Stochastic averaging principle for multi-valued McKean-Vlasov stochastic differential equations *

Guangjun Shen†, Jie Xiang†, Jiang-Lun Wu ‡,§

† Department of Mathematics, Anhui Normal University, Wuhu 241000, China.
‡ Department of Mathematics, Computational Foundry, Swansea University, Swansea SA1 8EN, UK

Abstract

In this paper, we are concerned with stochastic averaging principle for multi-valued McKean-Vlasov stochastic differential equations. Under certain averaging conditions, we show that solutions of multi-valued McKean-Vlasov stochastic differential equations can be approximated by solutions of the associated averaged multi-valued McKean-Vlasov stochastic differential equations in the sense of the mean square convergence.

Keywords: Multi-valued McKean-Vlasov stochastic differential equations; Stochastic averaging principle

1. Introduction

Stochastic differential equations (SDEs) play a significant role in modelling evolutions of dynamical systems when taking into account uncertainty features along time in diverse fields ranging from biology, chemistry, and physics, as well as economics and finance, etc. Recently, there has been an increasing interest to study multi-valued SDEs. There have been many fundamental studies addressing the existence and uniqueness, asymptotic behaviors of solutions ([1], [8], [2] and references therein). There are also increasing interest to investigate McKean-Vlasov SDEs (also known as distribution dependent SDEs or mean field SDEs), as the profound nonlinear Fokker-Planck equations can be characterised by McKean-Vlasov SDEs. A distinct feature of such systems is the appearance of probability laws in the coefficients of the resulting equations, due to this, usual and standard techniques developed for SDEs are no longer applicable to McKean-Vlasov SDEs. One has to utilise iteration in distributions in a proper manner to overcome the difficulties and to remedy challenge problems. More recently, Gong and Qiao [3] studied multi-valued McKean-Vlasov SDEs with non-Lipschitz coefficients by showing the existence and uniqueness of strong solutions. On the other hand, the averaging principle, initiated by Khasminskii in [5], is a very efficient and important tool in study of SDEs for modelling problems arising in many practical research areas. The averaging principle enables one to study complex equations with related averaged (yet comparably simpler) equations, which paves a convenient and easy way to study many important properties. To the best of our knowledge, the literature involving averaging principles for multi-valued SDEs is rather rare. In the distribution independent case, averaging principles for multi-valued SDEs driven by Brownian motion was obtained by Ngoran and Modeste [7], Xu and Liu [9]. Guo and Pei [4], Mao et al. [6] considered averaging principles for multi-valued SDEs driven by Poisson point processes. However, averaging principles for multi-valued McKean-Vlasov SDEs have yet not been considered.

*This research is supported by the National Natural Science Foundation of China (12071003).
§ Corresponding author: Jiang-Lun Wu (j.l.wu@swansea.ac.uk).

The objective of the present paper is to establish a stochastic averaging principle for the following multi-valued McKean-Vlasov SDEs with non-Lipschitz coefficients

$$dX_t \in -\mathcal{A}(X_t)dt + b(t, X_t, \mathscr{L}_{X_t})dt + \sigma(t, X_t, \mathscr{L}_{X_t})dB_t, \tag{1.1}$$

where $\mathscr{L}_{X_t}$ stands for the probability distribution of X_t, $\mathcal{A} : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ is a multi-valued maximal monotone operator, the coefficients $b : [0,T] \times \mathbb{R}^d \times \mathcal{M}_2(\mathbb{R}^d) \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \times \mathcal{M}_2(\mathbb{R}^d) \to \mathbb{R}^{d\times m}$ are Borel measurable, and $\{B_t\}_{t\geq 0}$ is an m-dimensional Brownian motion defined on a filtered probability space. Compared with the usual McKean-Vlasov SDEs, i.e., $\mathcal{A} = 0$, most of difficulties towards (1.1) come from the high singularity of multi-valued maximal monotone operator $\mathcal{A}$ which is neither bounded nor continuous.

