tilde{\eta}_i - \hat{\eta}) \\ &= \rho((\xi - \tilde{\eta}_i)^2) + \sup_{P \in \mathcal{P}} E_P[-(\tilde{\eta}_i - \hat{\eta})^2 + 2(\xi - \hat{\eta})(\tilde{\eta}_i - \hat{\eta})] \\ &\leq \beta + \frac{1}{2^{i-1}} + 2 \sup_{P \in \mathcal{P}} \|f^P\|_{L^q} \|\tilde{\eta}_i - \hat{\eta}\|_{L^{2p}} \|\xi - \hat{\eta}\|_{L^{2p}}. \end{aligned}$$

As (4.4) holds for any  $i \geq 1$ , we have that  $\rho((\xi - \hat{\eta})^2) = \beta$ . □

**4.2.2. Uniqueness.** In this subsection, we prove that the optimal solution of problem (4.1) is unique.

PROPOSITION 4.18. *Suppose that Assumption 4.9 holds. If  $\hat{\eta}$  is an optimal solution of (4.1), then there exists a  $\hat{P} \in \mathcal{P}$  such that  $\hat{\eta} = E_{\hat{P}}[\xi|\mathcal{C}]$ .*

*Proof.* If  $\hat{\eta}$  is an optimal solution of (4.1), then there exists a  $\hat{P} \in \mathcal{P}$  such that

$$\begin{aligned} & \sup_{P \in \mathcal{P}} [\mathbb{E}[f^P(\xi - \hat{\eta})^2] - \alpha(P)] \\ &= \min_{\eta \in L_C^{2p}(\mathbb{P})} \sup_{P \in \mathcal{P}} [\mathbb{E}[f^P(\xi - \eta)^2] - \alpha(P)] \\ &= \sup_{P \in \mathcal{P}} \min_{\eta \in L_C^{2p}(\mathbb{P})} [\mathbb{E}[f^P(\xi - \eta)^2] - \alpha(P)] \\ &= \max_{P \in \mathcal{P}} \inf_{\eta \in L_C^{2p}(\mathbb{P})} [\mathbb{E}[f^P(\xi - \eta)^2] - \alpha(P)] \\ &= \inf_{\eta \in L_C^{2p}(\mathbb{P})} [\mathbb{E}[f^{\hat{P}}(\xi - \eta)^2] - \alpha(\hat{P})] \end{aligned}$$

by Corollary 4.16, Theorem 4.17, and Theorem 5.1 in section 5. Thus, by Theorem 5.2 in section 5,  $(\hat{\eta}, \hat{P})$  is a saddle point, i.e., for all  $P \in \mathcal{P}, \eta \in L_C^{2p}(\mathbb{P})$ , we have

$$\mathbb{E}[f^P(\xi - \hat{\eta})^2] - \alpha(P) \leq \mathbb{E}[f^{\hat{P}}(\xi - \hat{\eta})^2] - \alpha(\hat{P}) \leq \mathbb{E}[f^{\hat{P}}(\xi - \eta)^2] - \alpha(\hat{P}).$$

This shows that if  $\hat{\eta}$  is an optimal solution, then there exists a  $\hat{P} \in \mathcal{P}$  such that  $\hat{\eta} = E_{\hat{P}}[\xi|\mathcal{C}]$  by the project property of conditional expectations.  $\square$

THEOREM 4.19 (uniqueness theorem). *Suppose that Assumption 4.9 holds. Then the optimal solution of (4.1) is unique.*

*Proof.* Suppose that there exist two optimal solutions  $\hat{\eta}_1$  and  $\hat{\eta}_2$ . Denote the corresponding probabilities in Proposition 4.18 by  $\hat{P}_1$  and  $\hat{P}_2$ , respectively. Then  $\hat{\eta}_1 = E_{\hat{P}_1}[\xi|\mathcal{C}]$  and  $\hat{\eta}_2 = E_{\hat{P}_2}[\xi|\mathcal{C}]$ . For  $\lambda \in (0, 1)$ , set

$$\begin{aligned} P^\lambda &= \lambda\hat{P}_1 + (1 - \lambda)\hat{P}_2, \\ \lambda_{\hat{P}_1} &= \lambda E_{P^\lambda} \left[ \frac{d\hat{P}_1}{dP^\lambda} | \mathcal{C} \right], \\ \lambda_{\hat{P}_2} &= (1 - \lambda) E_{P^\lambda} \left[ \frac{d\hat{P}_2}{dP^\lambda} | \mathcal{C} \right]. \end{aligned}$$

