



ACADÉMIE
DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus

Mathématique


Huijie Qiao and Shengqing Zhu

A limit theorem of nonlinear filtering for multiscale McKean–Vlasov stochastic systems

Volume 362 (2024), p. 1287-1299

Online since: 14 November 2024

<https://doi.org/10.5802/crmath.637>

 This article is licensed under the
CREATIVE COMMONS ATTRIBUTION 4.0 INTERNATIONAL LICENSE.
<http://creativecommons.org/licenses/by/4.0/>



*The Comptes Rendus. Mathématique are a member of the
Mersenne Center for open scientific publishing*
www.centre-mersenne.org — e-ISSN : 1778-3569



Research article / *Article de recherche*
Probability theory / *Probabilités*

A limit theorem of nonlinear filtering for multiscale McKean–Vlasov stochastic systems

Théorème limite du filtrage non linéaire pour les systèmes stochastiques McKean–Vlasov à plusieurs échelles

Huijie Qiao ^{*,a} and Shengqing Zhu ^a

^a School of Mathematics, Southeast University, Nanjing, Jiangsu 211189, P.R.China
E-mails: hjqiaogean@seu.edu.cn, 220211787@seu.edu.cn (Shengqing Zhu)

Abstract. The work concerns about multiscale McKean–Vlasov stochastic systems. First of all, we prove an average principle for these systems in the L^2 sense. Moreover, a convergence rate is presented. Then we define the nonlinear filtering of these systems and establish a limit theorem about nonlinear filtering of them in the L^2 sense.

Résumé. Ce travail concerne les systèmes stochastiques McKean–Vlasov multi-échelles. Tout d’abord, nous prouvons un principe de moyenne pour ces systèmes au sens L^2 . De plus, un taux de convergence est présenté. Ensuite, nous définissons le filtrage non linéaire de ces systèmes et établissons un théorème limite sur le filtrage non linéaire de ces systèmes au sens L^2 .

Keywords. Multiscale McKean–Vlasov stochastic systems, average principle, nonlinear filtering, limit theorem.

2020 Mathematics Subject Classification. 60G35.

Funding. This work was partly supported by NSF of China (No.12071071).

Manuscript received 26 November 2023, accepted 31 March 2024.

1. Introduction

In this paper, we study the following McKean–Vlasov stochastic system $(X_t^\varepsilon, Y_t^\varepsilon)$ on $\mathbb{R}^n \times \mathbb{R}^m$: for any $T > 0$

$$\begin{cases} dX_t^\varepsilon = \frac{1}{\varepsilon} b(X_t^\varepsilon) dt + \frac{1}{\sqrt{\varepsilon}} \sigma(X_t^\varepsilon) dB_t, & 0 \leq t \leq T, \\ X_0^\varepsilon = \xi, \\ dY_t^\varepsilon = h(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) dt + dW_t, & 0 \leq t \leq T, \\ Y_0^\varepsilon = 0, \end{cases} \quad (1)$$

*Corresponding author

where $\{B_t, t \geq 0\}, \{W_t, t \geq 0\}$ are d - and m -dimensional standard Brownian motions defined on a complete filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, T]}, \mathbb{Q})$, respectively, and mutually independent. Here these coefficients $b : \mathbb{R}^n \mapsto \mathbb{R}^n$, $\sigma : \mathbb{R}^n \mapsto \mathbb{R}^{n \times d}$ and $h : \mathbb{R}^n \times \mathcal{P}_2(\mathbb{R}^n) \mapsto \mathbb{R}^m$ are Borel measurable, ξ is a \mathcal{F}_0 -measurable random variable with $\mathbb{E}|\xi|^2 < \infty$ and independent of B . and W ., and $\mathcal{L}_{X_t^\varepsilon}^\mathbb{Q}$ denotes the distribution of X_t^ε under the probability measure \mathbb{Q} (See Section 2 for the definition of $\mathcal{P}_2(\mathbb{R}^n)$). ε is a small parameter which characterizes the ratio of timescales between processes X^ε and Y^ε .

The system (1) is usually called a multiscale or fast-slow McKean–Vlasov stochastic system, where $X^\varepsilon, Y^\varepsilon$ are fast and slow parts, respectively. These systems are widely used in engineering and science fields. For these systems, the average principle is an effective tool in capturing the main behaviors of the slow components while avoiding the complexities caused by the details of the fast components of them. If $h(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^\mathbb{Q})$ is independent of $\mathcal{L}_{X_t^\varepsilon}^\mathbb{Q}$, that is, the system (1) does not depend on the distribution, about average principles of the system (1) and more general systems there have been many related results (cf. [3, 4, 6]). If $h(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^\mathbb{Q})$ is replaced by $\tilde{h}(Y_t^\varepsilon, \mathcal{L}_{Y_t^\varepsilon}^\mathbb{Q}, X_t^\varepsilon)$, where $\tilde{h} : \mathbb{R}^m \times \mathcal{P}_2(\mathbb{R}^m) \times \mathbb{R}^n \mapsto \mathbb{R}^m$ is Borel measurable, there also have been a lot of average principle results on the system (1) and more general systems (cf. [1, 8, 10–12, 16]). However, for the system (1), as far as we know, there are few related average principle results. Therefore, in this paper, we establish an average principle for the system (1). Concretely speaking, we prove that for $0 < \gamma < 1$

$$\mathbb{E} \sup_{t \in [0, T]} |Y_t^\varepsilon - \bar{Y}_t|^2 \leq C(\varepsilon + \varepsilon^{1-\gamma} + \varepsilon^\gamma),$$

where

$$\bar{Y}_t := \bar{h}t + W_t \tag{2}$$

and $\bar{h} := \int_{\mathbb{R}^n} h(x, \nu) \nu(dx)$ (See Subsection 3.1 for the definition of ν).

Note that in [2], Gao, Hong and Liu considered the following McKean–Vlasov stochastic system

$$\begin{cases} dX_t^{\varepsilon, \delta} = \frac{1}{\varepsilon} b(X_t^{\varepsilon, \delta}) dt + \frac{1}{\sqrt{\varepsilon}} \sigma(X_t^{\varepsilon, \delta}) dB_t, & 0 \leq t \leq T, \\ X_0^{\varepsilon, \delta} = \xi, \\ dY_t^{\varepsilon, \delta} = \hat{h}(Y_t^{\varepsilon, \delta}, \mathcal{L}_{Y_t^{\varepsilon, \delta}}^\mathbb{Q}, X_t^{\varepsilon, \delta}, \mathcal{L}_{X_t^{\varepsilon, \delta}}^\mathbb{Q}) dt + \sqrt{\delta} dW_t, & 0 \leq t \leq T, \\ Y_0^\varepsilon = 0, \end{cases} \tag{3}$$

where $\hat{h} : \mathbb{R}^m \times \mathcal{P}_2(\mathbb{R}^m) \times \mathbb{R}^n \times \mathcal{P}_2(\mathbb{R}^n) \mapsto \mathbb{R}^m$ is Borel measurable and $\delta > 0$ is a small parameter. There they proved that

$$\mathbb{E} \sup_{t \in [0, T]} |Y_t^{\varepsilon, \delta} - \bar{Y}_t^0|^2 \leq C(1 + \mathbb{E}|\xi|^2)(\varepsilon^{1/3} + \delta), \tag{4}$$

where $C > 0$ is a constant independent of ε, δ , \bar{Y} solves the following ordinary differential equation

$$\begin{cases} d\bar{Y}_t^0 = \bar{h}(\bar{Y}_t^0, \mathcal{L}_{\bar{Y}_t^0}) dt, \\ \bar{Y}_0^0 = 0, \end{cases}$$

and $\bar{h}(y, \mu) := \int_{\mathbb{R}^n} \hat{h}(y, \mu, x, \nu) \nu(dx)$. Comparing the system (3) with the system (1), we find that the former is more general. But if $\delta = 1$, by (4), we only obtain that

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \sup_{t \in [0, T]} |Y_t^{\varepsilon, 1} - \bar{Y}_t|^2 \leq C(1 + \mathbb{E}|\xi|^2).$$

That is, the result in [2] does not cover our result (Theorem 2).

