$L^p(p > 2)$ -strong convergence of an averaging principle for two-time-scales jump-diffusion stochastic differential equations *

Jie Xu, Yu Miao

College of Mathematics and Information Science, Henan Normal University, Xinxiang, Henan 453007, P.R. China

Abstract

Multiscale jump-diffusion stochastic differential equations arise as models for various complex systems. In this paper we prove the averaging principle for a class of two-time-scales jump-diffusion stochastic differential equations. Under suitable conditions, it is shown that the slow component $L^p(p > 2)$ -strongly converges to the solution of the corresponding averaging equation. As a product, the rate of the convergence is also presented.

Keywords: Stochastic averaging principle, L^p -strong convergence, Fast-slow SDEs with jumps

2010 MSC: 60H15, 70K65, 70K70

1. Introduction

The theory of averaging principles has played an important role in study of the qualitative behaviors for complex systems with multiscales. There is a big amount of literature devoted to the study of averaging principle for stochastic ordinary differential equations. Let us mention a few of them. In 1968, the averaging principle of Itô type stochastic ordinary equations

*This work was supported by China NSF Grant No.11471104, and IRT-STHN:14IRSTHN023, PhD research startup foundation of Henan Normal University 5101019170120, Youth Science Foundation of Henan Normal University 5101019279032.

*College of Mathematics and Information Science, Henan Normal University, Xinxiang, Henan 453007, P.R. China. Tel.:+8615516552382.

Email addresses: xujiescu@163.com (Jie Xu), yumiao@htu.cn (Yu Miao)

with different time-scales driven by Brownian noise was first established by Khasminskii in [18]. In 2003, Vanden-Eijnden in [13] proposed a numerical scheme for stochastic dynamical systems with multiple time-scales. In 2005, E, Liu and Vanden-Eijnden in [14] derived an estimate for the rate of L^1 strong convergence to the solution of a couple fast variable ordinary differential system. After that, Liu in [20] proposed a generalization of the averaging principle for the case where both the slow and fast component were described by an Itô diffusion process. In 2007, Givon in [15] established the L^2 (also called mean-square) strong averaging principle for a system in which slow and fast dynamics were driven by Brownian noises and Poisson noises under suitable assumptions. Furthermore, Liu in [21] extended the result of [15] to couple fast variable in slow component case. Besides, the mean-square type convergence was also treated in [10, 16, 17]. Quite recently, Wainrib in [25] developed an averaging principle for systems of slow-fast stochastic differential equations, where the fast variable drift was periodically modulated on a fast time-scale. The above work mainly discussed L^1 convergence for the two-scale diffusion process and L^2 convergence for the two-scale jump-diffusion process. We observe that L^1 convergence and L^2 convergence do not imply $L^p(p > 2)$ convergence in general. But, $L^p(p > 2)$ convergence may imply L^1 convergence and L^2 convergence. Generally speaking, this is an irreversible process. If $L^p(p > 2)$ convergence is proved, then we may obtain a much bigger degree of freedom for parameter q in study of $L^q(0 < q \leq p)$ convergence, which possesses a good robustness. This point can be applied in numerical computations. On the other hand, people often need to study the asymptotic estimate of higher order moments in statistics, finance and bond. So, $L^p(p > 2)$ convergence is desired by people. However, to the best of our knowledge, $L^p(p > 2)$ -strong convergence of the averaging principle for two-time-scales jump-diffusion stochastic ordinary differential equations has not been considered.

In this paper, based on above discussions, we will study the $L^p(p > 2)$ -strong convergence of the averaging principle for two-time-scales jump-diffusion stochastic differential equations:

$$\begin{cases} \dot{X}_t^\epsilon = a(X_t^\epsilon, Y_t^\epsilon) + b(X_t^\epsilon)\dot{B}_t + c(X_t^\epsilon)\dot{P}_t, \\ \dot{Y}_t^\epsilon = \frac{1}{\epsilon}f(X_t^\epsilon, Y_t^\epsilon) + \frac{1}{\sqrt{\epsilon}}g(X_t^\epsilon, Y_t^\epsilon)\dot{W}_t + h(X_t^\epsilon, Y_t^\epsilon)\dot{N}_t^\epsilon \end{cases} \quad (1.1)$$

for $t > 0$, with initial conditions $X_0^\epsilon = x_0$, $Y_0^\epsilon = y_0$, where X_t^ϵ is an

n -dimensional jump-diffusion process and Y_t^ϵ is an m -dimensional jump-diffusion process. The driving processes B_t and W_t are d_1, d_2 -dimensional independent Wiener processes, while P_t is a scalar simple Poisson process with intensity λ_1 , and N_t^ϵ is a scalar simple Poisson process with intensity $\frac{\lambda_2}{\epsilon}$. ϵ is a small positive parameter describing the ratio of time scale between the process X^ϵ and Y^ϵ . With this time scale the variable X^ϵ is referred as the *slow* component and Y^ϵ as the *fast* component. For the drift coefficients $a(u, v), f(u, v) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ and the diffusion coefficients $b(u) : \mathbb{R}^n \rightarrow \mathbb{R}^n \times \mathbb{R}^{d_1}$, $c(u) : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $g(u, v) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m \times \mathbb{R}^{d_2}$, $h(u, v) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$, we adopt the following assumptions throughout the paper.

