STOCHASTIC AVERAGING PRINCIPLE FOR DISTRIBUTION DEPENDENT STOCHASTIC DIFFERENTIAL EQUATIONS*

GUANGJUN SHEN[†], JIE SONG[†], JIANG-LUN WU[‡,§]

[†] *Department of Mathematics, Anhui Normal University, Wuhu 241000, China*
E-mails: gjshen@163.com (G. Shen); Songjieahnu2596@163.com(J. Song)
and
[‡] *Department of Mathematics, Computational Foundry, Swansea University*
Swansea, SA1 8EN, UK
E-mail: j.l.wu@swansea.ac.uk

Abstract. Due to the intrinsic link with (kinetic) nonlinear Fokker-Planck equations and many diverse applications, distribution dependent stochastic differential equations have been investigated intensively in recent years. The appearance of the probability distributions (or laws) of the random variables of solutions in the coefficients is a distinct feature of distribution dependent stochastic differential equations. In this paper, under certain averaging conditions, we establish a stochastic averaging principle for distribution dependent stochastic differential equations.

1. Introduction

In the seminal papers [4, 5], Kac proposed the "propagation of chaos" of mean field particle systems in order to study nonlinear PDEs in Vlasov's kinetic theory. This motivated McKean [9] to study nonlinear Fokker-Planck equations utilising stochastic differential equations with distribution dependent drift coefficients. In general, nonlinear Fokker-Planck equations can be characterised by distribution dependent stochastic differential equations (DDSDEs for short), which are also named as McKean-Vlasov SDEs or mean field SDEs. A distinct feature of such systems is the appearance of probability laws in the coefficients of the resulting equations. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a given complete, filtered probability space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions and let $d, m \in \mathbb{N}$ be fixed. We use the following standard notations: $\mathbb{E}$ is used for the expectation with respect to $\mathbb{P}, \mathbb{R}^d$ denotes the d-dimensional Euclidean space and $|x|$ stands for the Euclidean norm of a vector $x \in \mathbb{R}^d$. For a matrix A, we denote the Frobenius norm by $\|A\| = \sqrt{\mathrm{tr}[AA^T]}$. In this paper, we are concerned with the following DDSDEs:

$$dX^\epsilon(t) = f\big(\frac{t}{\epsilon}, X^\epsilon(t), \mathscr{L}(X^\epsilon(t))\big)dt + g\big(X^\epsilon(t), \mathscr{L}(X^\epsilon(t))\big)dB(t), 0 < t \leq T \tag{1.1}$$

$X^\epsilon(0) = x_0$, with the sufficiently small parameter $0 < \epsilon \ll 1$, where $B(t)$ is an m-dimensional $\{\mathcal{F}_t\}$-Browinan motion defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}), \mathscr{L}(X^\epsilon(t))$ stands for the probability law or distribution of the random variable $X^\epsilon(t)$, $f : [0, \infty) \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \mapsto \mathbb{R}^d$ and $g : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \mapsto \mathbb{R}^{d\times m}$ are Borel measurable functions, $\mathcal{P}(\mathbb{R}^d)$ denotes the space of all probability measures on $\mathbb{R}^d$ equipped with the weak topology, the initial data x_0 is an $\mathcal{F}_0$-random variable satisfying $\mathbb{E}|x_0|^\beta < \infty$ for any $\beta > 0$.

[§] Corresponding author.

2010 *Mathematics Subject Classification.* 60H10; 34C29; 35Q83.

Key words and phrases. Stochastic averaging principle; distribution dependent stochastic differential equations; Wasserstein distance.