The rest of the paper is organised as follows. Section 2 recalls basic notations and introduces maximal monotone operators. Section 3 is denoted to the proof of our stochastic averaging principle which states that the solution of the averaged equation will converge to that of the concerned equation in the sense of the mean square.

2. Preliminaries

We use $|\cdot|$ and $\|\cdot\|$ for norms of vectors and matrices, respectively. Let $\langle\cdot,\cdot\rangle$ denote the scalar product in $\mathbb{R}^d$ and let $C(\mathbb{R}^d)$ be the collection of continuous functions on $\mathbb{R}^d$ and $C^2(\mathbb{R}^d)$ be the space of continuous functions on $\mathbb{R}^d$ which have continuous partial derivatives of order up to 2. Define the Banach space

$$C_\rho(\mathbb{R}^d) := \Big\{\varphi \in C(\mathbb{R}^d), \|\varphi\|_{C_\rho(\mathbb{R}^d)} := \sup_{x\in\mathbb{R}^d} \frac{|\varphi(x)|}{(1+|x|)^2} + \sup_{x\neq y} \frac{|\varphi(x)-\varphi(y)|}{|x-y|} < \infty\Big\}.$$

Let $\mathscr{B}(\mathbb{R}^d)$ be the Borel σ-algebra on $\mathbb{R}^d$ and $\mathcal{P}(\mathbb{R}^d)$ be the space of all probability measures defined on $\mathscr{B}(\mathbb{R}^d)$ carrying the usual topology of weak convergence. Let $\mathcal{M}_2(\mathbb{R}^d)$ be the set of probability measures on $\mathscr{B}(\mathbb{R}^d)$ with finite second order moments. That is,

$$\mathcal{M}_2(\mathbb{R}^d) := \Big\{\mu \in \mathcal{P}(\mathbb{R}^d) : \|\mu\|_2^2 := \int_{\mathbb{R}^d} |x|^2 \mu(dx) < \infty\Big\}.$$

Define the following metric on $\mathcal{M}_2(\mathbb{R}^d)$:

$$\rho(\mu,\nu) := \sup_{\|\varphi\|_{C_\rho(\mathbb{R}^d)}\leq 1} \Big|\int_{\mathbb{R}^d} \varphi(x)\mu(dx) - \int_{\mathbb{R}^d} \varphi(x)\nu(dx)\Big|, \quad \mu,\nu \in \mathcal{M}_2(\mathbb{R}^d).$$

It is clear that $(\mathcal{M}_2(\mathbb{R}^d), \rho)$ is a complete metric space and $\rho(\mathscr{L}_X, \mathscr{L}_Y) \leq (\mathbb{E}|X-Y|^2)^{\frac{1}{2}}$, and further that the convergence with respect to the metric ρ is equivalent to the weak convergence (see e.g. [3]). Denote by $2^{\mathbb{R}^d}$ the set of all subsets of $\mathbb{R}^d$, a map $\mathcal{A} : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ is called a multi-valued operator on $\mathbb{R}^d$. Given such a multi-valued operator $\mathcal{A}$, we define

$$D(\mathcal{A}) = \{x \in \mathbb{R}^d : \mathcal{A}(x) \neq \emptyset\}, \quad Gr(\mathcal{A}) = \{(x,y) \in \mathbb{R}^{2d} : x \in D(\mathcal{A}), y \in \mathcal{A}(x)\}.$$

$\mathcal{A}^{-1}$ is defined by $y \in \mathcal{A}^{-1}(x) \Leftrightarrow x \in \mathcal{A}(y)$.