(H1). The measurable functions a, b, c, f, g and h satisfy the global Lipschitz conditions, i.e., there is a positive constant L such that

$$\begin{aligned} & \|a(u_1, v_1) - a(u_2, v_2)\|^2 + \|b(u_1) - b(u_2)\|^2 + \|c(u_1) - c(u_2)\|^2 \\ & + \|f(u_1, v_1) - f(u_2, v_2)\|^2 + \|g(u_1, v_1) - g(u_2, v_2)\|^2 \\ & + \|h(u_1, v_1) - h(u_2, v_2)\|^2 \\ & \leq L(\|u_1 - u_2\|^2 + \|v_1 - v_2\|^2) \end{aligned} \quad (1.2)$$

for all $u_i \in \mathbb{R}^n, v_i \in \mathbb{R}^m, i = 1, 2$.

Remark 1.1. With the help of **(H1)**, it immediately follows that there is a positive constant K such that

$$\begin{aligned} & \|a(u, v)\|^2 + \|b(u)\|^2 + \|c(u)\|^2 + \|f(u, v)\|^2 + \|g(u, v)\|^2 + \|h(u, v)\|^2 \\ & \leq K(1 + \|u\|^2 + \|v\|^2) \end{aligned}$$

for $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$. Thus a, b, c, f, g and h satisfy the sublinear growth condition.

(H2). a, g and h are globally bounded.

(H3). There exist constants $\beta_1 > 0$ and $\beta_j \in \mathbb{R}, j = 2, 3, 4$, which are all independent of (u_1, v_1, v_2) , such that

$$v_1 \cdot f(u_1, v_1) \leq -\beta_1 \|v_1\|^2 + \beta_2, \quad (1.3)$$

$$(f(u_1, v_1) - f(u_1, v_2))(v_1 - v_2) \leq \beta_3 \|v_1 - v_2\|^2 \quad (1.4)$$

and

$$(h(u_1, v_1) - h(u_1, v_2))(v_1 - v_2) \leq \beta_4 \|v_1 - v_2\|^2 \quad (1.5)$$

for all $u_1 \in \mathbb{R}^n$ and $v_1, v_2 \in \mathbb{R}^m$.

(H4). $\eta := -(2\beta_3 + 2\lambda_2\beta_4 + C_g + \lambda_2 C_h) > 0$, here β_3 and β_4 take from (1.4) and (1.5), λ_2 is from N_t^ϵ with intensity λ_2/ϵ , C_g and C_h are the Lipschitz coefficients for g and h respectively, i.e.,

$$\|g(u_1, v_1) - g(u_2, v_2)\|^2 \leq C_g(\|u_1 - u_2\|^2 + \|v_1 - v_2\|^2) \quad (1.6)$$

and

$$\|h(u_1, v_1) - h(u_2, v_2)\|^2 \leq C_h(\|u_1 - u_2\|^2 + \|v_1 - v_2\|^2) \quad (1.7)$$

for all $u_1, u_2 \in \mathbb{R}^n, v_1, v_2 \in \mathbb{R}^m$.

(H5). There exists a constant $\gamma > 0$, which is independent of (u, v) , such that

$$v^T g(u, v) g^T(u, v) v \geq \gamma \|v\|^2 \quad (1.8)$$

for all $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$.

An example that satisfies the **(H1)**–**(H5)** is $a(u, v) = \frac{1}{1+(u+v)^2}$, $b(u) = e^{-u^2}$, $c(u) = \sin u$, $f(u, v) = -1.5(\lambda_2 + 1)v$, $g(u, v) = \frac{3 + \sin u + \sin v}{\sqrt{2}}$ and $h(u, v) = \frac{\sin u + \sin v}{\sqrt{2}}$.

Existence and uniqueness of the solution for stochastic system (1.1) can be guaranteed by **(H1)** and **(H2)**. **(H3)**–**(H5)** are interpreted as strong dissipative conditions which yield a unique invariant measure possessing exponentially mixing property for a Markov semigroup associated to the fast variable equation.

Here, we will prove that when $\epsilon \rightarrow 0$, X_t^ϵ $L^p(p > 2)$ -strongly converges to an effect dynamics equation of the form:

$$\begin{cases} \dot{\bar{X}}_t = \bar{a}(\bar{X}_t) + b(\bar{X}_t)\dot{B}_t + c(\bar{X}_t)\dot{P}_t, \\ \bar{X}_0 = x_0, \end{cases}$$

with

$$\bar{a}(x) = \int_{\mathbb{R}^m} a(x, y) \mu^x(dy), \quad x \in \mathbb{R}^n.$$