Recently, there has been an increasing interest to study existence and uniqueness results for the solutions of DDSDEs. Wang [13] established strong well-posedness of DDSDEs with one-sided Lipschitz continuous drift and Lipschitz-continuous dispersion coefficients. Under integrability conditions on distribution dependent coefficients, Huang and Wang [3] obtained the existence and uniqueness for DDSDEs with non-degenerate noise. Li et al [7] studied existence and uniqueness of solutions to McKean-Vlasov SDEs under one-sided local Lipschitz condition on the drift and local Lipschitz condition on the diffusion coefficient with respect to the state variable. Mehri and Stannat [10] proposed a Lyapunov-type approach to solve the problem of existence and uniqueness of general law-dependent SDEs.

On the other hand, the averaging principle, initiated by Khasminskii in [6], is a very efficient and important tool in study of SDEs for modelling problems arising in many practical research areas. It in fact provides a powerful tool for simplifying dynamical systems, and for obtaining approximating solutions to differential equations. The averaging principle enables one to study complex equations with related averaging equations, which paves a convenient and easy way to study many important properties (see, e.g., [14], [1], [8], [11], [2], [12]). Although there exist many investigations in the literature devoted to studying stochastic averaging principle for SDEs driven by Brownian motion, fractional Brownian motion, Lévy processes as well as more general stochastic measures inducing semimartingales, etc., as we know, there is not any consideration of averaging principle for DDSDEs. Moreover, due to their distribution dependent feature, they are potentially useful and important for modelling complex systems in diverse areas of applications. Comparing to the classical SDEs driven by Brownian motion, fractional Brownian motion, and Lévy processes, the DDSDEs are much more complex, therefore, a stochastic averaging principle for such SDEs is naturally interesting and would also be very useful. This motivates us to carry out the present paper, aiming to establish a stochastic averaging principle for the DDSDEs.

The rest of the paper is organised as follows. In Section 2, we present some preliminaries and our assumptions for this paper. In Section 3, we will prove an approximation theorem as an averaging principle for the solutions of the considered DDSDEs.

2. Preliminaries

In this section, we briefly give preliminaries and assumptions which will be used in the sequel. For technical reasons, we will work on the following subspace of $\mathcal{P}(\mathbb{R}^d)$ for any fixed $\theta \in [2,\infty)$, $\mathcal{P}_\theta(\mathbb{R}^d) := \left\{\mu \in \mathcal{P}(\mathbb{R}^d) : \mu(|\cdot|^\theta) := \int_{\mathbb{R}^d} |x|^\theta \mu(dx) < \infty\right\}$, which is a Polish space under the L^θ-Wasserstein distance

$$\mathbb{W}_\theta(\mu_1,\mu_2) := \inf_{\pi\in\mathscr{C}(\mu_1,\mu_2)} \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^\theta \pi(dx,dy)\right)^{\frac{1}{\theta}}, \mu_1,\mu_2 \in \mathcal{P}_\theta(\mathbb{R}^d),$$

where $\mathscr{C}(\mu_1,\mu_2)$ is the set of all couplings for μ_1 and μ_2.

Note that for any $x \in \mathbb{R}^d$, the Dirac measure δ_x belongs to $\mathcal{P}_\theta(\mathbb{R}^d)$ for any $\theta \in [2,\infty)$ and if $\mu_1 = \mathscr{L}(X)$, $\mu_2 = \mathscr{L}(Y)$ are the corresponding distributions of random variables X and Y respectively, then $(\mathbb{W}_\theta(\mu_1,\mu_2))^\theta \le \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^\theta \mathscr{L}((X,Y))(dx,dy) = \mathbb{E}|X-Y|^\theta$, in which $\mathscr{L}((X,Y))$ represents the joint distribution of the random pair (X,Y). Next, we impose the following conditions on the coefficients of (1.1).