Definition 2.1. (i) A multi-valued operator $\mathcal{A}$ is called monotone if

$$\langle x_1 - x_2, y_1 - y_2\rangle \geq 0, \quad \forall (x_1,y_1), (x_2,y_2) \in Gr(\mathcal{A}).$$

(ii) A monotone operator $\mathcal{A}$ is called maximal monotone if and only if

$$(x_1,y_1) \in Gr(\mathcal{A}) \Leftrightarrow \langle y_1 - y_2, x_1 - x_2\rangle \geq 0, \quad \forall (x_2,y_2) \in Gr(\mathcal{A}).$$

Given $T > 0$. Let $\mathscr{V}_0$ be the set of all continuous functions $K : [0,T] \to \mathbb{R}^d$ with finite variations and $K_0 = 0$. For $K \in \mathscr{V}_0$ and $s \in [0,T]$, we shall use $|K|_0^s$ to denote the variation of K on $[0,s]$. Set

$$\mathscr{A} := \Big\{(X,K) : X \in C([0,T], \overline{D(\mathcal{A})}), K \in \mathscr{V}_0 \ \ and \ \langle X_t - x, dK_t - ydt\rangle \geq 0 \, for\, any\, (x,y) \in Gr(\mathcal{A})\Big\}.$$

For more details about the maximal monotone operator, we refer to ([1], [10])

Lemma 2.2. *For $X \in C([0,T], \overline{D(\mathcal{A})})$ and $K \in \mathscr{V}_0$, the following statements are equivalent:*

(i) $(X,K) \in \mathscr{A}$*;*

(ii) *For any $(x,y) \in C([0,T], \mathbb{R}^d)$, $(x_t, y_t) \in Gr(\mathcal{A})$, it holds that $\langle X_t - x_t, dK_t - y_t dt\rangle \geq 0$;*

(iii) *For any $(X^{'}, K^{'}) \in \mathscr{A}$, it holds that $\langle X_t - X_t^{'}, dK_t - dK_t^{'}\rangle \geq 0$.*

Lemma 2.3. *Assume that $Int(D(\mathcal{A})) \neq \emptyset$, where $Int(D(\mathcal{A}))$ denotes the interior of the set $D(\mathcal{A})$. For any $a \in Int(D(\mathcal{A}))$, there exists constants $\gamma_1 > 0$, and $\gamma_2, \gamma_2 \geq 0$ such that for any $(X,K) \in \mathscr{A}$ and $0 \leq s < t \leq T$, $\int_s^t \langle X_r - a, dK_r\rangle \geq \gamma_1 |K|_s^t - \gamma_2 \int_s^t |X_r - a| dr - \gamma_3(t-s)$.*

Lemma 2.4. *Assume that $\{K^n, n \in \mathbb{N}\} \subset \mathscr{V}_0$ converges to some K in $C([0,T];\mathbb{R}^d)$ and $\sup_{n\in\mathbb{N}} |K^n|_0^T < \infty$. Then $K \in \mathscr{V}_0$, and $\lim_{n\to\infty} \int_0^T \langle X_s^n, dK_s^n\rangle = \int_0^T \langle X_s, dK_s\rangle$, where the sequence $\{X^n\} \subset C([0,T], \mathbb{R}^d)$ converges to some X in $C([0,T], \mathbb{R}^d)$.*

Definition 2.5. We say that Eq.(1.1) admits a strong solution with the initial value ξ if there exists a pair of adapted processes $(X_\cdot, K_\cdot)$ on a filtered probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\in[0,T]}, \mathbb{P})$ such that

(i) $\mathbb{P}(X_0 = \xi) = 1$, (ii) $X_t \in \mathscr{F}_t$, where $\mathscr{F}_t$ stands for the σ-field filtration generated by$(B_s)_{s\leq t}$ and ξ,

(iii) $(X_\cdot(\omega), K_\cdot(\omega)) \in \mathscr{A}$ a.s. $\mathbb{P}$,

(iv) it holds that $\mathbb{P}\Big\{ \int_0^T (|b(s,X_s,\mathscr{L}_{X_s})| + \|\sigma(s,X_s,\mathscr{L}_{X_s})\|^2) ds < +\infty \Big\} = 1$, and

$$X_t = \xi - K_t + \int_0^t b(s,X_s,\mathscr{L}_{X_s}) ds + \int_0^t \sigma(s,X_s,\mathscr{L}_{X_s}) dB_s, \quad t \in [0,T].$$

Utilising the Carathéodory approximation technique and Bihari's inequality, one can establish the existence and uniqueness theorem for the solution of multi-valued McKean-Vlasov SDEs (1.1) under the following Assumption.