It is easy to verify that  $\lambda_{\hat{P}_1} + \lambda_{\hat{P}_2} = 1$  and  $E_{P^\lambda}[\xi|\mathcal{C}] = \lambda_{\hat{P}_1}\hat{\eta}_1 + \lambda_{\hat{P}_2}\hat{\eta}_2$ . Noticing that  $E_{\hat{P}_i}[(\xi - \hat{\eta}_i)|\mathcal{C}] = 0, i = 1, 2$ , then we have the following equation (details of the calculation can be found in Lemma 5.5 in section 5):

$$\begin{aligned} & E_{P^\lambda}[(\xi - E_{P^\lambda}[\xi|\mathcal{C}])^2] - \alpha(P^\lambda) \\ &= \lambda E_{\hat{P}_1}[(\xi - \hat{\eta}_1)^2 - \lambda_{\hat{P}_2}((\xi - \hat{\eta}_1)^2 - (\xi - \hat{\eta}_2)^2 + (\hat{\eta}_1 - \hat{\eta}_2)^2) + \lambda_{\hat{P}_2}^2(\hat{\eta}_1 - \hat{\eta}_2)^2] \\ &\quad + (1 - \lambda) E_{\hat{P}_2}[\lambda_{\hat{P}_1}((\xi - \hat{\eta}_1)^2 - (\xi - \hat{\eta}_2)^2 - (\hat{\eta}_1 - \hat{\eta}_2)^2) \\ &\quad + (\xi - \hat{\eta}_2)^2 + \lambda_{\hat{P}_1}^2(\hat{\eta}_1 - \hat{\eta}_2)^2] - \alpha(P^\lambda) \\ &= \lambda E_{\hat{P}_1}[(\xi - \hat{\eta}_1)^2] + (1 - \lambda) E_{\hat{P}_2}[(\xi - \hat{\eta}_2)^2] + \lambda E_{\hat{P}_1}[\lambda_{\hat{P}_2}^2(\hat{\eta}_1 - \hat{\eta}_2)^2] \\ &\quad + (1 - \lambda) E_{\hat{P}_2}[\lambda_{\hat{P}_1}^2(\hat{\eta}_1 - \hat{\eta}_2)^2] - \alpha(P^\lambda). \end{aligned}$$

Set  $\beta = \inf_{\eta \in L_C^{2p}(\mathbb{P})} \rho((\xi - \eta)^2)$ . By the above equation and the convexity of  $\alpha(\cdot)$ ,

$$\begin{aligned}
 (4.5) \quad & E_{P^\lambda}[(\xi - E_{P^\lambda}[\xi|\mathcal{C}])^2] - \alpha(P^\lambda) \\
 & \geq \lambda E_{\hat{P}_1}[(\xi - \hat{\eta}_1)^2] + (1 - \lambda)E_{\hat{P}_2}[(\xi - \hat{\eta}_2)^2] - [\lambda\alpha(\hat{P}_1) + (1 - \lambda)\alpha(\hat{P}_2)] \\
 & \quad + \lambda E_{\hat{P}_1}[\lambda_{\hat{P}_2}^2(\hat{\eta}_1 - \hat{\eta}_2)^2] + (1 - \lambda)E_{\hat{P}_2}[\lambda_{\hat{P}_1}^2(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
 & = \beta + \lambda E_{\hat{P}_1}[\lambda_{\hat{P}_2}^2(\hat{\eta}_1 - \hat{\eta}_2)^2] + (1 - \lambda)E_{\hat{P}_2}[\lambda_{\hat{P}_1}^2(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
 & \geq \beta.
 \end{aligned}$$