It is worth pointing out that under the Assumption A1- A4 which were used in [15], it seems not easy to extend L^2 -strong convergence of the averaging principle for two-time-scales jump-diffusion stochastic differential equations to $L^p(p > 2)$ -strong convergence case. The reason is that under the Assumption A1-A4 (cf.[15]), we can't prove the estimate (4.6) in this paper (see Lemma 4.2 below). In order to get the $L^p(p > 2)$ -strong convergence principle of two-time-scales jump-diffusion stochastic differential equations, we need to make some new assumptions on the drift and diffusion coefficients (see **(H2)** and **(H4)**). Our assumptions **(H1)**, (1.3) in **(H3)**, and **(H5)** are the same as Assumption A1 and Assumption A3-A4 (cf.[15]). In [15], Assumption A2 assumes that the functions f, g and h are of class C^∞ and have bounded derivatives of any order, $f(x, y)$ is a bounded function of x for all y , g and h are global bounded functions. We only keep a global boundedness of g and h in Assumption A2 (cf.[15]), and replace the other assumptions in Assumption A2 by **(H4)** and a global boundedness of a (see **(H2)**). The proof of X_t^ϵ $L^p(p > 2)$ -strong convergence to $\bar{X}_t$ can be realized by the ergodicity of the fast equation (see Section 2 below), the estimate of the differences between the process X_t^ϵ and $\hat{X}_t^\epsilon$ (see Lemma 3.5 below), and the estimate of the differences between the process $\hat{X}_t^\epsilon$ and an auxiliary process $\bar{X}_t$ (see Lemma 4.3 below).

The main contribution of the present paper is to prove the $L^p(p > 2)$ -stronger convergence than the convergence in probability and mean-square shown in [4, 9, 10, 14, 23, 15, 20, 21], and propose an explicit rate of $L^p(p > 2)$ -strong convergence to the averaged effective dynamics, which is crucial in numerical analysis and engineering fields. In addition, by Hölder's inequality, Chebyshev's inequality and our results, it is very easy to derive that the slow component can be approximated by solutions of the corresponding averaged equation in the sense of both convergence in $L^q(0 < q \leq p)$ and convergence in probability.

The organization of this paper is as follows. In Section 2, we present the ergodicity of the fast equation with frozen slow component. In Section 3, some a priori estimates are presented. These results will be utilized in the subsequent discussion. In Section 4, we first prove the $L^p(p > 2)$ -strong convergence version for stochastic averaging principle with an explicit convergence order under suitable assumptions (see Theorem 4.1 below), and then, by strengthening the conditions in Theorem 4.1, we may get $L^p(p > 2)$ -strong convergence and faster rate of convergence (see Theorem 4.2 below), which is shown to be optimal in L^2 (cf.[21]).

Throughout this paper, the letter C with or without subscripts will denote positive constants whose value may change in different occasions. We will write the dependence of constant on parameters explicitly if it is essential.

2. Ergodicity for the fast motion equation

In this section, we give an exponential ergodicity of the fast equation, which has an independent interest.

For fixed $x \in \mathbb{R}^n$ we consider the following problem associated with the fast equation with frozen slow component

$$\begin{cases} \dot{Y}_t = f(x, Y_t) + g(x, Y_t)\dot{W}_t + h(x, Y_t)\dot{N}_t, \\ Y_0 = y, \end{cases} \quad (2.1)$$

where W_t is a standard Wiener process of dimension d_2 , N_t is a scalar Poisson process with constant intensity λ_2 .

By arguing as before, for any fixed slow component $x \in \mathbb{R}^n$ and any initial data $y \in \mathbb{R}^m$, system (2.1) has a unique strong solution denoted by $Y_t^{x,y}$. For $\|Y_t^{x,y}\|^2$, we have by Itô's formula

$$\begin{aligned} \|Y_t^{x,y}\|^2 = \|y\|^2 &+ 2 \int_0^t (f(x, Y_s^{x,y}), Y_s^{x,y}) ds + 2 \int_0^t (g(x, Y_s^{x,y}), Y_s^\epsilon) dW_s \\ &+ 2 \int_0^t (h(x, Y_s^{x,y}), Y_s^{x,y}) dN_s + \int_0^t \|g(x, Y_s^{x,y})\|^2 ds \\ &+ \lambda_2 \int_0^t \|h(x, Y_s^{x,y})\|^2 ds \quad a.s., \end{aligned}$$

which implies

$$\begin{aligned} \frac{d}{dt} \mathbb{E} \|Y_t^{x,y}\|^2 &= 2\mathbb{E}(f(x, Y_t^{x,y}), Y_t^{x,y}) + 2\lambda_2 \mathbb{E}(h(x, Y_t^{x,y}), Y_t^{x,y}) \\ &\quad + \mathbb{E} \|g(x, Y_t^{x,y})\|^2 + \lambda_2 \mathbb{E} \|h(x, Y_t^{x,y})\|^2. \end{aligned} \quad (2.2)$$

By (1.3) in **(H3)** we have

$$2(f(x, Y_t^{x,y}), Y_t^{x,y}) \leq -2\beta_1 \|Y_t^{x,y}\|^2 + C. \quad (2.3)$$

By **(H2)** and Young's inequality we obtain

$$\begin{aligned} 2\lambda_2(h(x, Y_t^{x,y}), Y_t^{x,y}) &\leq \frac{\lambda_2^2}{\beta_1} \|h(x, Y_t^{x,y})\|^2 + \beta_1 \|Y_t^{x,y}\|^2 \\ &\leq C + \beta_1 \|Y_t^{x,y}\|^2, \end{aligned} \quad (2.4)$$

$$\|g(x, Y_t^{x,y})\|^2 \leq C \quad (2.5)$$

and

$$\lambda_2 \|h(x, Y_t^{x,y})\|^2 \leq C. \quad (2.6)$$

From (2.2)-(2.6), we immediately see that

$$\frac{d}{dt} \mathbb{E} \|Y_t^{x,y}\|^2 \leq -\beta_1 \mathbb{E} \|Y_t^{x,y}\|^2 + C,$$

which, according to Gronwall's inequality, means

$$\mathbb{E} \|Y_t^{x,y}\|^2 \leq C (e^{-\beta_1 t} \|y\|^2 + 1), \quad t > 0, \quad (2.7)$$

for some constant $C > 0$.