Assumption 2.6. *For any $x, y \in \mathbb{R}^d$, $\mu, \nu \in \mathcal{M}_2(\mathbb{R}^d)$ and $t \in [0,T]$, there exists an increasing bounded function $L : [0,\infty) \to (0,\infty)$ such that*

$$|b(t,x,\mu) - b(t,y,\nu)|^2 + \|\sigma(t,x,\mu) - \sigma(t,y,\nu)\|^2 \leq L(t)\kappa(|x-y|^2 + \rho^2(\mu,\nu)),$$

and $|b(t,0,\delta_0)|^2 + \|\sigma(t,0,\delta_0)\|^2 \leq L(t)$, where $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ is a concave non-decreasing function such that $\kappa(0) = 0$, $\int_{0+} \frac{1}{\kappa(u)} du = +\infty$.

Remark 2.7. By Assumption 2.6, it holds that for $t \in [0,T]$, $x \in \mathbb{R}^d$, $\mu \in \mathcal{M}_2(\mathbb{R}^d)$

$$|b(t,x,\mu)|^2 + \|\sigma(t,x,\mu)\|^2 \leq 2L(t)[\kappa(|x|^2 + \rho^2(\mu,\delta_0)) + 1].$$

3. Stochastic averaging principle

We aim to derive a stochastic averaging principle for the following multi-valued McKean-Vlasov SDE

$$X_t^\epsilon = \xi - \epsilon K_t + \epsilon \int_0^t b(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) ds + \sqrt{\epsilon} \int_0^t \sigma(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) dB_s \tag{3.1}$$

with the initial value $X_0^\epsilon = \xi$. Here the coefficients b and σ have the same conditions as in Assumption 2.6 and $\epsilon \in [0,\epsilon_0]$ is a positive small parameter with ϵ_0 is a fixed number. Thus, (3.1) has a unique solution $(X_t^\epsilon, K_t), t \in$

$[0,T]$. Moreover, this solution satisfies $\mathbb{E}(\sup_{0\le t\le T}|X_t^\epsilon|^2)\le C_\epsilon$, and $\mathbb{E}|X_t^\epsilon - X_s^\epsilon|^2 \le C_\epsilon(t-s)$, where C_ϵ is a positive constant depend on ϵ.

Our objective is to show that the solution $(X_t^\epsilon, K_t), t\in[0,T]$ could be approximated in certain sense by the solution $(Y_t^\epsilon, \bar{K}_t), t\in[0,T]$ of the following averaged equation

$$Y_t^\epsilon = \xi - \epsilon\bar{K}_t + \epsilon\int_0^t \bar{b}(Y_s^\epsilon, \mathscr{L}_{Y_s^\epsilon})ds + \sqrt{\epsilon}\int_0^t \bar{\sigma}(Y_s^\epsilon, \mathscr{L}_{Y_s^\epsilon})dB_s, \tag{3.2}$$

where $\bar{b}:\mathbb{R}^d\times\mathcal{M}_2(\mathbb{R}^d)\to\mathbb{R}^d$ and $\bar{\sigma}:\mathbb{R}^d\times\mathcal{M}_2(\mathbb{R}^d)\to\mathbb{R}^{d\times m}$ are Borel measurable functions.