On the other hand, since  $(\hat{\eta}_1, \hat{P}_1)$  is a saddle point, we have

$$E_{P^\lambda}[(\xi - E_{P^\lambda}[\xi|\mathcal{C}])^2] - \alpha(P^\lambda) \leq E_{P^\lambda}[(\xi - \hat{\eta}_1)^2] - \alpha(P^\lambda) \leq E_{\hat{P}_1}[(\xi - \hat{\eta}_1)^2] - \alpha(\hat{P}_1) = \beta.$$

It yields that  $E_{P^\lambda}[(\xi - E_{P^\lambda}[\xi|\mathcal{C}])^2] - \alpha(P^\lambda) = \beta$ . By (4.5), we deduce that  $\hat{\eta}_1 = \hat{\eta}_2$   $\mathbb{P}$ -a.s. □

Finally, we will list some properties of the MMSE  $\rho(\xi|\mathcal{C})$ .

PROPOSITION 4.20. *If Assumption 4.9 holds, then for any  $\xi \in L_{\mathcal{F}}^{4p}(\mathbb{P})$ , we have*

- (i) *if  $C_1 \leq \xi(\omega) \leq C_2$  for two constants  $C_1$  and  $C_2$ , then  $C_1 \leq \rho(\xi|\mathcal{C}) \leq C_2$ ;*
- (ii)  $\rho(-\xi|\mathcal{C}) = -\rho(\xi|\mathcal{C})$ ;
- (iii) *For any given  $\eta_0 \in L_C^{2p}(\mathbb{P})$ , we have  $\rho(\xi + \eta_0|\mathcal{C}) = \rho(\xi|\mathcal{C}) + \eta_0$ ;*
- (iv) *If  $\xi$  is independent of the sub  $\sigma$ -algebra  $\mathcal{C}$  under every probability measure  $P \in \mathcal{P}$ , then  $\rho(\xi|\mathcal{C})$  is a constant.*

*Proof.* (i) If  $C_1 \leq \xi(\omega) \leq C_2$ , then for any  $P \in \mathcal{P}$ ,  $C_1 \leq E_P[\xi|\mathcal{C}] \leq C_2$ . According to the proof of Proposition 4.18,  $\rho(\xi|\mathcal{C}) \in \{E_P[\xi|\mathcal{C}]; P \in \mathcal{P}\}$  which leads to  $C_1 \leq \rho(\xi|\mathcal{C}) \leq C_2$ .

(ii) Since

$$\rho((\xi - \rho(\xi|\mathcal{C}))^2) = \inf_{\eta \in L_C^{2p}(\mathbb{P})} \rho((\xi - \eta)^2) = \inf_{\eta \in L_C^{2p}(\mathbb{P})} \rho((\xi + \eta)^2) = \inf_{\eta \in L_C^{2p}(\mathbb{P})} \rho((-\xi - \eta)^2),$$

it induces that

$$\begin{aligned}
 \rho((-\xi + \rho(\xi|\mathcal{C}))^2) &= \rho((-\xi - (-\rho(\xi|\mathcal{C})))^2) \\
 &= \inf_{\eta \in L_C^{2p}(\mathbb{P})} \rho((-\xi - \eta)^2) \\
 &= \rho((-\xi - \rho(-\xi|\mathcal{C}))^2).
 \end{aligned}$$

By Theorem 4.19, we have  $-\rho(\xi|\mathcal{C}) = \rho(-\xi|\mathcal{C})$ .

(iii) Note that

$$\begin{aligned}
 \rho((\xi + \eta_0 - (\eta_0 + \rho(\xi|\mathcal{C})))^2) &= \rho((\xi - \rho(\xi|\mathcal{C}))^2) \\
 &= \inf_{\eta \in L_C^{2p}(\mathbb{P})} \rho((\xi - \eta)^2) = \inf_{\eta \in L_C^{2p}(\mathbb{P})} \rho((\xi + \eta_0 - \eta)^2).
 \end{aligned}$$

By Theorem 4.19, we have  $\eta_0 + \rho(\xi|\mathcal{C}) = \rho(\xi + \eta_0|\mathcal{C})$ .