Assumption 2.1. (a) *There exists a positive constant L_1 such that $|f(t,x,\mu)-f(t,y,\mu)| \le L_1|x-y|$, for any $t \in [0,T]$, $x,y \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.*

(b) *There exists a positive constant L_2 such that $||g(x,\mu)-g(y,\mu)|| \le L_2|x-y|$, for any $x,y\in\mathbb{R}^d$ and $\mu\in\mathcal{P}_2(\mathbb{R}^d)$.*
(c) *There exists a positive constant K_1 such that $|f(t,x,\mu)-f(t,x,\nu)| \vee ||g(x,\mu)-g(x,\nu)|| \le K_1\mathbb{W}_2(\mu,\nu)$, for any $t\in[0,T]$, $x\in\mathbb{R}^d$ and $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d)$.*

Assumption 2.2. *There exists a positive constant K_2 such that for any $t\in[0,T]$, $x\in\mathbb{R}^d$ and $\mu\in\mathcal{P}_2(\mathbb{R}^d)$, $|f(t,x,\mu)| \vee ||g(x,\mu)|| \le K_2(1+|x|+\mathbb{W}_2(\mu,\delta_0))$, where δ_0 denotes the Dirac measure at 0.*

Remark 2.1. *One can conclude by Li et al [7] that under Assumptions 2.1–2.2, there exists a unique solution $X^\epsilon(t)\in L^2(\Omega;\mathbb{R}^d), t\in[0,T]$, to the equation (1.1) for any initial value $X^\epsilon(0)=x_0$ satisfying $\mathbb{E}|x_0|^\beta<\infty$ for any $\beta>0$. Moreover, for any $p\ge 2$, the solution fulfils*

$$\mathbb{E}\left[\sup_{0\le t\le T}|X^\epsilon(t)|^p\right] \le C_{p,\epsilon},\ \ \mathbb{E}|X^\epsilon(t)-X^\epsilon(s)|^2 \le C_{2,\epsilon}|t-s| \tag{2.1}$$

for some positive constants $C_{p,\epsilon}$ and $C_{2,\epsilon}$ depend on ϵ.

Throughout this paper, the letter C will denote a positive constant whose value may change in different occasions. We will write the dependence of constant on parameters explicitly if it is essential.

3. Stochastic averaging principle for DDSDEs

In this section, we aim to establish a stochastic averaging principle for the following DDSDE (i.e. the integral formulation of Equation (1.1))

$$X^\epsilon(t) = X^\epsilon(0) + \int_0^t f\left(\frac{s}{\epsilon}, X^\epsilon(s), \mathscr{L}(X^\epsilon(s))\right)ds + \int_0^t g(X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))dB(s), \tag{3.1}$$

where $X^\epsilon(0)=x_0\in\mathbb{R}^d$ is a random vector and $\epsilon\in(0,\epsilon_0]$ is a positive parameter with $\epsilon_0>0$ being fixed. According to Remark 2.1, Equation (3.1) has a unique solution $X^\epsilon(t), t\in[0,T]$. Our objective is to show that the solution $X^\epsilon(t), t\in[0,T]$, could be approximated in certain sense by the solutions $\bar{X}(t), t\in[0,T]$, of the following averaged equation

$$\bar{X}(t) = X^\epsilon(0) + \int_0^t \bar{f}(\bar{X}(s),\mathscr{L}(\bar{X}(s)))ds + \int_0^t g(\bar{X}(s),\mathscr{L}(\bar{X}(s)))dB(s), \tag{3.2}$$

where $\bar{f}:\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)\mapsto\mathbb{R}^d$ is Borel measurable function. Moreover, we assume the following condition hold.

Assumption 3.1. *(Averaging condition) There is a positive bounded function $\varphi:(0,\infty)\to(0,\infty)$ with $\lim_{t\to\infty}\varphi(t)=0$, such that for any $x\in\mathbb{R}^d$, $\mu\in\mathcal{P}_2(\mathbb{R}^d)$*

$$\sup_{t\ge 0}\left|\frac{1}{T}\int_t^{t+T}[f(s,x,\mu)-\bar{f}(x,\mu)]ds\right|^2 \le \varphi(T)(1+|x|^2).$$