Assumption 3.1. *(Averaging condition) For any $x\in\mathbb{R}^d$, $\mu\in\mathcal{M}_2(\mathbb{R}^d)$ and $T_1>0$, there exist two positive bounded functions $\psi_i:(0,\infty)\to(0,\infty), i=1,2$ with $\lim_{T_1\to\infty}\psi_i(T_1)=0$, such that*

$$\frac{1}{T_1}\int_0^{T_1}|b(s,x,\mu)-\bar{b}(x,\mu)|^2ds \le \psi_1(T_1)(1+|x|^2+\rho^2(\mu,\delta_0)),$$
$$\frac{1}{T_1}\int_0^{T_1}\|\sigma(s,x,\mu)-\bar{\sigma}(x,\mu)\|^2ds \le \psi_2(T_1)(1+|x|^2+\rho^2(\mu,\delta_0)).$$

Remark 3.2. For any $x,y\in\mathbb{R}^d$, $\mu,\nu\in\mathcal{M}_2(\mathbb{R}^d)$ and $T_1>0$, we have

$$\begin{aligned}&|\bar{b}(x,\mu)-\bar{b}(y,\nu)|^2+\|\bar{\sigma}(x,\mu)-\bar{\sigma}(y,\nu)\|^2\\&\le 3(\psi_1(T_1)+\psi_2(T_1))(2+|x|^2+|y|^2+\rho^2(\mu,\delta_0)+\rho^2(\nu,\delta_0))+3L(T_1)\kappa(|x-y|^2+\rho^2(\mu,\nu)),\end{aligned}$$

$$|\bar{b}(0,\delta_0)|^2+\|\bar{\sigma}(0,\delta_0)\|^2\le 2(\psi_1(T_1)+\psi_2(T_1))+2L(T_1).$$

Taking $T_1\to\infty$, because L(t) is bounded, there exists a constant M, such that

$$|\bar{b}(x,\mu)-\bar{b}(y,\nu)|^2+\|\bar{\sigma}(x,\mu)-\bar{\sigma}(y,\nu)\|^2\le M\kappa(|x-y|^2+\rho^2(\mu,\nu)),\ \ |\bar{b}(0,\delta_0)|^2+\|\bar{\sigma}(0,\delta_0)\|^2\le M.$$

Thus, the coefficients $\bar{b},\bar{\sigma}$ satisfy the Assumptions 2.6. Therefore, there is a unique solution $(Y_\cdot^\epsilon,\bar{K}_\cdot^\epsilon)$ to the averaged equation (3.2). Moreover, this solution satisfies $\mathbb{E}(\sup_{0\le t\le T}|Y_t^\epsilon|^2)\le C_\epsilon$.

Theorem 3.3. *Suppose that $Int(D(\mathcal{A}))\neq\emptyset$ and $\mathbb{E}|\xi|^2<+\infty$. Then under Assumption 2.6 and 3.1, the averaging principle holds $\lim_{\epsilon\to 0}\mathbb{E}(\sup_{0\le t\le T}|X_t^\epsilon-Y_t^\epsilon|^2)=0$.*

Proof. From (3.1) and (3.2), we have

$$\begin{aligned}X_t^\epsilon-Y_t^\epsilon=-\epsilon[K_t-\bar{K}_t]&+\epsilon\int_0^t[b(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon})-\bar{b}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})]ds\\&+\sqrt{\epsilon}\int_0^t[\sigma(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon})-\bar{\sigma}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})]dB_s.\end{aligned}$$

By Itô formula, we have

$$\begin{aligned}|X_t^\epsilon-Y_t^\epsilon|^2&=-2\epsilon\int_0^t\langle X_s^\epsilon-Y_s^\epsilon,dK_s-d\bar{K}_s\rangle+2\epsilon\int_0^t\langle X_s^\epsilon-Y_s^\epsilon,b(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon})-\bar{b}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})\rangle ds\\&+2\sqrt{\epsilon}\int_0^t\langle X_s^\epsilon-Y_s^\epsilon,(\sigma(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon})-\bar{\sigma}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon}))dB_s\rangle+\epsilon\int_0^t\|\sigma(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon})-\bar{\sigma}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})\|^2ds.\end{aligned}$$

Then, according to Definition 2.5 and Lemma 2.2, it is immediate to conclude that $\int_0^t\langle X_s^\epsilon-Y_s^\epsilon,dK_s-d\bar{K}_s\rangle\ge 0$.