Next, we define the nonlinear filtering of $(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^\mathbb{Q})$ with respect to $\{Y_s^\varepsilon, 0 \leq s \leq t\}$ as follows:

$$\pi_t^\varepsilon(F) := \mathbb{E} \left[F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^\mathbb{Q}) \mid \mathcal{F}_t^{Y^\varepsilon} \right], \quad F \in \mathcal{B}_b(\mathbb{R}^n \times \mathcal{P}_2(\mathbb{R}^n)),$$

where $\mathcal{F}_t^{Y^\varepsilon}$ is the \mathbb{Q} -augmentation σ -algebra generated by $\{Y_s^\varepsilon, 0 \leq s \leq t\}$. And $\pi_t^\varepsilon(F)$ satisfies a space-distribution dependent Kushner–Stratonovich equation (cf. [5, 13, 14]). Here we show that $\pi_t^\varepsilon(F)$ converges to $\bar{\pi}(F)$ in the L^2 sense (See Subsection 3.2 for the definition of $\bar{\pi}(F)$). Note that if h, F don't depend on the distribution, our result (Theorem 4) is the same to [7, Theorem 3.1(ii)]. Therefore, our result is more general.

The novelty of this work lies in two folds. The first fold is that we prove an average principle for the system (1) in the L^2 sense. Moreover, a convergence rate is presented, which is important for numerical simulation. The second fold is that we establish a limit theorem about nonlinear filtering of the system (1) in the L^2 sense. Comparing Theorem 4 with some known results (cf. [10, 11]), we find that our result concerns about the nonlinear filtering of the whole signal process and better characterizes the asymptotic behavior of the whole system.

This paper is arranged as follows. In the next section, we introduce some notations used in the sequel. Then the main results are stated in Section 3. The proofs of two main theorems are placed in Section 4 and 5, respectively. In Section 6, we give an example to explain our results.

The following convention will be used throughout the paper: C , with or without indices, will denote different positive constants whose values may change from one place to another.

2. Preliminaries

In this section, we introduce some notations used in the sequel.

Let $|\cdot|$ and $\|\cdot\|$ be norms of vectors and matrices, respectively. Let A^* be the transpose of the matrix A .

Let $\mathcal{B}_b(\mathbb{R}^n)$ be the set of bounded Borel measurable functions on \mathbb{R}^n . Let $C(\mathbb{R}^n)$ be the space of continuous functions on \mathbb{R}^n , and $C_b(\mathbb{R}^n)$ be the collection of functions in $C(\mathbb{R}^n)$ which are bounded.

Let $\mathcal{B}(\mathbb{R}^n)$ be the Borel σ -field on \mathbb{R}^n . Let $\mathcal{P}(\mathbb{R}^n)$ be the space of probability measures defined on $\mathcal{B}(\mathbb{R}^n)$ and $\mathcal{P}_2(\mathbb{R}^n)$ be the collection of the probability measures μ on $\mathcal{B}(\mathbb{R}^n)$ satisfying

$$\mu(|\cdot|^2) := \int_{\mathbb{R}^n} |x|^2 \mu(dx) < \infty.$$

We put on $\mathcal{P}_2(\mathbb{R}^n)$ a topology induced by the following 2-Wasserstein metric:

$$\mathbb{W}_2^2(\mu_1, \mu_2) := \inf_{\pi \in \mathcal{C}(\mu_1, \mu_2)} \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2 \pi(dx, dy), \quad \mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^n),$$

where $\mathcal{C}(\mu_1, \mu_2)$ denotes the set of the probability measures whose marginal distributions are μ_1, μ_2 , respectively. Moreover, if ξ, ζ are two random variables with distributions $\mathcal{L}_\xi^{\mathbb{Q}}, \mathcal{L}_\zeta^{\mathbb{Q}}$ under \mathbb{Q} , respectively,

$$\mathbb{W}_2^2(\mathcal{L}_\xi^{\mathbb{Q}}, \mathcal{L}_\zeta^{\mathbb{Q}}) \leq \mathbb{E}|\xi - \zeta|^2,$$

where \mathbb{E} stands for the expectation with respect to \mathbb{Q} .

3. Main results

In this section, we formulate the main results in this paper.

3.1. The average principle for multiscale McKean–Vlasov stochastic systems

In this subsection, we present an average principle for multiscale McKean–Vlasov stochastic systems.

Assume:

$(\mathbf{H}_{b,\sigma}^1)$ There is a constant $L_1 > 0$ such that for $x_1, x_2 \in \mathbb{R}^n$,

$$|b(x_1) - b(x_2)|^2 + \|\sigma(x_1) - \sigma(x_2)\|^2 \leq L_1|x_1 - x_2|^2.$$

(\mathbf{H}_h) h is bounded, and there is a constant $L_2 > 0$ such that for any $x_i \in \mathbb{R}^n, \mu_i \in \mathcal{P}_2(\mathbb{R}^n), i = 1, 2$

$$|h(x_1, \mu_1) - h(x_2, \mu_2)|^2 \leq L_2(|x_1 - x_2|^2 + \mathbb{W}_2^2(\mu_1, \mu_2)).$$

$(\mathbf{H}_{b,\sigma}^2)$ There exists a constant $\beta > 0$ satisfying $\beta > 2L_1$ such that for $x_i \in \mathbb{R}^n, i = 1, 2$,

$$2\langle x_1 - x_2, b(x_1) - b(x_2) \rangle + \|\sigma(x_1) - \sigma(x_2)\|^2 \leq -\beta|x_1 - x_2|^2.$$

Remark 1. By $(\mathbf{H}_{b,\sigma}^1)$ and $(\mathbf{H}_{b,\sigma}^2)$, it holds that for $x \in \mathbb{R}^n$,

$$2\langle x, b(x) \rangle + \|\sigma(x)\|^2 \leq -\alpha|x|^2 + C, \tag{5}$$

where $\alpha := \beta - 2L_1$ and $C > 0$ is a constant.

Under $(\mathbf{H}_{b,\sigma}^1)$ (\mathbf{H}_h) , the system (1) has a unique strong solution $(X^\varepsilon, Y^\varepsilon)$. Then consider the following stochastic differential equation:

$$\begin{cases} d\widehat{X}_t = b(\widehat{X}_t)dt + \sigma(\widehat{X}_t)d\widehat{B}_t, & 0 \leq t \leq T, \\ \widehat{X}_0 = \widehat{\xi}, \end{cases} \tag{6}$$

where \widehat{B} is a d -dimensional standard Brownian motion defined on the other complete filtered probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \{\widehat{\mathcal{F}}_t\}_{t \in [0, T]}, \widehat{\mathbb{Q}})$ and $\widehat{\xi}$ is a $\widehat{\mathcal{F}}_0$ -measurable random variable independent of \widehat{B} . with $\mathcal{L}_{\widehat{\xi}}^{\widehat{\mathbb{Q}}} = \mathcal{L}_{\widehat{\xi}}^{\mathbb{Q}}$. Under $(\mathbf{H}_{b,\sigma}^1)$, (6) has a unique strong solution $\widehat{X}^{\widehat{\xi}}$. Moreover, under $(\mathbf{H}_{b,\sigma}^2)$, by [15, Theorem 3.1], one could obtain that there exists a unique invariant probability measure $\nu \in \mathcal{P}_2(\mathbb{R}^n)$ for (6). So, for \bar{Y} defined in (2), we have the following theorem which is the first main result.

Theorem 2. Suppose that $(\mathbf{H}_{b,\sigma}^1)$, (\mathbf{H}_h) , $(\mathbf{H}_{b,\sigma}^2)$ hold. Then it holds that for $0 < \gamma < 1$

$$\mathbb{E} \sup_{t \in [0, T]} |Y_t^\varepsilon - \bar{Y}_t|^2 \leq C(\varepsilon + \varepsilon^{1-\gamma} + \varepsilon^\gamma). \tag{7}$$

The proof of the above theorem is placed in Section 4.

Remark 3. Let $\gamma = \frac{1}{2}$, and we obtain that

$$\left(\mathbb{E} \sup_{t \in [0, T]} |Y_t^\varepsilon - \bar{Y}_t|^2 \right)^{1/2} \leq C\varepsilon^{1/4}.$$

That is, the convergence rate is $\frac{1}{4}$.