Remark 3.1. (i) *Noting that*

$$\sup_{t\ge 0}\left|\frac{1}{T}\int_t^{t+T}[f(s,x,\mu)-\bar{f}(x,\mu)]ds\right|^2 \le \sup_{t\ge 0}\frac{1}{T}\int_t^{t+T}|f(s,x,\mu)-\bar{f}(x,\mu)|^2ds,$$

this shows that Assumption 3.1 is weaker than the following averaging condition (for example, [14]) $\sup_{t\ge 0}\frac{1}{T}\int_t^{t+T}|f(s,x,\mu)-\bar{f}(x,\mu)|^2ds \le \varphi(T)(1+|x|^2)$. Hence, we need to overcome the difficulties with the weaker condition to obtain the averaging principle for the concerned DDSDEs.

(ii) *for any $x, y \in \mathbb{R}^d$, and any $T > 0$, by Assumptions 2.1 and 3.1, we have*

$$\begin{aligned}
|\bar{f}(x,\mu) - \bar{f}(y,\mu)| &\leq \left|\frac{1}{T}\int_0^T [f(s,x,\mu) - \bar{f}(x,\mu)]ds\right| \\
&+ \left|\frac{1}{T}\int_0^T [f(s,y,\mu) - \bar{f}(y,\mu)]ds\right| + \left|\frac{1}{T}\int_0^T [f(s,x,\mu) - f(s,y,\mu)]ds\right| \\
&\leq \sqrt{\varphi(T)}\big(\sqrt{1+|x|^2} + \sqrt{1+|y|^2}\big) + L_1|x-y|.
\end{aligned} \tag{3.3}$$

Taking $T \to \infty$, we have $\bar{f}$ satisfies the condition (a) in Assumptions 2.1. Similarly, for any $x, y \in \mathbb{R}^d$, and any $T > 0$, we have

$$|\bar{f}(x,\mu) - \bar{f}(x,\nu)| \leq 2\sqrt{\varphi(T)}\big(\sqrt{1+|x|^2}\big) + K_1\mathbb{W}_2(\mu,\nu). \tag{3.4}$$

Taking $T \to \infty$, we have $\bar{f}$ satisfies the condition (c) in Assumptions 2.1.

On the other hand, by (3.3), (3.4) and

$$|\bar{f}(x,\mu)| \leq |\bar{f}(x,\mu) - \bar{f}(x,\delta_0)| + |\bar{f}(x,\delta_0) - \bar{f}(0,\delta_0)| + |\bar{f}(0,\delta_0)|,$$

we have $\bar{f}$ satisfies the Assumptions 2.2. Hence, there is a unique solution $\bar{X}(t)$ to the averaged equation (3.2).

Now we are in the position to present our main result.

Theorem 3.1. *Assume that $\mathbb{E}|X_0^\epsilon|^2 < +\infty$. Then, under Assumptions 2.1-2.2 and Assumption 3.1, the following averaging principle holds*

$$\lim_{\epsilon\to 0}\mathbb{E}\bigg(\sup_{0\leq t\leq T}|X^\epsilon(t) - \bar{X}(t)|^2\bigg) = 0. \tag{3.5}$$

We proceed by first proving the following proposition which is important in the proof of Theorem 3.1.

Proposition 3.1. *Suppose that Assumptions 2.1–2.2 and Assumption 3.1 hold, $\mathbb{E}|X_0^\epsilon|^2 < +\infty$. Then, we have*

$$\lim_{\epsilon\to 0}\mathbb{E}\bigg(\sup_{0\leq t\leq T}\Big|\int_0^t [f(\frac{s}{\epsilon}, X^\epsilon(s), \mathscr{L}(X^\epsilon(s))) - \bar{f}(X^\epsilon(s), \mathscr{L}(X^\epsilon(s)))]ds\Big|^2\bigg) = 0. \tag{3.6}$$

Proof. In order to simplify the calculation, let $\{t_1, t_2, \cdots, t_N\}$ be a partition of $[0,T]$, that is, $t_i = i\sqrt{\epsilon},\ 0 \leq i \leq N-1,\ 0 < T - t_{N-1} \leq \sqrt{\epsilon},\ t_N = T$. Then, it is easy to see that $T \leq N\sqrt{\epsilon} < T + \sqrt{\epsilon}$.