(iv) If  $\xi$  is independent of the sub  $\sigma$ -algebra  $\mathcal{C}$  under every  $P \in \mathcal{P}$ , then  $E_P[\xi|\mathcal{C}]$  is a constant for any  $P \in \mathcal{P}$ . Since  $\rho(\xi|\mathcal{C}) \in \{E_P[\xi|\mathcal{C}]; P \in \mathcal{P}\}$ , we know that  $\rho(\xi|\mathcal{C})$  is a constant. This completes the proof. □

**5. Appendix.** For the convenience of the reader, we list the main theorems used in our proofs.

**THEOREM 5.1** (Fan [18, Theorem 2]). *Let  $\mathcal{X}$  be a compact Hausdorff space and  $\mathcal{Y}$  be an arbitrary set. Let  $F$  be a real valued function defined on  $\mathcal{X} \times \mathcal{Y}$  such that, for every  $y \in \mathcal{Y}$ ,  $F(x, y)$  is an l.s.c on  $\mathcal{X}$ . If  $F$  is convex on  $\mathcal{X}$  and concave on  $\mathcal{Y}$ , then*

$$\min_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} F(x, y) = \sup_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} F(x, y).$$

*Proof.* Refer to Theorem 2 in [18]. □

**THEOREM 5.2** (Zălinescu [43, Theorem 2.10.1]). *Let  $A$  and  $B$  be two nonempty sets and  $f$  from  $A \times B$  to  $\mathbb{R} \cup \{\infty\}$ . Then  $f$  has saddle points; i.e., there exists  $(\bar{x}, \bar{y}) \in A \times B$  such that*

$$\forall x \in A, \forall y \in B : f(x, \bar{y}) \leq f(\bar{x}, \bar{y}) \leq f(\bar{x}, y)$$

*if and only if*

$$\inf_{y \in B} f(\bar{x}, y) = \max_{x \in A} \inf_{y \in B} f(x, y) = \min_{y \in B} \sup_{x \in A} f(x, y) = \sup_{x \in A} f(x, \bar{y}).$$

**THEOREM 5.3** (Girsanov [20]). *Let  $T > 0$  be a fixed time horizon,  $\mathcal{F}_t$  be a natural filtration of the standard Brownian motion  $w(\cdot)$  up to time  $t$ , and  $\mathcal{F} = \mathcal{F}_T$ . Denote by  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leq t \leq T}, \mathbb{P})$  be a complete filtered probability space satisfying usual conditions. We suppose that  $\phi(t, \omega)$  satisfies the following conditions:*

- (1)  $\phi(\cdot, \cdot)$  are measurable in both variables.
- (2)  $\phi(t, \cdot)$  is  $\mathcal{F}_t$ -measurable for fixed  $t$ .
- (3)  $\int_0^T |\phi(t, \omega)|^2 dt < \infty$  almost everywhere, and  $0 < c_1 \leq |\phi(t, \omega)| \leq c_2$  for almost all  $(t, \omega)$ , where  $c_1$  and  $c_2$  are given constants. Then, for any given constant  $\alpha > 1$ ,  $\exp[\alpha \zeta_s^t(\phi)]$  is integrable and the following inequality holds:

$$(5.1) \quad \exp \left[ \frac{(\alpha^2 - \alpha)}{2} (t - s) c_1^2 \right] \leq \mathbb{E}[\exp[\alpha \zeta_s^t(\phi)]] \leq \exp \left[ \frac{(\alpha^2 - \alpha)}{2} (t - s) c_2^2 \right],$$

where  $\zeta_s^t(\phi) = \int_s^t \phi(u, \omega) dw(u) - \frac{1}{2} \int_s^t \phi^2(u, \omega) du$ .

**THEOREM 5.4** (Yosida [42]). *Let  $(X, \|\cdot\|)$  be a Banach space and  $\{x_n\}_{n \in \mathbb{N}}$  be a sequence in  $X$  that converges weakly to some  $x \in X$ . Then there exists, for any  $\epsilon > 0$ , a convex combination  $\sum_{j=1}^n \alpha_j x_j$ , ( $\alpha_j \geq 0$ ,  $\sum_{j=1}^n \alpha_j = 1$ ) such that  $\|x - \sum_{j=1}^n \alpha_j x_j\| \leq \epsilon$ .*