3.2. A limit theorem of nonlinear filtering for multiscale McKean–Vlasov stochastic systems

In this subsection, we define nonlinear filtering for multiscale McKean–Vlasov stochastic systems and average systems, and state a limit theorem.

Set

$$(\Lambda_t^\varepsilon)^{-1} := \exp \left\{ -\int_0^t h^i(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^{\mathbb{Q}}) dW_s^i - \frac{1}{2} \int_0^t |h(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^{\mathbb{Q}})|^2 ds \right\}.$$

Here and hereafter, we use the convention that repeated indices imply summation. Then by (\mathbf{H}_h) , we know that $(\Lambda_t^\varepsilon)^{-1}$ is an exponential martingale. Define a probability measure $\widetilde{\mathbb{Q}}^\varepsilon$ via

$$d\widetilde{\mathbb{Q}}^\varepsilon = (\Lambda_T^\varepsilon)^{-1} d\mathbb{Q},$$

and under the probability measure $\tilde{\mathbb{Q}}^\varepsilon$,

$$Y_t^\varepsilon = W_t + \int_0^t h(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^\mathbb{Q}) ds$$

is an (\mathcal{F}_t) -adapted Brownian motion.

Set for any $F \in \mathcal{B}_b(\mathbb{R}^n \times \mathcal{P}_2(\mathbb{R}^n))$

$$\pi_t^\varepsilon(F) := \mathbb{E} \left[F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^\mathbb{Q}) \mid \mathcal{F}_t^{Y^\varepsilon} \right],$$

where $\mathcal{F}_t^{Y^\varepsilon}$ is the \mathbb{Q} -augmentation σ -algebra generated by $\{Y_s^\varepsilon, 0 \leq s \leq t\}$, and the Kallianpur-Striebel formula gives

$$\pi_t^\varepsilon(F) = \frac{\tilde{\mathbb{P}}_t^\varepsilon(F)}{\tilde{\mathbb{P}}_t^\varepsilon(1)},$$

where

$$\tilde{\mathbb{P}}_t^\varepsilon(F) := \tilde{\mathbb{E}} \left[F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^\mathbb{Q}) \Lambda_t^\varepsilon \mid \mathcal{F}_t^{Y^\varepsilon} \right],$$

and $\tilde{\mathbb{E}}$ denotes the expectation under the probability measure $\tilde{\mathbb{Q}}^\varepsilon$. Here $\tilde{\mathbb{P}}_t^\varepsilon(F)$, $\pi_t^\varepsilon(F)$ are called the unnormalized filtering and the normalized filtering of $(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^\mathbb{Q})$ with respect to $\mathcal{F}_t^{Y^\varepsilon}$, respectively.

Also set for any $F \in \mathcal{B}_b(\mathbb{R}^n \times \mathcal{P}_2(\mathbb{R}^n))$

$$\begin{aligned} \bar{\Lambda}_t &:= \exp \left\{ \bar{h}^t \bar{Y}_t^i - \frac{1}{2} |\bar{h}|^2 t \right\}, & \bar{F} &:= \int_{\mathbb{R}^n} F(x, \nu) \nu(dx), \\ \bar{\mathbb{P}}_t(\bar{F}) &:= \tilde{\mathbb{E}}[\bar{F} \bar{\Lambda}_t \mid \mathcal{F}_t^{\bar{Y}}], & \bar{\pi}_t(\bar{F}) &:= \frac{\bar{\mathbb{P}}_t(\bar{F})}{\bar{\mathbb{P}}_t(1)}. \end{aligned}$$

Then the relationship between π^ε and $\bar{\pi}$ is presented in the following theorem which is the main result in this subsection.

Theorem 4. Assume that $(\mathbf{H}_{b,\sigma}^1)$, (\mathbf{H}_h) , $(\mathbf{H}_{b,\sigma}^2)$ hold. Then we have that for any $t \in [0, T]$

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left| \pi_t^\varepsilon(F) - \bar{\pi}_t(\bar{F}) \right|^2 = 0, \quad F \in \mathcal{A}_\eta,$$

where for $\eta > 0$ \mathcal{A}_η is defined as follows:

$$\mathcal{A}_\eta := \left\{ F \in C(\mathbb{R}^n \times \mathcal{P}_2(\mathbb{R}^n)) : \sup_{\varepsilon \in (0,1)} \sup_{t \in [0,T]} \mathbb{E} \left| F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^\mathbb{Q}) \right|^{2+\eta} < \infty, \quad \text{and} \quad \int_{\mathbb{R}^n} |F(x, \nu)|^2 \nu(dx) < \infty \right\}.$$

The proof of the above theorem is placed in Section 5.

Remark 5. Theorem 4 can be roughly understood as follows: As $\varepsilon \rightarrow 0$, $\mathbb{E}[F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^\mathbb{Q}) \mid \mathcal{F}_t^{Y^\varepsilon}]$ converges to “ $\mathbb{E}[\bar{F} \mid \mathcal{F}_t^{\bar{Y}}]$ ” in the L^2 sense. This is also reasonable.

Remark 6. It is easy to see that $C_b(\mathbb{R}^n \times \mathcal{P}_2(\mathbb{R}^n)) \subset \mathcal{A}_\eta$.

Remark 7. We mention that if both h and the test function F are independent of μ , Theorem 4 is [7, Theorem 3.1(ii)]. Therefore, our result is more general.

4. Proof of Theorem 2

In this section, we prove Theorem 2. And we begin with some key estimates.

Lemma 8. Under $(\mathbf{H}_{b,\sigma}^1)$ and $(\mathbf{H}_{b,\sigma}^2)$, it holds that for any $t \geq 0$

$$\widehat{\mathbb{E}}|\widehat{X}_t^{\hat{\xi}}|^2 \leq \widehat{\mathbb{E}}|\widehat{\xi}|^2 e^{-\alpha t} + \frac{C}{\alpha}, \tag{8}$$

$$\widehat{\mathbb{E}}|\widehat{X}_t^{\hat{\xi}_1} - \widehat{X}_t^{\hat{\xi}_2}|^2 \leq \widehat{\mathbb{E}}|\widehat{\xi}_1 - \widehat{\xi}_2|^2 e^{-\beta t}. \tag{9}$$

Proof. By applying the Itô formula to $|\widehat{X}_t^{\widehat{\xi}}|^2 e^{\lambda_1 t}$ with $\lambda_1 = \alpha$, it holds that

$$|\widehat{X}_t^{\widehat{\xi}}|^2 e^{\lambda_1 t} = |\widehat{\xi}|^2 + \lambda_1 \int_0^t e^{\lambda_1 s} |\widehat{X}_s^{\widehat{\xi}}|^2 ds + 2 \int_0^t e^{\lambda_1 s} \langle \widehat{X}_s^{\widehat{\xi}}, b(\widehat{X}_s^{\widehat{\xi}}) \rangle ds + 2 \int_0^t e^{\lambda_1 s} \langle \widehat{X}_s^{\widehat{\xi}}, \sigma(\widehat{X}_s^{\widehat{\xi}}) d\widehat{B}_s \rangle + \int_0^t e^{\lambda_1 s} \|\sigma(\widehat{X}_s^{\widehat{\xi}})\|^2 ds.$$

Taking the expectation on two sides, by (5), we have that

$$\begin{aligned} \widehat{\mathbb{E}}|\widehat{X}_t^{\widehat{\xi}}|^2 e^{\lambda_1 t} &\leq \widehat{\mathbb{E}}|\widehat{\xi}|^2 + \lambda_1 \int_0^t e^{\lambda_1 s} \widehat{\mathbb{E}}|\widehat{X}_s^{\widehat{\xi}}|^2 ds + \int_0^t e^{\lambda_1 s} (-\alpha \widehat{\mathbb{E}}|\widehat{X}_s^{\widehat{\xi}}|^2 + C) ds \\ &\leq \widehat{\mathbb{E}}|\widehat{\xi}|^2 + C \int_0^t e^{\lambda_1 s} ds, \end{aligned}$$

and

$$\widehat{\mathbb{E}}|\widehat{X}_t^{\widehat{\xi}}|^2 \leq \widehat{\mathbb{E}}|\widehat{\xi}|^2 e^{-\alpha t} + \frac{C}{\alpha},$$

which completes the proof of (8).