Let $Y_t^{x,y'}$ be the solution of system (2.1)-(2.1) with initial value $Y_0 = y'$. Thanks to (1.2), (1.4), (1.5), (1.6) and (1.7), we can derive that for any $t \geq 0$,

$$\begin{aligned} &\mathbb{E} \|Y_t^{x,y} - Y_t^{x,y'}\|^2 \\ = &\|y - y'\|^2 + 2\mathbb{E} \int_0^t (f(x, Y_s^{x,y}) - f(x, Y_s^{x,y'}), Y_s^{x,y} - Y_s^{x,y'}) ds \\ &+ 2\lambda_2 \mathbb{E} \int_0^t (h(x, Y_s^{x,y}) - h(x, Y_s^{x,y'}), Y_s^{x,y} - Y_s^{x,y'}) ds \\ &+ \mathbb{E} \int_0^t \|g(x, Y_s^{x,y}) - g(x, Y_s^{x,y'})\|^2 ds \\ &+ \lambda_2 \mathbb{E} \int_0^t \|h(x, Y_s^{x,y}) - h(x, Y_s^{x,y'})\|^2 ds \\ \leq &\|y - y'\|^2 - \eta \int_0^t \mathbb{E} \|Y_s^{x,y} - Y_s^{x,y'}\|^2 ds. \end{aligned}$$

By Gronwall's inequality we have

$$\mathbb{E}\|Y_t^{x,y} - Y_t^{x,y'}\|^2 \leq \|y - y'\|^2 e^{-\eta t}, \quad (2.8)$$

where $\eta := -(2\beta_3 + 2\lambda_2\beta_4 + C_g + \lambda_2 C_h) > 0$.

For any $x \in \mathbb{R}^n$ denote by P_t^x the Markov semigroup associated to equation (2.1)-(2.1) defined by

$$P_t^x \psi(y) = \mathbb{E}\psi(Y_t^{x,y}), \quad t \geq 0, \quad y \in \mathbb{R}^m,$$

for any $\psi \in \mathcal{B}_b(\mathbb{R}^m)$ the space of bounded functions on $\mathbb{R}^m$. We also recall a probability μ^x on $\mathbb{R}^m$ which is called an invariant measure for $(P_t^x)_{t \geq 0}$ if

$$\int_{\mathbb{R}^m} P_t^x \psi d\mu^x = \int_{\mathbb{R}^m} \psi d\mu^x, \quad t \geq 0,$$

for any bounded function $\psi \in \mathcal{B}_b(\mathbb{R}^m)$. Thanks to (2.7), by adapting the arguments used in [20] and [21], it is possible to show that there exists a unique invariant measure μ^x for the semigroup P_t^x , which satisfies

$$\int_{\mathbb{R}^m} \|y\|^2 \mu^x(dy) \leq C(1 + \|x\|^2). \quad (2.9)$$

Set $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ and define the averaging function

$$\bar{a}(x) := \int_{\mathbb{R}^m} a(x, y) \mu^x(dy).$$

By the invariant property of μ^x , (1.2) and (2.8) we have

$$\begin{aligned} \|\mathbb{E}a(x, Y_t^{x,y}) - \bar{a}(x)\|^2 &= \left\| \int_{\mathbb{R}^m} \mathbb{E}(a(x, Y_t^{x,y}) - a(x, Y_t^{x,z})) \mu^x(dz) \right\|^2 \\ &\leq C \int_{\mathbb{R}^m} \mathbb{E} \|Y_t^{x,y} - Y_t^{x,z}\|^2 \mu^x(dz) \\ &\leq C e^{-\eta t} \int_{\mathbb{R}^m} \|y - z\|^2 \mu^x(dz) \\ &\leq C e^{-\eta t} (\|y\|^2 + \int_{\mathbb{R}^m} \|z\|^2 \mu^x(dz)) \\ &\leq C e^{-\eta t} (1 + \|x\|^2 + \|y\|^2). \end{aligned} \quad (2.10)$$

□

In this position, we sketch an outline of the proof of the main result for the reader's convenience. First of all, we show one estimate about the strong convergence speed of X_t^ϵ to $\hat{X}_t^\epsilon$ (see Lemma 3.5 in Section 3). Secondly, we prove the other estimate about the strong convergence speed of $\hat{X}_t^\epsilon$ to $\bar{X}_t$ (see Lemma 4.3 in Section 4). Lastly, based on above two estimates, it is very easy to arrive at X_t^ϵ $L^p(p > 2)$ -strong convergence to $\bar{X}_t$, which is our main result.