Denote $X_i := \int_{t_i}^{t_{i+1}}[f(\frac{s}{\epsilon}, X^\epsilon(s), \mathscr{L}(X^\epsilon(s))) - \bar{f}(X^\epsilon(s), \mathscr{L}(X^\epsilon(s)))]ds$. We have

$$\begin{aligned}
&\left|\int_0^t [f(\frac{s}{\epsilon}, X^\epsilon(s), \mathscr{L}(X^\epsilon(s))) - \bar{f}(X^\epsilon(s), \mathscr{L}(X^\epsilon(s)))]ds\right|^2 \\
&\leq\ N\left|\int_{[\frac{t}{\sqrt{\epsilon}}]\sqrt{\epsilon}}^t [f(\frac{s}{\epsilon}, X^\epsilon(s), \mathscr{L}(X^\epsilon(s))) - \bar{f}(X^\epsilon(s), \mathscr{L}(X^\epsilon(s)))]ds\right|^2 + N\sum_{i=0}^{N-2}|X_i|^2.
\end{aligned} \tag{3.7}$$

By Assumption 2.2, we have

$$\begin{aligned}
&\left|\int_{[\frac{t}{\sqrt{\epsilon}}]\sqrt{\epsilon}}^t [f(\frac{s}{\epsilon}, X^\epsilon(s), \mathscr{L}(X^\epsilon(s))) - \bar{f}(X^\epsilon(s), \mathscr{L}(X^\epsilon(s)))]ds\right|^2 \\
&\leq\ C|t - [\frac{t}{\sqrt{\epsilon}}]\sqrt{\epsilon}|^2(1 + \sup_{0\leq t\leq T}|X^\epsilon(t)|^2 + \sup_{0\leq t\leq T}\mathbb{E}|X^\epsilon(t)|^2) \\
&\leq\ C\epsilon(1 + \sup_{0\leq t\leq T}|X^\epsilon(t)|^2 + \sup_{0\leq t\leq T}\mathbb{E}|X^\epsilon(t)|^2),
\end{aligned} \tag{3.8}$$

where we use the fact $(\mathbb{W}_2(\mathscr{L}(X(t)),\delta_0))^2 \le \mathbb{E}|X(t)|^2$. By (2.1) and (3.8), we get

$$\mathbb{E}\Big(\sup_{0\le t\le T}\Big|\int_0^t[f(\tfrac{s}{\epsilon},X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))-\bar f(X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))]ds\Big|^2\Big)$$
$$\le C\epsilon N + N\mathbb{E}\sum_{i=0}^{N-2}|X_i|^2 \le C\sqrt{\epsilon}(T+\sqrt{\epsilon}) + N\sum_{i=0}^{N-2}\mathbb{E}|X_i|^2.$$

By Assumptions 2.1 and 3.1, we have

$$\begin{aligned}|X_i|^2 \le\ & 5\Big|\epsilon\int_{\frac{t_i}{\epsilon}}^{\frac{t_{i+1}}{\epsilon}}[f(s,X^\epsilon(t_i),\mathscr{L}(X^\epsilon(t_i)))-\bar f(X^\epsilon(t_i),\mathscr{L}(X^\epsilon(t_i)))]ds\Big|^2\\ &+10L_1^2\sqrt{\epsilon}\int_{t_i}^{t_{i+1}}|X^\epsilon(s)-X^\epsilon(t_i)|^2ds+10K_1^2\sqrt{\epsilon}\int_{t_i}^{t_{i+1}}\mathbb{E}|X^\epsilon(s)-X^\epsilon(t_i)|^2ds\\ \le\ & C\epsilon\varphi(\frac{1}{\sqrt{\epsilon}})(1+\sup_{0\le t\le T}|X^\epsilon(t)|^2)\\ &+10L_1^2\sqrt{\epsilon}\int_{t_i}^{t_{i+1}}|X^\epsilon(s)-X^\epsilon(t_i)|^2ds+10K_1^2\sqrt{\epsilon}\int_{t_i}^{t_{i+1}}\mathbb{E}|X^\epsilon(s)-X^\epsilon(t_i)|^2ds,\end{aligned}$$