Next, assume that $\widehat{X}_t^{\widehat{\xi}_1}, \widehat{X}_t^{\widehat{\xi}_2}$ are solutions of (6) with the initial values $\widehat{\xi}_1, \widehat{\xi}_2$, respectively. Then it holds that

$$\widehat{X}_t^{\widehat{\xi}_1} - \widehat{X}_t^{\widehat{\xi}_2} = \widehat{\xi}_1 - \widehat{\xi}_2 + \int_0^t (b(\widehat{X}_s^{\widehat{\xi}_1}) - b(\widehat{X}_s^{\widehat{\xi}_2})) ds + \int_0^t (\sigma(\widehat{X}_s^{\widehat{\xi}_1}) - \sigma(\widehat{X}_s^{\widehat{\xi}_2})) d\widehat{B}_s.$$

Applying the Itô formula to $|\widehat{X}_t^{\widehat{\xi}_1} - \widehat{X}_t^{\widehat{\xi}_2}|^2 e^{\lambda_2 t}$ for $\lambda_2 = \beta$ and taking the expectation on two sides, we obtain that

$$\begin{aligned} \widehat{\mathbb{E}}|\widehat{X}_t^{\widehat{\xi}_1} - \widehat{X}_t^{\widehat{\xi}_2}|^2 e^{\lambda_2 t} &= \widehat{\mathbb{E}}|\widehat{\xi}_1 - \widehat{\xi}_2|^2 + \lambda_2 \widehat{\mathbb{E}} \int_0^t e^{\lambda_2 s} |\widehat{X}_s^{\widehat{\xi}_1} - \widehat{X}_s^{\widehat{\xi}_2}|^2 ds \\ &\quad + 2 \widehat{\mathbb{E}} \int_0^t e^{\lambda_2 s} \langle \widehat{X}_s^{\widehat{\xi}_1} - \widehat{X}_s^{\widehat{\xi}_2}, b(\widehat{X}_s^{\widehat{\xi}_1}) - b(\widehat{X}_s^{\widehat{\xi}_2}) \rangle ds + \widehat{\mathbb{E}} \int_0^t e^{\lambda_2 s} \|\sigma(\widehat{X}_s^{\widehat{\xi}_1}) - \sigma(\widehat{X}_s^{\widehat{\xi}_2})\|^2 ds \\ &\leq \widehat{\mathbb{E}}|\widehat{\xi}_1 - \widehat{\xi}_2|^2 + (\lambda_2 - \beta) \widehat{\mathbb{E}} \int_0^t e^{\lambda_2 s} |\widehat{X}_s^{\widehat{\xi}_1} - \widehat{X}_s^{\widehat{\xi}_2}|^2 ds, \end{aligned}$$

which implies that

$$\widehat{\mathbb{E}}|\widehat{X}_t^{\widehat{\xi}_1} - \widehat{X}_t^{\widehat{\xi}_2}|^2 \leq \widehat{\mathbb{E}}|\widehat{\xi}_1 - \widehat{\xi}_2|^2 e^{-\beta t}.$$

The proof is complete. □

Lemma 9. Under $(\mathbf{H}_{b,\sigma}^1)$ and $(\mathbf{H}_{b,\sigma}^2)$, it holds that for any $t \geq 0$

$$\mathbb{W}_2^2(\mathcal{L}_{\widehat{X}_t^{\widehat{\xi}}}^{\widehat{\mathcal{Q}}}, \nu) \leq 2e^{-\beta t} (\widehat{\mathbb{E}}|\widehat{\xi}|^2 + \nu(|\cdot|^2)). \tag{10}$$

Proof. Let $\widehat{\zeta}$ be a $\widehat{\mathcal{F}}_0$ -measurable random variable with $\mathcal{L}_{\widehat{\zeta}}^{\widehat{\mathcal{Q}}} = \nu$. And by the definition of ν , we know that $\mathcal{L}_{\widehat{X}_t^{\widehat{\xi}}}^{\widehat{\mathcal{Q}}} = \nu$ for any $t \geq 0$. So, (9) implies that

$$\mathbb{W}_2^2(\mathcal{L}_{\widehat{X}_t^{\widehat{\xi}}}^{\widehat{\mathcal{Q}}}, \nu) = \mathbb{W}_2^2(\mathcal{L}_{\widehat{X}_t^{\widehat{\xi}}}^{\widehat{\mathcal{Q}}}, \mathcal{L}_{\widehat{X}_t^{\widehat{\xi}}}^{\widehat{\mathcal{Q}}}) \leq \widehat{\mathbb{E}}|\widehat{X}_t^{\widehat{\xi}} - \widehat{X}_t^{\widehat{\zeta}}|^2 \leq \widehat{\mathbb{E}}|\widehat{\xi} - \widehat{\zeta}|^2 e^{-\beta t} \leq 2e^{-\beta t} (\widehat{\mathbb{E}}|\widehat{\xi}|^2 + \nu(|\cdot|^2)).$$

□

Lemma 10. Under $(\mathbf{H}_{b,\sigma}^1)$ and $(\mathbf{H}_{b,\sigma}^2)$, it holds that for any $t \geq 0$

$$\left| \widehat{\mathbb{E}}h(\widehat{X}_t^{\widehat{\xi}}, \nu) - \bar{h} \right| \leq 2L_2 e^{-\beta t} (\widehat{\mathbb{E}}|\widehat{\xi}|^2 + \nu(|\cdot|^2)). \tag{11}$$

Proof. From the definition of v and (9), it follows that

$$\begin{aligned} \left| \widehat{\mathbb{E}}h(\widehat{X}_t^\xi, v) - \bar{h} \right|^2 &= \left| \widehat{\mathbb{E}}h(\widehat{X}_t^\xi, v) - \int_{\mathbb{R}^n} h(x, v) \nu(dx) \right|^2 = \left| \widehat{\mathbb{E}}h(\widehat{X}_t^\xi, v) - \int_{\mathbb{R}^n} \widehat{\mathbb{E}}h(\widehat{X}_t^x, v) \nu(dx) \right|^2 \\ &\leq \widehat{\mathbb{E}} \int_{\mathbb{R}^n} \left| h(\widehat{X}_t^\xi, v) - h(\widehat{X}_t^x, v) \right|^2 \nu(dx) \leq L_2 \int_{\mathbb{R}^n} \widehat{\mathbb{E}}|\widehat{X}_t^\xi - \widehat{X}_t^x|^2 \nu(dx) \\ &\leq L_2 e^{-\beta t} \int_{\mathbb{R}^n} \widehat{\mathbb{E}}|\widehat{\xi} - x|^2 \nu(dx) \leq 2L_2 e^{-\beta t} (\widehat{\mathbb{E}}|\widehat{\xi}|^2 + v(|\cdot|^2)), \end{aligned}$$

which completes the proof. □

Proof of Theorem 2. First of all, by (1) it holds that for any $t \in [0, T]$

$$\begin{aligned} |Y_t^\varepsilon - \bar{Y}_t|^2 &= \left| \int_0^t \left(h(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) - \bar{h} \right) ds \right|^2 \\ &\leq 2 \left| \int_0^t \left(h(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) - h(X_s^\varepsilon, v) \right) ds \right|^2 + 2 \left| \int_0^t \left(h(X_s^\varepsilon, v) - \bar{h} \right) ds \right|^2 \\ &=: I_1 + I_2. \end{aligned}$$

For I_1 , the Hölder inequality and (\mathbf{H}_h) imply that

$$\begin{aligned} \mathbb{E} \sup_{t \in [0, T]} I_1 &\leq 2TL_2 \int_0^T \mathbb{W}_2^2(\mathcal{L}_{X_s^\varepsilon}^Q, v) ds \leq 2TL_2 \int_0^T \mathbb{W}_2^2(\mathcal{L}_{\widehat{X}_s^\xi}^Q, v) ds \\ &\stackrel{(10)}{\leq} 2TL_2 \int_0^T 2e^{-\beta s} (\widehat{\mathbb{E}}|\widehat{\xi}|^2 + v(|\cdot|^2)) ds \\ &\leq 4TL_2 \beta^{-1} (\widehat{\mathbb{E}}|\widehat{\xi}|^2 + v(|\cdot|^2)) \varepsilon. \end{aligned} \tag{12}$$

For I_2 , we divide it into two parts and obtain that

$$\mathbb{E} \sup_{t \in [0, T]} I_2 \leq 4\mathbb{E} \sup_{t \in [0, T]} \left| \int_0^{\lfloor \frac{t}{\delta} \rfloor \delta} \left(h(X_s^\varepsilon, v) - \bar{h} \right) ds \right|^2 + 4\mathbb{E} \sup_{t \in [0, T]} \left| \int_{\lfloor \frac{t}{\delta} \rfloor \delta}^t \left(h(X_s^\varepsilon, v) - \bar{h} \right) ds \right|^2 = I_{21} + I_{22}, \tag{13}$$

where δ is a fixed positive number depending on ε and $\lfloor \frac{t}{\delta} \rfloor$ denotes the integer part of $\frac{t}{\delta}$.