**LEMMA 5.5.** *Let  $\hat{\eta}_1 = E_{\hat{P}_1}[\xi|\mathcal{C}]$ ,  $\hat{\eta}_2 = E_{\hat{P}_2}[\xi|\mathcal{C}]$ ,  $P^\lambda = \lambda \hat{P}_1 + (1 - \lambda) \hat{P}_2$ ,  $\lambda_{\hat{P}_1} = \lambda E_{P^\lambda} \left[ \frac{d\hat{P}_1}{dP^\lambda} | \mathcal{C} \right]$ ,  $\lambda_{\hat{P}_2} = (1 - \lambda) E_{P^\lambda} \left[ \frac{d\hat{P}_2}{dP^\lambda} | \mathcal{C} \right]$ . Then we have*

$$\begin{aligned} & E_{P^\lambda} [(\xi - \lambda_{\hat{P}_1} \hat{\eta}_1 - \lambda_{\hat{P}_2} \hat{\eta}_2)^2] \\ &= \lambda E_{\hat{P}_1} [(\xi - \hat{\eta}_1)^2] + (1 - \lambda) E_{\hat{P}_2} [(\xi - \hat{\eta}_2)^2] \\ & \quad + \lambda E_{\hat{P}_1} [\lambda_{\hat{P}_2}^2 (\hat{\eta}_1 - \hat{\eta}_2)^2] + (1 - \lambda) E_{\hat{P}_2} [\lambda_{\hat{P}_1}^2 (\hat{\eta}_1 - \hat{\eta}_2)^2]. \end{aligned}$$

*Proof.*

$$\begin{aligned}
 (5.2) \quad & E_{P^\lambda}[(\xi - \lambda_{\hat{P}_1} \hat{\eta}_1 - \lambda_{\hat{P}_2} \hat{\eta}_2)^2] \\
 &= E_{P^\lambda}[(\lambda_{\hat{P}_1}(\xi - \hat{\eta}_1) + \lambda_{\hat{P}_2}(\xi - \hat{\eta}_2))^2] \\
 &= E_{P^\lambda}[\lambda_{\hat{P}_1}^2(\xi - \hat{\eta}_1)^2 + \lambda_{\hat{P}_2}^2(\xi - \hat{\eta}_2)^2 + 2\lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\xi - \hat{\eta}_1)(\xi - \hat{\eta}_2)] \\
 &= E_{P^\lambda}[\lambda_{\hat{P}_1}(\xi - \hat{\eta}_1)^2 + \lambda_{\hat{P}_2}(\xi - \hat{\eta}_2)^2 - \lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
 &= \lambda E_{\hat{P}_1}[\lambda_{\hat{P}_1}(\xi - \hat{\eta}_1)^2 + \lambda_{\hat{P}_2}(\xi - \hat{\eta}_2)^2 - \lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
 &\quad + (1 - \lambda)E_{\hat{P}_2}[\lambda_{\hat{P}_1}(\xi - \hat{\eta}_1)^2 + \lambda_{\hat{P}_2}(\xi - \hat{\eta}_2)^2 - \lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
 &= \lambda E_{\hat{P}_1}[(\xi - \hat{\eta}_1)^2] + (1 - \lambda)E_{\hat{P}_2}[(\xi - \hat{\eta}_2)^2] \\
 &\quad + \lambda E_{\hat{P}_1}[(\xi - \hat{\eta}_2)^2 - \lambda_{\hat{P}_1}(\xi - \hat{\eta}_2)^2 - \lambda_{\hat{P}_2}(\xi - \hat{\eta}_1)^2 - \lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
 &\quad + (1 - \lambda)E_{\hat{P}_2}[(\xi - \hat{\eta}_1)^2 - \lambda_{\hat{P}_2}(\xi - \hat{\eta}_1)^2 - \lambda_{\hat{P}_1}(\xi - \hat{\eta}_2)^2 - \lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2].
 \end{aligned}$$