Taking the expectation on both sides, it follows that for any $t \in [0,T]$,

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t}|X_s^\epsilon - Y_s^\epsilon|^2\Big) \le\ & 2\epsilon\mathbb{E}\int_0^t |X_s^\epsilon - Y_s^\epsilon||b(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) - \bar{b}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})|ds\\ &+ \epsilon\mathbb{E}\int_0^t \|\sigma(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) - \bar{\sigma}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})\|^2 ds\\ &+ 2\sqrt{\epsilon}\mathbb{E}\Big(\sup_{0\le s\le t}\int_0^s \langle X_r^\epsilon - Y_r^\epsilon, (\sigma(r,X_r^\epsilon,\mathscr{L}_{X_r^\epsilon}) - \bar{\sigma}(Y_r^\epsilon,\mathscr{L}_{Y_r^\epsilon}))dB_r\rangle\Big).\end{aligned} \tag{3.3}$$

By the basic inequality $2|a||b| \le |a|^2 + |b|^2$, we can obtain

$$\begin{aligned}&2\epsilon\mathbb{E}\int_0^t |X_s^\epsilon - Y_s^\epsilon||b(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) - \bar{b}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})|ds\\ &\le \epsilon\mathbb{E}\int_0^t |X_s^\epsilon - Y_s^\epsilon|^2 ds + \epsilon\mathbb{E}\int_0^t |b(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) - \bar{b}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})|^2 ds.\end{aligned} \tag{3.4}$$

Burkholder-Davis-Gundy's inequality implies that

$$\begin{aligned}&2\sqrt{\epsilon}\mathbb{E}\Big(\sup_{0\le s\le t}\int_0^s \langle X_r^\epsilon - Y_r^\epsilon, (\sigma(r,X_r^\epsilon,\mathscr{L}_{X_r^\epsilon}) - \bar{\sigma}(Y_r^\epsilon,\mathscr{L}_{Y_r^\epsilon}))dB_r\rangle\Big)\\ &\le C\sqrt{\epsilon}\mathbb{E}\Big[\int_0^t |X_s^\epsilon - Y_s^\epsilon|^2\|\sigma(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) - \bar{\sigma}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})\|^2 ds\Big]^{\frac{1}{2}}\\ &\le \frac{1}{2}\mathbb{E}\Big(\sup_{0\le s\le t}|X_s^\epsilon - Y_s^\epsilon|^2\Big) + C\epsilon\mathbb{E}\int_0^t \|\sigma(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) - \bar{\sigma}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})\|^2 ds.\end{aligned} \tag{3.5}$$

Combing with (3.3)-(3.5), we obtain

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t}|X_s^\epsilon - Y_s^\epsilon|^2\Big) \le\ & 2\epsilon\mathbb{E}\int_0^t |X_s^\epsilon - Y_s^\epsilon|^2 ds + 2\epsilon\mathbb{E}\int_0^t |b(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) - \bar{b}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})|^2 ds\\ &+ C\epsilon\mathbb{E}\int_0^t \|\sigma(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) - \bar{\sigma}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})\|^2 ds.\end{aligned} \tag{3.6}$$