Let us deal with I_{21} . Some computation implies that

$$\begin{aligned} I_{21} &\leq 4\mathbb{E} \sup_{t \in [0, T]} \left| \sum_{k=0}^{\lfloor \frac{t}{\delta} \rfloor - 1} \int_{k\delta}^{(k+1)\delta} \left(h(X_s^\varepsilon, v) - \bar{h} \right) ds \right|^2 \\ &\leq 4 \left[\frac{T}{\delta} \right] \mathbb{E} \sum_{k=0}^{\lfloor \frac{T}{\delta} \rfloor - 1} \left| \int_{k\delta}^{(k+1)\delta} \left(h(X_s^\varepsilon, v) - \bar{h} \right) ds \right|^2 \\ &\leq 4 \left(\frac{T}{\delta} \right)^2 \sup_{0 \leq k \leq \lfloor \frac{T}{\delta} \rfloor - 1} \mathbb{E} \left| \int_{k\delta}^{(k+1)\delta} \left(h(X_s^\varepsilon, v) - \bar{h} \right) ds \right|^2 \\ &= 4 \left(\frac{T}{\delta} \right)^2 \varepsilon^2 \sup_{0 \leq k \leq \lfloor \frac{T}{\delta} \rfloor - 1} \mathbb{E} \left| \int_0^{\delta/\varepsilon} \left(h(X_{s\varepsilon+k\delta}^\varepsilon, v) - \bar{h} \right) ds \right|^2 \\ &= 8 \left(\frac{T}{\delta} \right)^2 \varepsilon^2 \sup_{0 \leq k \leq \lfloor \frac{T}{\delta} \rfloor - 1} \int_0^{\delta/\varepsilon} \int_r^{\delta/\varepsilon} \mathbb{E} \langle h(X_{s\varepsilon+k\delta}^\varepsilon, v) - \bar{h}, h(X_{r\varepsilon+k\delta}^\varepsilon, v) - \bar{h} \rangle ds dr. \end{aligned} \tag{14}$$

Set $0 \leq r \leq s \leq \delta/\varepsilon$

$$\Psi(s, r) := \mathbb{E} \left\langle h(X_{s\varepsilon+k\delta}^\varepsilon, \nu) - \bar{h}, h(X_{r\varepsilon+k\delta}^\varepsilon, \nu) - \bar{h} \right\rangle,$$

and it holds that

$$\begin{aligned} \Psi(s, r) &= \mathbb{E} \left[\mathbb{E} \left[\langle h(X_{s\varepsilon+k\delta}^\varepsilon, \nu) - \bar{h}, h(X_{r\varepsilon+k\delta}^\varepsilon, \nu) - \bar{h} \rangle \mid \mathcal{F}_{k\delta} \right] \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\langle h(X_{s\varepsilon}^{\varepsilon, x}, \nu) - \bar{h}, h(X_{r\varepsilon}^{\varepsilon, x}, \nu) - \bar{h} \rangle \mid_{x=X_{k\delta}^\varepsilon} \right] \right], \end{aligned}$$

where $X_{\varepsilon s}^{\varepsilon, x}$ satisfies the following equation

$$X_{\varepsilon s}^{\varepsilon, x} = x + \frac{1}{\varepsilon} \int_0^{\varepsilon s} b(X_u^{\varepsilon, x}) du + \frac{1}{\sqrt{\varepsilon}} \int_0^{\varepsilon s} \sigma(X_u^{\varepsilon, x}) dB_u.$$

Besides, we notice that for $0 \leq s \leq \delta/\varepsilon$

$$X_{\varepsilon s}^{\varepsilon, x} = x + \int_0^s b(X_{\varepsilon v}^{\varepsilon, x}) dv + \int_0^s \sigma(X_{\varepsilon v}^{\varepsilon, x}) d\tilde{B}_v,$$

where $\tilde{B}_v = \frac{1}{\sqrt{\varepsilon}} B_{\varepsilon v}$. Since \tilde{B} is a d -dimensional standard Brownian motion, $X_{\varepsilon s}^{\varepsilon, x}$ and the solution \hat{X}_s^x of (6) have the same distributions for $0 \leq s \leq \delta/\varepsilon$. Thus,

$$\begin{aligned} \mathbb{E} \left[\langle h(X_{s\varepsilon}^{\varepsilon, x}, \nu) - \bar{h}, h(X_{r\varepsilon}^{\varepsilon, x}, \nu) - \bar{h} \rangle \right] &= \hat{\mathbb{E}} \langle h(\hat{X}_s^x, \nu) - \bar{h}, h(\hat{X}_r^x, \nu) - \bar{h} \rangle \\ &= \hat{\mathbb{E}} \left[\langle \hat{\mathbb{E}} [h(\hat{X}_s^x, \nu) \mid \mathcal{F}_r^{\hat{B}}] - \bar{h}, h(\hat{X}_r^x, \nu) - \bar{h} \rangle \right] \\ &\leq \left(\hat{\mathbb{E}} \left[\hat{\mathbb{E}} [h(\hat{X}_{s-r}^\theta, \nu) \mid_{\theta=\hat{X}_r^x} - \bar{h}]^2 \right] \right)^{1/2} \left(\hat{\mathbb{E}} [h(\hat{X}_r^x, \nu) - \bar{h}]^2 \right)^{1/2}. \end{aligned}$$

On one hand, by (11) and (8), it holds that

$$\begin{aligned} \left(\hat{\mathbb{E}} \left[\hat{\mathbb{E}} [h(\hat{X}_{s-r}^\theta, \nu) \mid_{\theta=\hat{X}_r^x} - \bar{h}]^2 \right] \right)^{1/2} &\leq C e^{-\beta(s-r)/2} (\hat{\mathbb{E}} |\hat{X}_r^x|^2 + \nu(|\cdot|^2))^{1/2} \\ &\leq C e^{-\beta(s-r)/2} (1 + |x|^2 + \nu(|\cdot|^2))^{1/2}. \end{aligned}$$

On the other hand, (\mathbf{H}_h) and (8) imply that

$$\begin{aligned} \left(\hat{\mathbb{E}} |h(\hat{X}_r^x, \nu) - \bar{h}|^2 \right)^{1/2} &= \left(\hat{\mathbb{E}} |h(\hat{X}_r^x, \nu) - \int_{\mathbb{R}^n} h(u, \nu) \nu(du)|^2 \right)^{1/2} \\ &\leq \left(\int_{\mathbb{R}^n} \hat{\mathbb{E}} |\hat{X}_r^x - u|^2 \nu(du) \right)^{1/2} \\ &\leq C(1 + |x|^2 + \nu(|\cdot|^2))^{1/2}. \end{aligned}$$