Since

$$\begin{aligned}
 \lambda E_{\hat{P}_1}[(\xi - \hat{\eta}_2)^2] &= \lambda E_{\hat{P}_1}[(\lambda_{\hat{P}_1} + \lambda_{\hat{P}_2})(\xi - \hat{\eta}_2)^2], \\
 (1 - \lambda)E_{\hat{P}_2}[(\xi - \hat{\eta}_1)^2] &= (1 - \lambda)E_{\hat{P}_2}[(\lambda_{\hat{P}_1} + \lambda_{\hat{P}_2})(\xi - \hat{\eta}_1)^2],
 \end{aligned}$$

it results in that

$$\begin{aligned}
 (5.2) &= \lambda E_{\hat{P}_1}[(\xi - \hat{\eta}_1)^2] + (1 - \lambda)E_{\hat{P}_2}[(\xi - \hat{\eta}_2)^2] \\
 &\quad + \lambda E_{\hat{P}_1}[\lambda_{\hat{P}_2}(\xi - \hat{\eta}_2)^2 - \lambda_{\hat{P}_2}(\xi - \hat{\eta}_1)^2 - \lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2] \\
 &\quad + (1 - \lambda)E_{\hat{P}_2}[\lambda_{\hat{P}_1}(\xi - \hat{\eta}_1)^2 - \lambda_{\hat{P}_1}(\xi - \hat{\eta}_2)^2 - \lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2].
 \end{aligned}$$

Firstly, we calculate the items with respect to the expectation  $\lambda E_{\hat{P}_1}[\cdot]$ ; the following relations hold:

$$\begin{aligned}
 &\lambda_{\hat{P}_2}(\xi^2 + \hat{\eta}_2^2 - 2\xi\hat{\eta}_2) - \lambda_{\hat{P}_2}(\xi^2 + \hat{\eta}_1^2 - 2\xi\hat{\eta}_1) - \lambda_{\hat{P}_1}\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2 \\
 &= \lambda_{\hat{P}_2}(\xi^2 + \hat{\eta}_2^2 - 2\xi\hat{\eta}_2) - \lambda_{\hat{P}_2}(\xi^2 + \hat{\eta}_1^2 - 2\xi\hat{\eta}_1) - (1 - \lambda_{\hat{P}_2})\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)^2 \\
 &= \lambda_{\hat{P}_2}(-2\hat{\eta}_1^2 + 2\hat{\eta}_1\hat{\eta}_2 + 2\xi(\hat{\eta}_1 - \hat{\eta}_2)) + \lambda_{\hat{P}_2}^2(\hat{\eta}_1 - \hat{\eta}_2)^2 \\
 &= \lambda_{\hat{P}_2}[-2\hat{\eta}_1(\hat{\eta}_1 - \hat{\eta}_2) + 2\xi(\hat{\eta}_1 - \hat{\eta}_2)] + \lambda_{\hat{P}_2}^2(\hat{\eta}_1 - \hat{\eta}_2)^2 \\
 &= \lambda_{\hat{P}_2}[2(\xi - \hat{\eta}_1)(\hat{\eta}_1 - \hat{\eta}_2)] + \lambda_{\hat{P}_2}^2(\hat{\eta}_1 - \hat{\eta}_2)^2.
 \end{aligned}$$

Since  $\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)$  is  $\mathcal{C}$ -measurable and  $(\xi - \hat{\eta}_1)$  is orthogonal with  $\sigma$ -algebra  $\mathcal{C}$  under probability measure  $\hat{P}_1$ , it results that

$$\lambda E_{\hat{P}_1}[\lambda_{\hat{P}_2}2(\xi - \hat{\eta}_1)(\hat{\eta}_1 - \hat{\eta}_2)] = \lambda E_{\hat{P}_1}[\lambda_{\hat{P}_2}(\hat{\eta}_1 - \hat{\eta}_2)]E_{\hat{P}_1}[2(\xi - \hat{\eta}_1)] = 0.$$

Secondly, we can also similarly calculate the items with respect to the expectation  $(1 - \lambda)E_{\hat{P}_2}[\cdot]$ . Finally, (5.2) can be expressed as

$$\begin{aligned}
 &E_{P^\lambda}[(\xi - \lambda_{\hat{P}_1} \hat{\eta}_1 - \lambda_{\hat{P}_2} \hat{\eta}_2)^2] \\
 &= \lambda E_{\hat{P}_1}[(\xi - \hat{\eta}_1)^2] + (1 - \lambda)E_{\hat{P}_2}[(\xi - \hat{\eta}_2)^2] \\
 &\quad + \lambda E_{\hat{P}_1}[\lambda_{\hat{P}_2}^2(\hat{\eta}_1 - \hat{\eta}_2)^2] + (1 - \lambda)E_{\hat{P}_2}[\lambda_{\hat{P}_1}^2(\hat{\eta}_1 - \hat{\eta}_2)^2].
 \end{aligned}$$