By Assumption 2.6, Assumption 3.1 and Remark 3.2, we have

$$\begin{aligned}&\mathbb{E}\Big(\sup_{0\le s\le t}|X_s^\epsilon - Y_s^\epsilon|^2\Big) \le 2\epsilon\mathbb{E}\int_0^t |X_s^\epsilon - Y_s^\epsilon|^2 ds + 4\epsilon\mathbb{E}\int_0^t |b(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) - b(s,Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})|^2 ds\\ &\quad + C\epsilon\mathbb{E}\int_0^t \|\sigma(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}) - \sigma(s,Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})\|^2 ds + 4\epsilon\mathbb{E}\int_0^t |b(s,Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon}) - \bar{b}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})|^2 ds\\ &\quad + C\epsilon\mathbb{E}\int_0^t \|\sigma(s,Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon}) - \bar{\sigma}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon})\|^2 ds\\ &\le 2\epsilon\mathbb{E}\int_0^t |X_s^\epsilon - Y_s^\epsilon|^2 ds + C\epsilon\mathbb{E}\int_0^t L(s)\kappa(|X_s^\epsilon - Y_s^\epsilon|^2 + \rho^2(\mathscr{L}_{X_s^\epsilon},\mathscr{L}_{Y_s^\epsilon}))ds\\ &\quad +12\epsilon\mathbb{E}\int_0^t |b(s,Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon}) - b(s,0,\delta_0)|^2 ds + 12\epsilon\mathbb{E}\int_0^t |b(s,0,\delta_0) - \bar{b}(0,\delta_0)|^2 ds\\ &\quad +12\epsilon\mathbb{E}\int_0^t |\bar{b}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon}) - \bar{b}(0,\delta_0)|^2 ds + C\epsilon\mathbb{E}\int_0^t \|\sigma(s,Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon}) - \sigma(s,0,\delta_0)\|^2 ds\\ &\quad +C\epsilon\mathbb{E}\int_0^t \|\sigma(s,0,\delta_0) - \bar{\sigma}(0,\delta_0)\|^2 ds + C\epsilon\mathbb{E}\int_0^t \|\bar{\sigma}(Y_s^\epsilon,\mathscr{L}_{Y_s^\epsilon}) - \bar{\sigma}(0,\delta_0)\|^2 ds\\ &\le 2\epsilon\mathbb{E}\int_0^t |X_s^\epsilon - Y_s^\epsilon|^2 ds + C\epsilon\int_0^t L(s)\kappa(2\mathbb{E}|X_s^\epsilon - Y_s^\epsilon|^2)ds + C\epsilon\int_0^t (L(s)+M)\kappa(2\mathbb{E}|Y_s^\epsilon|^2)ds\\ &\quad +C\epsilon t(\psi_1(t) + \psi_2(t)).\end{aligned}$$

Letting $\gamma(x)=\kappa(x)+x$, we get

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t}|X_s^\epsilon-Y_s^\epsilon|^2\Big)\le C\epsilon\int_0^t(L(s)+1)\gamma(2\mathbb{E}(\sup_{0\le r\le s}|X_r^\epsilon-Y_r^\epsilon|^2))ds\\+C\epsilon\int_0^t(L(s)+M)\kappa(2\mathbb{E}(\sup_{0\le r\le s}|Y_r^\epsilon|^2))ds+C\epsilon t(\psi_1(t)+\psi_2(t)).\end{aligned}\tag{3.7}$$

Given that $\kappa(\cdot)$ is concave and increasing, there must exist a positive number a such that

$$\kappa(x)\le a(1+x).$$

Hence, we have

$$\begin{aligned}\mathbb{E}\Big(\sup_{0\le s\le t}|X_s^\epsilon-Y_s^\epsilon|^2\Big)\le C\epsilon\int_0^t(L(s)+1)\gamma(2\mathbb{E}(\sup_{0\le r\le s}|X_r^\epsilon-Y_r^\epsilon|^2))ds\\+C\epsilon t(L(T)+M)(1+2\mathbb{E}(\sup_{0\le r\le s}|Y_r^\epsilon|^2))+C\epsilon t(\psi_1(t)+\psi_2(t)).\end{aligned}$$

Remark 3.2 and the boundness of $L(t),\psi_i(t),i=1,2$ yield that

$$\mathbb{E}\Big(\sup_{0\le s\le t}|X_s^\epsilon-Y_s^\epsilon|^2\Big)\le C\epsilon\int_0^t(L(s)+1)\gamma(2\mathbb{E}(\sup_{0\le r\le s}|X_r^\epsilon-Y_r^\epsilon|^2))ds+C\epsilon t.$$