Combining the two calculations, we have that

$$\mathbb{E} \left[\langle h(X_{s\varepsilon+k\delta}^\varepsilon, \nu) - \bar{h}, h(X_{r\varepsilon+k\delta}^\varepsilon, \nu) - \bar{h} \rangle \right] \leq C e^{-\beta(s-r)/2} (1 + |x|^2 + \nu(|\cdot|^2)),$$

and

$$\begin{aligned} \Psi(s, r) &\leq C e^{-\beta(s-r)/2} (1 + \mathbb{E} |X_{k\delta}^\varepsilon|^2 + \nu(|\cdot|^2)) = C e^{-\beta(s-r)/2} \left(1 + \hat{\mathbb{E}} |\hat{X}_{k\delta/\varepsilon}^{\hat{\xi}}|^2 + \nu(|\cdot|^2) \right) \\ &\stackrel{(8)}{\leq} C e^{-\beta(s-r)/2} (1 + \hat{\mathbb{E}} |\hat{\xi}|^2 + \nu(|\cdot|^2)), \end{aligned}$$

which together with (14) yields that

$$\begin{aligned}
 I_{21} &\leq 8 \left(\frac{T}{\delta}\right)^2 \varepsilon^2 \sup_{0 \leq k \leq \lfloor \frac{T}{\delta} \rfloor - 1} \int_0^{\delta/\varepsilon} \int_r^{\delta/\varepsilon} C e^{-\beta(s-r)/2} (1 + \widehat{\mathbb{E}}|\widehat{\xi}|^2 + v(|\cdot|^2)) ds dr \\
 &\leq 8T^2 C (1 + \widehat{\mathbb{E}}|\widehat{\xi}|^2 + v(|\cdot|^2)) \frac{\varepsilon}{\delta}.
 \end{aligned}
 \tag{15}$$

Now we treat I_{22} . Based on (\mathbf{H}_h) and the Hölder inequality, it holds that

$$I_{22} \leq 4\delta \mathbb{E} \sup_{t \in [0, T]} \int_{\lfloor \frac{t}{\delta} \rfloor \delta}^t |h(X_s^\varepsilon, \nu) - \bar{h}|^2 ds \leq 4\delta \int_0^T \mathbb{E} |h(X_s^\varepsilon, \nu) - \bar{h}|^2 ds \leq C\delta.
 \tag{16}$$

Finally, collecting (12) (15) (16), one can conclude that

$$\mathbb{E} \sup_{t \in [0, T]} |Y_t^\varepsilon - \bar{Y}_t|^2 \leq C(\varepsilon + \frac{\varepsilon}{\delta} + \delta).$$

Taking $\delta = \varepsilon^\gamma$ for $0 < \gamma < 1$, we obtain the required estimate. The proof is complete. □

5. Proof of Theorem 4

In this section, we prove Theorem 4. First of all, we make some preparations.

Lemma 11. *Under the assumption (\mathbf{H}_h) , there exists a constant $C > 0$ such that*

$$\mathbb{E} |\mathbb{P}_t^\varepsilon(1)|^{-r} \leq \exp\{(2r^2 + r + 1)CT/2\}, \quad r > 1.$$

Since its proof is similar to that of [9, Lemma 3.6], we omit it.

Proposition 12. *Assume that $(\mathbf{H}_{b,\sigma}^1)$, (\mathbf{H}_h) , $(\mathbf{H}_{b,\sigma}^2)$ hold. Then we have that for any $t \in [0, T]$*

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left| \pi_t^\varepsilon(F) - \bar{\pi}_t(\bar{F}) \right|^2 = 0, \quad F \in C_b(\mathbb{R}^n \times \mathcal{D}_2(\mathbb{R}^n)).$$

Proof. We divide the proof into two parts. In the first step, we prove the main result. In the second step, we show a key convergence which is used in the first step.

Step 1. We prove that $\lim_{\varepsilon \rightarrow 0} \mathbb{E} |\pi_t^\varepsilon(F) - \bar{\pi}_t(\bar{F})|^2 = 0$.

Note that

$$\mathbb{E} |\pi_t^\varepsilon(F) - \bar{\pi}_t(\bar{F})|^2 \leq 2\mathbb{E} |\pi_t^\varepsilon(F) - \pi_t^\varepsilon(\bar{F})|^2 + 2\mathbb{E} |\pi_t^\varepsilon(\bar{F}) - \bar{\pi}_t(\bar{F})|^2,$$

and

$$\pi_t^\varepsilon(\bar{F}) = \bar{F}, \quad \bar{\pi}_t(\bar{F}) = \frac{\bar{\mathbb{P}}_t(\bar{F})}{\bar{\mathbb{P}}_t(1)} = \frac{\bar{F}\bar{\Lambda}_t}{\bar{\Lambda}_t} = \bar{F}.$$

Thus, we only need to prove

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} |\pi_t^\varepsilon(F) - \pi_t^\varepsilon(\bar{F})|^2 = 0.$$

On one hand, by the result in **Step 2**, we know that $|\pi_t^\varepsilon(F) - \pi_t^\varepsilon(\bar{F})| \xrightarrow{Q} 0$ as $\varepsilon \rightarrow 0$. On the other hand, F and \bar{F} are bounded. Thus, the bounded dominated convergence theorem yields that

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} |\pi_t^\varepsilon(F) - \pi_t^\varepsilon(\bar{F})|^2 = 0.$$

Step 2. We prove that $|\pi_t^\varepsilon(F) - \pi_t^\varepsilon(\bar{F})| \xrightarrow{Q} 0$ as $\varepsilon \rightarrow 0$.

By the Hölder inequality, it holds that

$$\mathbb{E} \left| \pi_t^\varepsilon(F) - \pi_t^\varepsilon(\bar{F}) \right| = \widetilde{\mathbb{E}} \left| \pi_t^\varepsilon(F) - \pi_t^\varepsilon(\bar{F}) \right| \Lambda_T^\varepsilon \leq \left(\widetilde{\mathbb{E}} |\pi_t^\varepsilon(F) - \pi_t^\varepsilon(\bar{F})|^{r_1} \right)^{1/r_1} \left(\widetilde{\mathbb{E}} (\Lambda_T^\varepsilon)^{r_2} \right)^{1/r_2},$$

where $1 < r_1 < 2, r_2 > 1, 1/r_1 + 1/r_2 = 1$. On one hand, it is not difficult to prove that

$$\left(\widetilde{\mathbb{E}} (\Lambda_T^\varepsilon)^{r_2} \right)^{1/r_2} \leq \exp\{CT\}.$$

On the other hand, the Kallianpur–Striebel formula and the Hölder inequality imply that

$$\begin{aligned} \left(\tilde{\mathbb{E}}|\pi_t^\varepsilon(F) - \pi_t^\varepsilon(\bar{F})|^{r_1}\right)^{1/r_1} &= \left(\tilde{\mathbb{E}}|\mathbb{P}_t^\varepsilon(F) - \mathbb{P}_t^\varepsilon(\bar{F})|^{r_1} \mathbb{P}_t^\varepsilon(1)^{-r_1}\right)^{1/r_1} \\ &\leq \left(\tilde{\mathbb{E}}|\mathbb{P}_t^\varepsilon(F) - \mathbb{P}_t^\varepsilon(\bar{F})|^2\right)^{1/2} \left(\tilde{\mathbb{E}}\mathbb{P}_t^\varepsilon(1)^{-2r_1/(2-r_1)}\right)^{\frac{2-r_1}{2r_1}} \\ &\leq C \left(\tilde{\mathbb{E}}|\mathbb{P}_t^\varepsilon(F) - \mathbb{P}_t^\varepsilon(\bar{F})|^2\right)^{1/2}, \end{aligned}$$

where the last inequality is based on Lemma 11.