This completes the proof. □

THEOREM 5.6. Let  $f_n, f, n \in \mathbb{N}$  be the real-valued measurable functions on measure space  $(\Omega, \mathcal{F}, \mu)$  such that

$$f_n \xrightarrow{\mu} f.$$

$g(\cdot)$  is a real-valued function defined on a subset  $D$  of real-space and  $\forall \omega \in \Omega, f_n(\omega) \in D, f(\omega) \in D$ . If  $\mu$  is a finite measure on  $\mathcal{F}$ ,  $D$  is a open set,  $g(\cdot)$  is a continuous function on  $D$ ; then  $g(f_n) \rightarrow_{\mu} g(f)$ .

## REFERENCES

- [1] A. ALLAN AND S. COHEN, *Parameter uncertainty in the Kalman–Bucy filter*, SIAM J. Control Optim., 57 (2019), pp. 1646–1671.
- [2] B. D. O. ANDERSON AND J. B. MOORE, *Optimal Control: Linear Quadratic Methods*, Prentice Hall, Englewood Cliffs, NJ, 1990.
- [3] T. ARAI AND M. FUKASAWA, *Convex risk measure for good deal bounds*, Math. Finance, 24 (2014), pp. 464–484.
- [4] K. J. ÅSTRÖM, *Introduction to Stochastic Control Theory*, Math. Sci. Engrg. 70, Academic Press, New York, 1970.
- [5] R. J. BARTON AND H. V. POOR, *An RKHS approach to robust  $L^2$  estimation and signal detection*, IEEE Trans. Automat. Control, 36 (1990), pp. 485–501.
- [6] A. BENSOUSSAN, *Stochastic Control of Partially Observable Systems*, Cambridge University Press, Cambridge, 2004.
- [7] A. V. BORISOV AND A. R. PANKOV, *Minimax filtering in uncertain-stochastic systems described by stochastic differential equation with a measure*, IFAC Proceedings Volumes, 30 (1994), pp. 175–178.
- [8] A. V. BORISOV, *Minimax a posteriori estimation of the Markov processes with finite state spaces*, Autom. Remote Control, 69 (2008), pp. 233–246.
- [9] A. V. BORISOV, *The Wonham filter under uncertainty: A game-theoretic approach*, Automatica, 47 (2011), pp. 1015–1019.
- [10] Z. CHEN AND L. EPSTEIN, *Ambiguity, risk, and asset returns in continuous time*, Econometrica, 70 (2002), pp. 1403–1443.
- [11] G. CHEN AND Y. SHEN, *Robust  $H_\infty$  filter design for neutral stochastic uncertain systems with time-varying delay*, J. Math. Anal. Appl., 353 (2009), pp. 196–204.
- [12] W. CHE AND G. YANG,  *$H_\infty$  filter design for continuous-time systems with quantised signals*, Int. J. Systems Sci., 44 (2013), pp. 265–274.
- [13] M. H. A. DAVIS, *Linear Estimation and Stochastic Control*, Chapman and Hall, London, 1977.
- [14] F. DELBAEN, S. PENG, AND E. R. GIANIN, *Representation of the penalty term of dynamic concave utilities*, Finance Stoch., 14 (2010), pp. 449–472.
- [15] N. EL KAROUI, S. PENG, AND M. QUENEZ, *Backward stochastic differential equations in finance*, Math. Finance, 7 (1997), pp. 1–71.
- [16] L. EPSTEIN AND S. JI, *Ambiguous volatility, possibility and utility in continuous time*, J. Math. Economics, 50 (2014), pp. 269–282.
- [17] L. EPSTEIN AND S. JI, *Ambiguous volatility and asset pricing in continuous time*, Rev. Finan. Stud., 26 (2013) pp. 1740–1786.
- [18] K. FAN, *Minimax theorems*, Proc. Nat. Acad. Sci. U.S.A., 39 (1953), pp. 42–47.
- [19] H. FÖLLMER AND A. SCHIED, *Stochastic Finance, An Introduction in Discrete Time*, Walter de Gruyter, Berlin, 2002.
- [20] I. V. GIRSANOV, *On transforming a certain class of stochastic processes by absolutely continuous substitution of measures*, Theory. Probab. Appl., 5 (1960), pp. 285–301.
- [21] L. GUO, *Estimation, control, and games of dynamical systems with uncertainty*, SCIENCE CHINA Information Sciences, 50 (2020).
- [22] P. HUANG, G. WANG, AND H. ZHANG, *A partial information linear-quadratic optimal control problem of backward stochastic differential equation with its applications*, Sci. China Inf. Sci., 63 (2020), 192204.
- [23] A. H. JAZWINSKI, *Stochastic Processes and Filtering Theory*, Academic Press, Cambridge, MA, 1970.
- [24] S. JI, C. KONG, AND C. SUN, *A filtering problem with uncertainty in observation*, System Control Lett., 135 (2020), 104589.
- [25] S. JI, C. KONG, AND C. SUN, *A robust Kalman–Bucy filtering problem*, Automatica, 122 (2020), 109252, <https://doi.org/10.1016/j.automatica.2020.109252>.