Obviously, $\gamma(x)$ is nondecreasing function on $\mathbb{R}_+$ and $\gamma(0)=0$. Setting $G(t)=\int_1^t\frac{ds}{\gamma(s)}$, it follows from Bihari's inequality that $\mathbb{E}\Big(\sup_{0\le s\le t}|X_s^\epsilon-Y_s^\epsilon|^2\Big)\le\frac{1}{2}G^{-1}(G(C\epsilon t)+C\epsilon(L(T)+1)T)$. Noting that $C\epsilon t\to 0$ as $\epsilon\to 0$. Recalling the condition $\int_{0+}\frac{ds}{\gamma(s)}=+\infty$, we can conclude that $G(C\epsilon t)+C\epsilon(L(T)+1)T\to-\infty,\quad \epsilon\to 0$. On the other hand, because G is a strictly increasing function, then we obtain that G has an inverse function which is strictly increasing and $G^{-1}(-\infty)=0$. That is $G^{-1}(G(C\epsilon t)+C\epsilon(L(T)+1)T)\to 0,\quad \epsilon\to 0$. Consequently, we have $\lim_{\epsilon\to 0}\mathbb{E}\Big(\sup_{0\le s\le t}|X_s^\epsilon-Y_s^\epsilon|^2\Big)=0$. This completes the proof. □

Note that we can obtain the order of convergence, that is, for any small number $\delta_1>0$, there exist $L>0$, $\beta\in(0,1)$ and $\epsilon_1\in(0,\epsilon_0]$ such that for all $\epsilon\in(0,\epsilon_1]$, $\mathbb{E}(\sup_{t\in[0,L\epsilon^{\frac{1}{2}-\beta}]}|X_t^\epsilon-Y_t^\epsilon|^2)\le\delta_1$. Here we have established our main result in the sense of L^2 convergence. A very interesting question is that whether one can derive such kind of result in the sense of L^p convergence for $p>2$. We plan to address this problem in our forthcoming work.

Acknowledgements The authors would like to thank the reviewers and Professor Alan Tucker (Editor-in-Chief) for their constructive and insightful comments which have led to improve the presentation of the paper.

References

[1] E. Cépa. Problème de Skorohod multivoque. *Ann. Probab*. **26** (1998) 500-532.

[2] Z-Q, Chen, J. Wu. Averaging principle for stochastic variational inequalities with application to PDEs with nonlinear Neumann conditions. *J. Differential Equations.* **328** (2022) 157-201.

[3] J. Gong, H. Qiao. The stability for multivalued McKean-Vlasov SDEs with non-Lipschitz coefficients. https://arxiv.org/abs/2106.12080.

[4] R. Guo, B. Pei. Stochastic averaging principles for multi-valued stochastic differential equations driven by Poisson point processes. *Stoch. Anal. Appl*. **36** (2018) 751-766.

[5] R. Khasminskii. On the principle of averaging the Itô stochastic differential equations. *Kibernetika.* **4** (1968) 260-279.

[6] W. Mao, L. Hu, S. You and X. Mao. The averaging method for multivalued SDEs with jumps and non-Lipschitz coefficients. *Discrete Contin. Dyn. Syst. Ser. B*. **24** (2019) 4937-4954.

[7] L. Ngoran and N. Modeste. Averaging principle for multivalued stochastic differential equations. *Random Oper. Stoch. Equ.* **9** (2001) 399-407.

[8] J. Ren, S. Xu and X. Zhang. Large deviations for multivalued stochastic differential equations. *J. Theoret. Probab*. **23** (2010) 1142-1156.

[9] J. Xu, J. Liu. An averaging principle for multivalued stochastic differential equations. *Stoch. Anal. Appl*. **32** (2014) 962-974.

[10] X. Zhang. Skorohod problem and multivalued stochastic evolution equations in Banach spaces. *Bull. Sci. Math.* **131** (2007) 175-217.