3. Some a priori estimates

In this section, we present some a priori estimates for solution processes X^ϵ and Y^ϵ .

Lemma 3.1. *For any $p > 1$ and $T_0 > 0$ there exists a constant $C_1 > 0$ such that for all $\epsilon \in (0, 1)$,*

$$\mathbb{E} \sup_{0 \leq t \leq T_0} \|X_t^\epsilon\|^{2p} \leq C_1, \quad (3.1)$$

where C_1 is independent of ϵ .

Proof. For $\|X_t^\epsilon\|^{2p}$, we have by Itô's formula

$$\begin{aligned} \|X_t^\epsilon\|^{2p} = \|x_0\|^{2p} &+ 2p \int_0^t \|X_s^\epsilon\|^{2p-2} (a(X_s^\epsilon, Y_s^\epsilon), X_s^\epsilon) ds \\ &+ p \int_0^t \|X_s^\epsilon\|^{2p-2} \|b(X_s^\epsilon)\|^2 ds \\ &+ p\lambda_1 \int_0^t \|X_s^\epsilon\|^{2p-2} \|c(X_s^\epsilon)\|^2 ds \\ &+ 2p(p-1) \int_0^t \|X_s^\epsilon\|^{2(p-2)} (b(X_s^\epsilon), X_s^\epsilon)^2 ds \\ &+ 2p(p-1)\lambda_1 \int_0^t \|X_s^\epsilon\|^{2(p-2)} (c(X_s^\epsilon), X_s^\epsilon)^2 ds \\ &+ 2p \int_0^t \|X_s^\epsilon\|^{2p-2} (b(X_s^\epsilon), X_s^\epsilon) dB_s \\ &+ 2p \int_0^t \|X_s^\epsilon\|^{2p-2} (c(X_s^\epsilon), X_s^\epsilon) dP_s. \end{aligned} \quad (3.2)$$

By the boundedness of a in **(H2)**, we have

$$(a(X_s^\epsilon, Y_s^\epsilon), X_s^\epsilon) \leq C \|X_s^\epsilon\|. \quad (3.3)$$

In terms of Remark1.1, we have

$$\|b(X_s^\epsilon)\|^2 + \|c(X_s^\epsilon)\|^2 \leq C(1 + \|X_s^\epsilon\|^2) \quad (3.4)$$

and

$$(b(X_s^\epsilon), X_s^\epsilon)^2 + (c(X_s^\epsilon), X_s^\epsilon)^2 \leq C \|X_s^\epsilon\|^2 (1 + \|X_s^\epsilon\|^2). \quad (3.5)$$

Note that $\hat{P}_s = P_s - \lambda_1 s$ is a martingale. By (3.2)-(3.5) and $P_s = \hat{P}_s + \lambda_1 s$, we have

$$\begin{aligned} \|X_t^\epsilon\|^{2p} &\leq \|x_0\|^{2p} + C \int_0^t \|X_s^\epsilon\|^{2p-1} ds \\ &\quad + C \int_0^t \|X_s^\epsilon\|^{2p} ds + C \int_0^t \|X_s^\epsilon\|^{2p-2} ds \\ &\quad + C \int_0^t \|X_s^\epsilon\|^{2p-2} (b(X_s^\epsilon), X_s^\epsilon) dB_s \\ &\quad + C \int_0^t \|X_s^\epsilon\|^{2p-2} (c(X_s^\epsilon), X_s^\epsilon) d\hat{P}_s \\ &\quad + C \int_0^t \|X_s^\epsilon\|^{2p-2} (c(X_s^\epsilon), X_s^\epsilon) ds. \end{aligned} \quad (3.6)$$

By Young's inequality, we have

$$\int_0^t \|X_s^\epsilon\|^{2p-1} ds + \int_0^t \|X_s^\epsilon\|^{2p-2} ds \leq C + C \int_0^t \|X_s^\epsilon\|^{2p} ds. \quad (3.7)$$

By Remark1.1 and Young's inequality, we have

$$\begin{aligned} \int_0^t \|X_s^\epsilon\|^{2p-2} (c(X_s^\epsilon), X_s^\epsilon) ds &\leq \int_0^t \|X_s^\epsilon\|^{2p-1} \|c(X_s^\epsilon)\| ds \\ &\leq C \int_0^t \|X_s^\epsilon\|^{2p-1} (1 + \|X_s^\epsilon\|) ds \end{aligned}$$

$$\begin{aligned}
&= C \int_0^t \|X_s^\epsilon\|^{2p-1} ds + C \int_0^t \|X_s^\epsilon\|^{2p} ds \\
&\leq C + C \int_0^t \|X_s^\epsilon\|^{2p} ds.
\end{aligned} \tag{3.8}$$