Hence,

$$\begin{aligned}N\sum_{i=0}^{N-2}\mathbb{E}|X_i|^2 \le\ & C\epsilon N\sum_{i=0}^{N-2}\mathbb{E}\Big[\varphi(\frac{1}{\sqrt{\epsilon}})(1+\sup_{0\le t\le T}|X^\epsilon(t)|^2)\Big]\\ &+10(L_1^2+K_1^2)\sqrt{\epsilon}N\sum_{i=0}^{N-2}\int_{t_i}^{t_{i+1}}\mathbb{E}|X^\epsilon(s)-X^\epsilon(t_i)|^2ds\\ \le\ & C\epsilon N^2[\varphi(\frac{1}{\sqrt{\epsilon}})+\sqrt{\epsilon}] \le C(T+\sqrt{\epsilon})^2[\varphi(\frac{1}{\sqrt{\epsilon}})+\sqrt{\epsilon}],\end{aligned}\tag{3.9}$$

in the second inequality in (3.9) we have used inequality (2.1). Combining (3.9) into (3.7), we get

$$\begin{aligned}&\mathbb{E}\Big(\sup_{0\le t\le T}\Big|\int_0^t[f(\frac{s}{\epsilon},X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))-\bar f(X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))]ds\Big|^2\Big)\\ &\le C\sqrt{\epsilon}(T+\sqrt{\epsilon})+C(T+\sqrt{\epsilon})^2[\varphi(\frac{1}{\sqrt{\epsilon}})+\sqrt{\epsilon}]\to 0,\end{aligned}\tag{3.10}$$

as ϵ tends to zero. This completes the proof. □

The proof of Theorem 3.1. we have

$$\begin{aligned}&\mathbb{E}\sup_{0\le t\le T}|X^\epsilon(t)-\bar X(t)|^2\\ &\le 5\mathbb{E}\sup_{0\le t\le T}\Big|\int_0^t[f(\frac{s}{\epsilon},X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))-\bar f(X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))]ds\Big|^2\\ &\quad+5\mathbb{E}\sup_{0\le t\le T}\Big|\int_0^t[\bar f(X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))-\bar f(\bar X(s),\mathscr{L}(X^\epsilon(s)))]ds\Big|^2\\ &\quad+5\mathbb{E}\sup_{0\le t\le T}\Big|\int_0^t[\bar f(\bar X(s),\mathscr{L}(X^\epsilon(s)))-\bar f(\bar X(s),\mathscr{L}(\bar X(s)))]ds\Big|^2\\ &\quad+5\mathbb{E}\sup_{0\le t\le T}\Big|\int_0^t[g(X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))-g(\bar X(s),\mathscr{L}(X^\epsilon(s)))]dB(s)\Big|^2\\ &\quad+5\mathbb{E}\sup_{0\le t\le T}\Big|\int_0^t[g(\bar X(s),\mathscr{L}(X^\epsilon(s)))-g(\bar X(s),\mathscr{L}(\bar X(s)))]dB(s)\Big|^2.\end{aligned}\tag{3.11}$$

Using the Hölder inequality, the Doob martingale inequality and Assumption 2.1, for any $u \in [0,T]$ we have