Next, we are devoted to estimating $\tilde{\mathbb{E}}|\mathbb{P}_t^\varepsilon(F) - \mathbb{P}_t^\varepsilon(\bar{F})|^2$. In order to do this, we take an independent copy \hat{X}^ε of X^ε , which has the same distribution to that for X^ε and is independent of (X^ε, W) . It follows from the Jensen inequality that

$$\begin{aligned} &\tilde{\mathbb{E}}|\mathbb{P}_t^\varepsilon(F) - \mathbb{P}_t^\varepsilon(\bar{F})|^2 \\ &= \tilde{\mathbb{E}} \left[\left| \tilde{\mathbb{E}} \left[F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) \Lambda_t^\varepsilon \mid \mathcal{F}_t^{Y^\varepsilon} \right] - \tilde{\mathbb{E}} \left[\bar{F} \Lambda_t^\varepsilon \mid \mathcal{F}_t^{Y^\varepsilon} \right] \right|^2 \right] \\ &\leq \tilde{\mathbb{E}} \left[\tilde{\mathbb{E}} \left[\left| F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) \Lambda_t^\varepsilon - \bar{F} \Lambda_t^\varepsilon \right|^2 \mid \mathcal{F}_t^{Y^\varepsilon} \right] \right] \\ &= \tilde{\mathbb{E}} \left[\tilde{\mathbb{E}} \left[(F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F})(F(\hat{X}_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F}) \right. \right. \\ &\quad \times \exp \left\{ \int_0^t (h^i(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) + h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q)) d(Y_s^\varepsilon)^i \right. \\ &\quad \left. \left. - \frac{1}{2} \int_0^t (|h(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q)|^2 + |h(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q)|^2) ds \right\} \mid \mathcal{F}_t^{Y^\varepsilon} \right] \right] \\ &= \tilde{\mathbb{E}} \left[\tilde{\mathbb{E}} \left[(F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F})(F(\hat{X}_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F}) \right. \right. \\ &\quad \times \exp \left\{ \int_0^t (h^i(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) + h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q)) d(Y_s^\varepsilon)^i \right. \\ &\quad \left. \left. - \frac{1}{2} \int_0^t |h(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) + h(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q)|^2 ds \right\} \right. \right. \\ &\quad \left. \left. \times \exp \left\{ \int_0^t h^i(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right\} \mid \mathcal{F}_t^{X^\varepsilon, \hat{X}^\varepsilon} \right] \right] \\ &= \tilde{\mathbb{E}} \left[(F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F})(F(\hat{X}_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F}) \exp \left\{ \int_0^t h^i(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right\} \right] \\ &= \tilde{\mathbb{E}} \left[(F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F})(F(\hat{X}_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F}) \right. \\ &\quad \times \left(\exp \left\{ \int_0^t h^i(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right\} - \exp \left\{ \int_0^t \bar{h}^i h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right\} \right) \right. \\ &\quad \left. + \tilde{\mathbb{E}} \left[(F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F})(F(\hat{X}_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F}) \exp \left\{ \int_0^t \bar{h}^i h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right\} \right] \right] \\ &=: J_1 + J_2. \tag{17} \end{aligned}$$

For J_1 , by (\mathbf{H}_h) it holds that

$$J_1 \leq 8 \|F\|_\infty^2 e^{\|h\|_\infty^2 T} \tilde{\mathbb{E}} \left| \int_0^t (h^i(X_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) - \bar{h}^i) h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right|$$

where $\|F\|_\infty$ denotes the boundedness of F and we use the fact that $|e^u - e^v| \leq (e^u + e^v)|u - v|$ for any $u, v \in \mathbb{R}$. Since \hat{X}^ε and X^ε are independent, the ergodicity of X^ε yields that

$$\lim_{\varepsilon \rightarrow 0} J_1 = 0. \tag{18}$$

For J_2 , we obtain that for arbitrary small $\delta > 0$

$$\begin{aligned}
 J_2 &\leq 4\|F\|_\infty^2 \tilde{\mathbb{E}} \left| \exp \left\{ \int_0^t \bar{h}^i h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right\} - \exp \left\{ \int_0^{t-\delta} \bar{h}^i h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right\} \right| \\
 &\quad + \left| \tilde{\mathbb{E}} \left[(F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F}) \middle| \mathcal{F}_{t-\delta}^{X^\varepsilon} \vee \mathcal{F}_t^{\hat{X}^\varepsilon} \right] (F(\hat{X}_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F}) \times \exp \left\{ \int_0^{t-\delta} \bar{h}^i h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right\} \right| \right| \\
 &\leq 4\|F\|_\infty^2 \tilde{\mathbb{E}} \left| \exp \left\{ \int_0^t \bar{h}^i h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right\} - \exp \left\{ \int_0^{t-\delta} \bar{h}^i h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right\} \right| \\
 &\quad + 2\|F\|_\infty \tilde{\mathbb{E}} \left[\left| \tilde{\mathbb{E}} \left[(F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - \bar{F}) \middle| \mathcal{F}_{t-\delta}^{X^\varepsilon} \right] \right| \exp \left\{ \int_0^{t-\delta} \bar{h}^i h^i(\hat{X}_s^\varepsilon, \mathcal{L}_{X_s^\varepsilon}^Q) ds \right\} \right] \\
 &=: J_{21} + J_{22}.
 \end{aligned}$$

By (\mathbf{H}_h) and the dominated convergence theorem, we know that

$$\lim_{\delta \rightarrow 0} J_{21} = 0. \tag{19}$$

For J_{22} , the ergodicity of X^ε implies that for any $\delta > 0$

$$\lim_{\varepsilon \rightarrow 0} J_{22} = 0. \tag{20}$$

Finally, combining (18) (19) (20) with (17) and first letting $\varepsilon \rightarrow 0$ and $\delta \rightarrow 0$, we conclude that

$$\lim_{\varepsilon \rightarrow 0} \tilde{\mathbb{E}} \left| \mathbb{P}_t^\varepsilon(F) - \mathbb{P}_t^\varepsilon(\bar{F}) \right|^2 = 0,$$

and furthermore

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left| \pi_t^\varepsilon(F) - \pi_t^\varepsilon(\bar{F}) \right| = 0.$$

Since the L^1 convergence implies the convergence in probability, $|\pi_t^\varepsilon(F) - \pi_t^\varepsilon(\bar{F})|$ tends to 0 in probability as $\varepsilon \rightarrow 0$. The proof is complete. \square

Now, it is the position to prove Theorem 4.

Proof of Theorem 4. For any $F \in \mathcal{A}_\eta$, set for $x \in \mathbb{R}^n, \mu \in \mathcal{P}_2(\mathbb{R}^n)$

$$F_k(x, \mu) := \phi_k(F(x, \mu)), \quad \bar{F}_k = \int_{\mathbb{R}^n} F_k(x, \nu) \nu(dx), \quad k \in \mathbb{N},$$

where $\phi_k(z) = z, |z| \leq k$ and $\phi_k(z) = k \text{sign}(z), |z| > k$, and it holds that $|F_k(x, \mu)| \leq k$ and

$$\begin{aligned}
 |F_k(x, \mu) - F(x, \mu)|^{2+\eta/2} &\leq 2^{2+\eta/2} |F(x, \mu)|^{2+\eta/2} I_{\{|F(x, \mu)| > k\}}(x, \mu) \\
 &\leq 2^{2+\eta/2} k^{-\eta/2} |F(x, \mu)|^{2+\eta},
 \end{aligned} \tag{21}$$

$$\begin{aligned}
 |\bar{F}_k - \bar{F}|^2 &\leq \left| \int_{\mathbb{R}^n} F_k(x, \nu) \nu(dx) - \int_{\mathbb{R}^n} F(x, \nu) \nu(dx) \right|^2 \\
 &\leq 4 \int_{\mathbb{R}^n} |F(x, \mu)|^2 I_{\{|F(x, \mu)| > k\}}(x) \nu(dx).
 \end{aligned} \tag{22}$$

Since $F_k \in C_b(\mathbb{R}^n \times \mathcal{P}_2(\mathbb{R}^n))$, Proposition 12 implies that

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} |\pi_t^\varepsilon(F_k) - \bar{\pi}_t(\bar{F}_k)|^2 = 0. \tag{23}$$

Besides, by the Jensen inequality and the Hölder inequality, it holds that

$$\begin{aligned} \mathbb{E}|\pi_t^\varepsilon(F) - \pi_t^\varepsilon(F_k)|^2 &= \mathbb{E}\left|\mathbb{E}\left[F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) \mid \mathcal{F}_t^{Y^\varepsilon}\right] - \mathbb{E}\left[F_k(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) \mid \mathcal{F}_t^{Y^\varepsilon}\right]\right|^2 \\ &\leq \mathbb{E}\left|F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - F_k(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q)\right|^2 \\ &\leq \left(\mathbb{E}\left|F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) - F_k(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q)\right|^{2+\eta/2}\right)^{\frac{2}{2+\eta/2}} \\ &\stackrel{(21)}{\leq} 4k^{-\frac{\eta}{2+\eta/2}} \left(\mathbb{E}\left|F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q)\right|^{2+\eta}\right)^{\frac{2}{2+\eta/2}} \\ &\leq 4k^{-\frac{\eta}{2+\eta/2}} \left(\sup_{\varepsilon \in (0,1)} \mathbb{E}\left|F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q)\right|^{2+\eta}\right)^{\frac{2}{2+\eta/2}}, \end{aligned}$$

and furthermore for $\varepsilon \in (0, 1)$

$$\lim_{k \rightarrow \infty} \mathbb{E}|\pi_t^\varepsilon(F) - \pi_t^\varepsilon(F_k)|^2 = 0. \tag{24}$$