Then, by Burkholder's inequality (cf.[7]), Remark1.1, and Young's inequality twice we get for any $T \in [0, T_0]$ that

$$\begin{aligned}
&\mathbb{E} \sup_{0 \leq t \leq T} \int_0^t \|X_s^\epsilon\|^{2p-2} (b(X_s^\epsilon), X_s^\epsilon) dB_s + \mathbb{E} \sup_{0 \leq t \leq T} \int_0^t \|X_s^\epsilon\|^{2p-2} (c(X_s^\epsilon), X_s^\epsilon) d\hat{P}_s \\
&\leq C \mathbb{E} \left[\int_0^T \|X_s^\epsilon\|^{4p-4} (b(X_s^\epsilon), X_s^\epsilon)^2 ds \right]^{\frac{1}{2}} + C \mathbb{E} \left[\int_0^T \|X_s^\epsilon\|^{4p-4} (c(X_s^\epsilon), X_s^\epsilon)^2 ds \right]^{\frac{1}{2}} \\
&\leq C \mathbb{E} \left[\int_0^T \|X_s^\epsilon\|^{4p-2} (1 + \|X_s^\epsilon\|^2) ds \right]^{\frac{1}{2}} \\
&\leq C \mathbb{E} \left[\int_0^T \|X_s^\epsilon\|^{4p-2} ds \right]^{\frac{1}{2}} + C \mathbb{E} \left[\int_0^T \|X_s^\epsilon\|^{4p} ds \right]^{\frac{1}{2}} \\
&\leq C + C \mathbb{E} \left[\int_0^T \|X_s^\epsilon\|^{4p} ds \right]^{\frac{1}{2}} + C \mathbb{E} \left[\int_0^T \|X_s^\epsilon\|^{4p} ds \right]^{\frac{1}{2}} \\
&\leq C + C \mathbb{E} \left[\int_0^T \|X_s^\epsilon\|^{4p} ds \right]^{\frac{1}{2}} \\
&\leq C + C \mathbb{E} \left[\sup_{0 \leq t \leq T} \|X_s^\epsilon\|^{2p} \int_0^T \|X_s^\epsilon\|^{2p} ds \right]^{\frac{1}{2}} \\
&\leq C + \frac{1}{4} \mathbb{E} \sup_{0 \leq t \leq T} \|X_s^\epsilon\|^{2p} + \int_0^T \|X_s^\epsilon\|^{2p} ds
\end{aligned} \tag{3.9}$$

Moreover, we have by (3.6)-(3.9) for any $T \in [0, T_0]$

$$\begin{aligned}
\mathbb{E} \sup_{0 \leq t \leq T} \|X_t^\epsilon\|^{2p} &\leq \|x_0\|^{2p} + C + \frac{1}{4} \mathbb{E} \sup_{0 \leq r \leq T} \|X_r^\epsilon\|^{2p} + C \mathbb{E} \int_0^T \|X_s^\epsilon\|^{2p} ds \\
&\leq \|x_0\|^{2p} + C + \frac{1}{4} \mathbb{E} \sup_{0 \leq r \leq T} \|X_r^\epsilon\|^{2p} + C_p \int_0^T \mathbb{E} \sup_{0 \leq r \leq s} \|X_r^\epsilon\|^{2p} ds.
\end{aligned}$$

Therefore,

$$\mathbb{E} \sup_{0 \leq t \leq T} \|X_t^\epsilon\|^{2p} \leq \|x_0\|^{2p} + C + C \int_0^T \mathbb{E} \sup_{0 \leq r \leq s} \|X_r^\epsilon\|^{2p} ds.$$

By Gronwall's inequality, the desired estimate is obtained. $\square$

Lemma 3.2. *For any $p > 1$ and $T_0 > 0$ there exists a constant $C_2 > 0$ such that for all $\epsilon \in (0, 1)$,*

$$\sup_{0 \leq t \leq T_0} \mathbb{E} \|Y_t^\epsilon\|^{2p} \leq C_2, \quad (3.10)$$

where C_2 is independent of ϵ .

Proof. For $\|Y_t^\epsilon\|^{2p}$, we have by Itô's formula

$$\begin{aligned} \|Y_t^\epsilon\|^{2p} = \|y_0\|^{2p} &+ \frac{2p}{\epsilon} \int_0^t \|Y_s^\epsilon\|^{2p-2} (f(X_s^\epsilon, Y_s^\epsilon), Y_s^\epsilon) ds \\ &+ \frac{2p}{\sqrt{\epsilon}} \int_0^t \|Y_s^\epsilon\|^{2p-2} (g(X_s^\epsilon, Y_s^\epsilon), Y_s^\epsilon) dW_s \\ &+ 2p \int_0^t \|Y_s^\epsilon\|^{2p-2} (h(X_s^\epsilon, Y_s^\epsilon), Y_s^\epsilon) dN_s^\epsilon \\ &+ \frac{p}{\epsilon} \int_0^t \|Y_s^\epsilon\|^{2p-2} \|g(X_s^\epsilon, Y_s^\epsilon)\|^2 ds \\ &+ \frac{p\lambda_2}{\epsilon} \int_0^t \|Y_s^\epsilon\|^{2p-2} \|h(X_s^\epsilon, Y_s^\epsilon)\|^2 ds \\ &+ \frac{2p(p-1)}{\epsilon} \int_0^t \|Y_s^\epsilon\|^{2(p-2)} (g(X_s^\epsilon, Y_s^\epsilon), Y_s^\epsilon)^2 ds \\ &+ \frac{2p(p-1)\lambda_2}{\epsilon} \int_0^t \|Y_s^\epsilon\|^{2(p-2)} (h(X_s^\epsilon, Y_s^\epsilon), Y_s^\epsilon)^2 ds. \end{aligned} \quad (3.11)$$