(3.12)

$$\mathbb{E}\sup_{0\le t\le u}|X^\epsilon(t)-\bar{X}(t)|^2 \le 5\mathbb{E}\sup_{0\le t\le u}\left|\int_0^t [f(\frac{s}{\epsilon},X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))-\bar{f}(X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))]ds\right|^2$$
$$+5(L_1^2+K_1^2)T\int_0^u \mathbb{E}\sup_{0\le t\le s}|X^\epsilon(t)-\bar{X}(t)|^2 ds + 20(L_2^2+K_1^2)\int_0^u \mathbb{E}\sup_{0\le t\le s}|X^\epsilon(t)-\bar{X}(t)|^2 ds.$$

By the Gronwall inequality, we get

$$\begin{aligned}&\mathbb{E}\sup_{0\le t\le T}|X^\epsilon(t)-\bar{X}(t)|^2\\ &\le 5\mathbb{E}\left(\sup_{0\le t\le T}\left|\int_0^t [f(\frac{s}{\epsilon},X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))-\bar{f}(X^\epsilon(s),\mathscr{L}(X^\epsilon(s)))]ds\right|^2\right)\\ &\qquad\times e^{5[L_1^2T+4L_2^2+(4+T)K_1^2]T}.\end{aligned} \tag{3.13}$$

The proof is then completed by our Proposition 3.1 .

Remark 3.2. *By the Chebyshev-Markov inequality and Theorem 3.1, for any given number $\delta > 0$, we have*

$$\lim_{\epsilon\to 0}\mathbb{P}\left(\sup_{0\le t\le T}|X^\epsilon(t)-\bar{X}(t)|>\delta\right)\le \frac{1}{\delta^2}\lim_{\epsilon\to 0}\mathbb{E}\left(\sup_{0\le t\le T}|X^\epsilon(t)-\bar{X}(t)|^2\right)=0.$$

This implies the convergence in probability of the solutions $X^\epsilon(t)$ to the averaged solution $\bar{X}(t)$.

Acknowledgments 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. This research was supported by the National Natural Science Foundation of China (12071003).

References

[1] Z. Dong, X. Sun, H. Xiao and J. Zhai. Averaging principle for one dimensional stochastic Burgers equation. *J. Differential Equations.* **265** (2018) 4749-4797.

[2] Z. Guo, G. Lv, J. Wei. Averaging principle for stochastic differential equations under a weak condition. *Chaos.* **30** (2020) 123139.

[3] X. Huang, F.-Y. Wang. Distribution dependent SDEs with singular coefficients. *Stochastic Process. Appl.* **129** (2019) 4747–4770.

[4] M. Kac. Foudations of kinetic theory. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. 1954-1955. vol. III. pages 171-197. University of California Press. Berkeley and Los Angeles. 1956.

[5] M. Kac. Probability and Related Topics in the Physical Sciences. New York. 1959.

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

[7] Y. Li, X. Mao, Q. Song, F. Wu, G. Yin. Existence and uniqueness of solutions to McKean-Vlasov stochastic differential equations under local Lipschitz conditions of state variables. Submitted.

[8] D. Luo, Q. Zhu, Z. Luo. An averaging principle for stochastic fractional differential equations with time-delays. *Appl. Math. Lett.* **105** (2020) 106290.

[9] H. P. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA.* **56** (1966) 1907-1911.

[10] S. Mehri, W. Stannat. Weak solutions to Vlasov-McKean equations under Lyapunov-type conditions. *Stoch. Dyn.* **19** (2019) 1950042 (23 pages)

[11] B. Pei, Y. Xu and J.-L. Wu. Stochastic averaging for stochastic differential equations driven by fractional Brownian motion and standard Brownian motion. *Appl. Math. Lett.* **100** (2020) 106006.

[12] G. Shen, J.-L. Wu, X. Yin. Averaging principle for fractional heat equations driven by stochastic measures. *Appl. Math. Lett.* **106** (2020) 106404.

[13] F.-Y. Wang. Distribution dependent SDEs for Landau type equations. *Stochastic Process. Appl.* **128** (2018) 595–621.

[14] Y. Xu, J. Duan, W. Xu. An averaging principle for stochastic dynamical systems with Lévy noise. Phys. D **240** (2011) 1395-1401.