Finally, we notice that

$$\mathbb{E}|\pi_t^\varepsilon(F) - \bar{\pi}_t(\bar{F})|^2 \leq 3\mathbb{E}|\pi_t^\varepsilon(F) - \pi_t^\varepsilon(F_k)|^2 + 3\mathbb{E}|\pi_t^\varepsilon(F_k) - \bar{\pi}_t(\bar{F}_k)|^2 + 3\mathbb{E}|\bar{\pi}_t(\bar{F}_k) - \bar{\pi}_t(\bar{F})|^2.$$

Based on (22) (23) (24), it holds that

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E}|\pi_t^\varepsilon(F) - \bar{\pi}_t(\bar{F})|^2 = 0.$$

The proof is complete. □

6. An example

In this section, we give an example to explain our results.

Example 13. For $n = d = m = 1$, consider the following multiscale McKean–Vlasov stochastic system $(X_t^\varepsilon, Y_t^\varepsilon)$ on $\mathbb{R} \times \mathbb{R}$:

$$\begin{cases} dX_t^\varepsilon = -\frac{1}{\varepsilon} \frac{X_t^\varepsilon}{2} dt + \frac{1}{\sqrt{\varepsilon}} \sigma dB_t, & 0 \leq t \leq T, \\ X_0^\varepsilon = x_0, \\ dY_t^\varepsilon = \int_{\mathbb{R}} \sin|x + u| \nu(dx) + u | \mathcal{L}_{X_t^\varepsilon}^Q(du) dt + dW_t, & 0 \leq t \leq T, \\ Y_0^\varepsilon = 0, \end{cases} \tag{25}$$

where $x_0 \in \mathbb{R}$ and $\sigma > 0$ are two constants. It is easy to see that $b(x) = -x/2, \sigma(x) = \sigma, h(x, \mu) = \int_{\mathbb{R}} \sin|x + u| \mu(du)$ satisfy $(\mathbf{H}_{b,\sigma}^1), (\mathbf{H}_h), (\mathbf{H}_{b,\sigma}^2)$ with $L_1 = \frac{1}{4}, L_2 = 1, \beta = 1$. Note that the invariant measure $\nu(dx)$ of X^1 is $\frac{1}{\sqrt{\pi\sigma}} e^{-\frac{x^2}{\sigma^2}} dx$. Thus, by Theorem 2, we obtain that

$$\mathbb{E} \sup_{t \in [0, T]} |Y_t^\varepsilon - \bar{Y}_t|^2 \leq C(\varepsilon + \varepsilon^{1-\gamma} + \varepsilon^\gamma),$$

where $\bar{Y}_t = \int_{\mathbb{R}} \int_{\mathbb{R}} \sin|x + u| \nu(du) \nu(dx) t + W_t$.

Next, for any $F \in \mathcal{B}_b(\mathbb{R} \times \mathcal{P}_2(\mathbb{R}))$ define

$$\pi_t^\varepsilon(F) := \mathbb{E}\left[F(X_t^\varepsilon, \mathcal{L}_{X_t^\varepsilon}^Q) \mid \mathcal{F}_t^{Y^\varepsilon}\right],$$

and

$$\begin{aligned} \bar{\Lambda}_t &:= \exp\left\{\bar{h}^i \bar{Y}_t^i - \frac{1}{2} |\bar{h}|^2 t\right\}, & \bar{F} &= \int_{\mathbb{R}} F(x, \nu) \nu(dx), \\ \bar{\mathbb{P}}_t(\bar{F}) &:= \tilde{\mathbb{E}}\left[\bar{F} \bar{\Lambda}_t \mid \mathcal{F}_t^{\bar{Y}}\right], & \bar{\pi}_t(\bar{F}) &:= \frac{\bar{\mathbb{P}}_t(\bar{F})}{\bar{\mathbb{P}}_t(1)}, \end{aligned}$$

and by Theorem 4 it holds that

$$\lim_{\varepsilon \rightarrow 0} \mathbb{E} \left| \pi_t^\varepsilon(F) - \bar{\pi}_t(\bar{F}) \right|^2 = 0, \quad F \in \mathcal{A}_\eta.$$

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

References

- [1] Z. W. Bezemek and K. Spiliopoulos, “Rate of homogenization for fully-coupled McKean–Vlasov SDEs”, *Stoch. Dyn.* **23** (2023), no. 2, article no. 2350013 (65 pages).
- [2] J. Gao, W. Hong and W. Liu, “Small noise asymptotics of multi-scale McKean–Vlasov stochastic dynamical systems”, *J. Differ. Equations* **364** (2023), pp. 521–575.
- [3] R. Z. Has’minskiĭ, “On the principle of averaging the Itô’s stochastic differential equations”, *Kybernetika* **4** (1968), pp. 260–279.
- [4] D. Liu, “Strong convergence of principle of averaging for multiscale stochastic dynamical systems”, *Commun. Math. Sci.* **8** (2010), no. 4, pp. 999–1020.
- [5] M. Liu and H. Qiao, “Uniqueness and superposition of the space-distribution-dependent Zakai equations”, *Infin. Dimens. Anal. Quantum Probab. Relat. Top.* **27** (2024), no. 1, article no. 2350014 (24 pages).
- [6] W. Liu, M. Röckner, X. Sun and Y. Xie, “Averaging principle for slow-fast stochastic differential equations with time dependent locally Lipschitz coefficients”, *J. Differ. Equations* **268** (2020), no. 6, pp. 2910–2948.
- [7] A. Papanicolaou and K. Spiliopoulos, “Filtering the maximum likelihood for multiscale problems”, *Multiscale Model. Simul.* **12** (2014), no. 3, pp. 1193–1229.
- [8] H. Qiao, “Average principles and large deviation principles of multiscale multivalued McKean–Vlasov stochastic systems”, 2023. <http://arxiv.org/abs/2307.14561>.
- [9] H. Qiao, “Convergence of nonlinear filtering for multiscale systems with correlated Lévy noises”, *Stoch. Dyn.* **23** (2023), no. 2, article no. 2350016 (30 pages).
- [10] H. Qiao and W. Wei, “Strong approximation of nonlinear filtering for multiscale McKean–Vlasov stochastic systems”, 2022. <http://arxiv.org/abs/2206.05037>.
- [11] H. Qiao and W. Wei, “Weak approximation of nonlinear filtering for multiscale McKean–Vlasov stochastic systems”, 2022. <http://arxiv.org/abs/2212.00240>.
- [12] M. Röckner, X. Sun and Y. Xie, “Strong convergence order for slow-fast McKean–Vlasov stochastic differential equations”, *Ann. Inst. Henri Poincaré, Probab. Stat.* **57** (2021), no. 1, pp. 547–576.
- [13] N. Sen and P. E. Caines, “Nonlinear filtering for McKean–Vlasov type SDEs with application to mean field games”, in *Proceedings of the 21st International symposium on Mathematical Theory of Networks and Systems, Groningen, The Netherlands, 2014*, pp. 1344–1351.
- [14] N. Şen and P. E. Caines, “Nonlinear filtering theory for McKean–Vlasov type stochastic differential equations”, *SIAM J. Control Optim.* **54** (2016), no. 1, pp. 153–174.
- [15] F.-Y. Wang, “Distribution dependent SDEs for Landau type equations”, *Stochastic Processes Appl.* **128** (2018), no. 2, pp. 595–621.
- [16] J. Xu, J. Liu, J. Liu and Y. Miao, “Strong averaging principle for two-time-scale stochastic McKean–Vlasov equations”, *Appl. Math. Optim.* **84** (2021), no. suppl. 1, S837–S867.