By (1.3), we have

$$(f(X_s^\epsilon, Y_s^\epsilon), Y_s^\epsilon) \leq -\beta_1 \|Y_s^\epsilon\|^2 + \beta_2. \quad (3.12)$$

By Young's inequality and **(H2)** we can conclude that

$$2\lambda_2 (h(x, Y_s^{x,y}), Y_s^{x,y}) \leq \frac{\lambda_2^2}{\beta_1} \|h(x, Y_s^{x,y})\|^2 + \beta_1 \|Y_s^{x,y}\|^2$$

$$\leq C + \beta_1 \|Y_s^{x,y}\|^2, \quad (3.13)$$

$$(g(X_s^\epsilon, Y_s^\epsilon), Y_s^\epsilon)^2 \leq C \|Y_s^\epsilon\|^2 \quad (3.14)$$

and

$$(h(X_s^\epsilon, Y_s^\epsilon), Y_s^\epsilon)^2 \leq C \|Y_s^\epsilon\|^2. \quad (3.15)$$

Taking expectation on both sides of (3.11) and by (3.12)-(3.15) we have

$$\begin{aligned} \mathbb{E} \|Y_t^\epsilon\|^{2p} \leq \|y_0\|^{2p} & - \frac{p\beta_1}{\epsilon} \mathbb{E} \int_0^t \|Y_s^\epsilon\|^{2p} ds + \frac{C_{p,\beta_2\lambda_2}}{\epsilon} \mathbb{E} \int_0^t \|Y_s^\epsilon\|^{2p-2} ds \\ & + \frac{C_{p,\lambda_2}}{\epsilon} \mathbb{E} \int_0^t \|Y_s^\epsilon\|^{2p-1} ds. \end{aligned}$$

Furthermore, we have by taking $\kappa > 0$ small enough for Young's inequality in the form $ab \leq \kappa |b|^m + C_{\kappa,m} |a|^{m/(m-1)}$,

$$\mathbb{E} \|Y_t^\epsilon\|^{2p} \leq \|y_0\|^{2p} - \frac{C_{p,\beta_1,\beta_2,\lambda_2}}{\epsilon} \mathbb{E} \int_0^t \|Y_s^\epsilon\|^{2p} ds + \frac{C'_{p,\beta_1,\beta_2,\lambda_2}}{\epsilon} t,$$

which, with the help of Gronwall's inequality, yields

$$\mathbb{E} \|Y_t^\epsilon\|^{2p} \leq C_2.$$

□

Lemma 3.3. *For all $t \in [0, T_0]$, $h \in (0, 1)$ and $p > 1$ there exists a positive constant C such that*

$$\mathbb{E} \|X_{t+h}^\epsilon - X_t^\epsilon\|^{2p} \leq Ch^p, \quad (3.16)$$

where C is independent of (ϵ, h) .

Proof. It is clear that for all $t \in [0, T_0]$ and $h \in (0, 1)$,

$$\begin{aligned} X_{t+h}^\epsilon - X_t^\epsilon &= \int_t^{t+h} a(X_s^\epsilon, Y_s^\epsilon) ds + \int_t^{t+h} b(X_s^\epsilon) dB_s + \int_t^{t+h} c(X_s^\epsilon) dP_s \\ &:= I_1(t, h) + I_2(t, h) + I_3(t, h), \quad a.s.. \end{aligned}$$

By Hölder's inequality and the boundedness of the function a , we have

$$\begin{aligned}
\mathbb{E} \|I_1(t, h)\|^{2p} &= \mathbb{E} \left\| \int_t^{t+h} a(X_s^\epsilon, Y_s^\epsilon) ds \right\|^{2p} \\
&\leq h^{2p-1} \mathbb{E} \int_t^{t+h} \|a(X_s^\epsilon, Y_s^\epsilon)\|^{2p} ds \\
&\leq Ch^{2p}.
\end{aligned}$$

By $P_s = \hat{P}_s + \lambda_1 s$, Burkholder's inequality (cf.[7]), Hölder's inequality, Remark 1.1 and Lemma 3.1, we have