- [26] S. JI, C. KONG, AND C. SUN, *The minimum mean square estimator of integrable variables under sublinear operators*, *Stochastics*, 92 (2020), pp. 519–532.
- [27] R. E. KALMAN AND R. S. BUCY, *New results in linear filtering and prediction theory*, *J. Basic Engrg.*, 83 (1961), pp. 95–108.
- [28] M. KAINA AND L. RÜSCHENDORF, *On convex risk measures on  $L_p$ -spaces*, *Math. Methods Oper. Res.*, 69 (2009), pp. 475–495.
- [29] I. KARATZAS AND S. SHREVE, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, 2002.
- [30] P. LAKNER, *Utility maximization with partial information*, *Stoch. Process. Appl.*, 56 (1995), pp. 247–273.
- [31] R. S. LIPTSER AND A. N. SHIRYAEV, *Statistics of Random Processes: I. General Theory*, Springer Science and Business Media, New York, 2013.
- [32] K. M. NAGPAL AND P. P. KHARGONEKAR, *Filtering and smoothing in an  $H_\infty$  setting*, *IEEE Trans. Automat. Control*, 36 (1991), pp. 152–166.
- [33] A. R. PANKOV AND A. V. BORISOV, *Optimal Estimation in Uncertain Stochastic Discrete-Time Systems*, Technical Report 91-27, University of Michigan, Ann Arbor, MI, 1991.
- [34] A. R. PANKOV AND A. V. BORISOV, *A solution of the filtering and smoothing problems for uncertain-stochastic linear dynamic systems*, *Int. J. Control*, 60 (1994), pp. 413–423.
- [35] S. PENG, *BSDE and related  $g$ -expectations*, in *Backward Stochastic Differential Equations*, N. El Karoui and L. Mazliak, eds., Pitman Research Notes in Mathematics Series 364, Longman, Harlow, 1997, pp. 141–159.
- [36] H. PHAM, *Continuous-Time Stochastic Control and Optimization with Financial Applications*, Springer Science and Business Media, New York, 2009.
- [37] S. SIMONS, *From Hahn-Banach to Monotonicity*, Springer, New York, 2008.
- [38] C. SUN AND S. JI, *The least squares estimator of random variables under sublinear expectations*, *J. Math. Anal. Appl.*, 451 (2017), pp. 906–923.
- [39] J. XIONG, *An Introduction to Stochastic Filtering Theory*, Oxford University Press Oxford, 2008.
- [40] J. XIONG AND X. ZHOU, *Mean-variance portfolio selection under partial information*, *SIAM J. Control. Optim.*, 46 (2007), pp. 156–175.
- [41] J. YONG AND X. ZHOU, *Stochastic Controls: Hamiltonian Systems and HJB Equations*, Springer Science and Business Media, New York, 1999.
- [42] K. YOSIDA, *Functional Analysis*, Springer, Berlin, 1980.
- [43] C. ZĂLINESCU, *Convex Analysis in General Vector Spaces*, World Scientific, River Edge, NJ, 2002.