$$\begin{aligned}
\mathbb{E} \|I_2(t, h)\|^{2p} &= \mathbb{E} \left\{ \left\| \int_t^{t+h} b(X_s^\epsilon) dB_s \right\| \right\}^{2p} \\
&\leq \mathbb{E} \left\{ \int_t^{t+h} \|b(X_s^\epsilon)\|^2 ds \right\}^p \\
&\leq h^{p-1} \int_t^{t+h} \mathbb{E} \|b(X_s^\epsilon)\|^{2p} ds \\
&\leq h^{p-1} \int_t^{t+h} \mathbb{E} (1 + \|X_s^\epsilon\|^{2p}) ds \\
&\leq Ch^p
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E} \|I_3(t, h)\|^{2p} &= \mathbb{E} \left\| \int_t^{t+h} c(X_s^\epsilon) d\hat{P}_s + \lambda_1 \int_t^{t+h} c(X_s^\epsilon) ds \right\|^{2p} \\
&\leq C \mathbb{E} \left[\int_t^{t+h} \|c(X_s^\epsilon)\|^2 ds \right]^p + C \mathbb{E} \left[\left\| \lambda_1 \int_t^{t+h} c(X_s^\epsilon) ds \right\| \right]^{2p} \\
&\leq C \mathbb{E} \left[\int_t^{t+h} \|c(X_s^\epsilon)\|^2 ds \right]^p + C \mathbb{E} \left[\lambda_1 \int_t^{t+h} \|c(X_s^\epsilon)\| ds \right]^{2p} \\
&\leq Ch^{p-1} \mathbb{E} \int_t^{t+h} \|c(X_s^\epsilon)\|^{2p} ds + Ch^{2p-1} \mathbb{E} \int_t^{t+h} \|c(X_s^\epsilon)\|^{2p} ds \\
&\leq C(h^{p-1} + h^{2p-1}) \int_t^{t+h} \mathbb{E} (1 + \|X_s^\epsilon\|^{2p}) ds \\
&\leq C(h^p + h^{2p}) \\
&\leq Ch^p.
\end{aligned}$$

Combining above three inequalities, we have

$$\mathbb{E}\|X_{t+h}^\epsilon - X_t^\epsilon\|^{2p} \leq Ch^p.$$

□

Next, we introduce an auxiliary process $(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon) \in \mathbb{R}^n \times \mathbb{R}^m$. Fix a positive number $\delta < 1$ and do a partition of time interval $[0, T_0]$ of size δ . We construct a process $\hat{Y}_t^\epsilon$, with initial datum $\hat{Y}_0^\epsilon = y_0$, by means of the equations

$$d\hat{Y}_t^\epsilon = \frac{1}{\epsilon} f(X_{k\delta}^\epsilon, \hat{Y}_t^\epsilon) dt + \frac{1}{\sqrt{\epsilon}} g(X_{k\delta}^\epsilon, \hat{Y}_t^\epsilon) dW_t + h(X_{k\delta}^\epsilon, \hat{Y}_t^\epsilon) dN_t^\epsilon, \quad \hat{Y}_{k\delta}^\epsilon = Y_{k\delta}^\epsilon$$

for $t \in [k\delta, \min\{(k+1)\delta, T_0\})$, $k \geq 0$, where $X_{k\delta}^\epsilon$ and $Y_{k\delta}^\epsilon$ are slow and fast solution processes at time $k\delta$, respectively. Denote $\lfloor \cdot \rfloor$ to be the integer function and define the process $\hat{X}_t^\epsilon$ by integral

$$\hat{X}_t^\epsilon = x_0 + \int_0^t a(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon) ds + \int_0^t b(X_s^\epsilon) dB_s + \int_0^t c(X_s^\epsilon) dP_s$$

for $t \in [0, T_0]$, where $s(\delta) = \lfloor s/\delta \rfloor \delta$ is the nearest breakpoint preceding s . We will establish convergence of the auxiliary processes $\hat{Y}_t^\epsilon$ to the fast solution process Y_t^ϵ and $\hat{X}_t^\epsilon$ to the slow solution process X_t^ϵ , respectively.

Lemma 3.4. *For all $t \in [0, T_0]$ and $p > 1$ there is a constant C_{p, λ_2} such that*

$$\mathbb{E}\|Y_t^\epsilon - \hat{Y}_t^\epsilon\|^{2p} \leq \frac{C_{p, \lambda_2}}{\epsilon} \delta^{p+1} e^{C_{p, \lambda_2} \frac{\delta}{\epsilon}}. \quad (3.17)$$

Proof. Set $Z_t^\epsilon := Y_t^\epsilon - \hat{Y}_t^\epsilon$, where $t \in [0, T_0]$ with $t \in [k\delta, (k+1)\delta)$. Then

$$\begin{aligned} Z_t^\epsilon &= \frac{1}{\epsilon} \int_{k\delta}^t (f(X_s^\epsilon, Y_s^\epsilon) - f(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon)) ds \\ &\quad + \frac{1}{\sqrt{\epsilon}} \int_{k\delta}^t (g(X_s^\epsilon, Y_s^\epsilon) - g(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon)) dW_s \\ &\quad + \int_{k\delta}^t (h(X_s^\epsilon, Y_s^\epsilon) - h(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon)) dN_s^\epsilon. \end{aligned}$$

By Itô's formula we have

$$\begin{aligned}
\mathbb{E}\|Z_t^\epsilon\|^{2p} &= \frac{2p}{\epsilon} \mathbb{E} \int_{k\delta}^t \|Z_s^\epsilon\|^{2p-2} (f(X_s^\epsilon, Y_s^\epsilon) - f(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon), Z_s^\epsilon) ds \\
&\quad + \frac{2p\lambda_2}{\epsilon} \mathbb{E} \int_{k\delta}^t \|Z_s^\epsilon\|^{2p-2} (h(X_s^\epsilon, Y_s^\epsilon) - h(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon), Z_s^\epsilon) ds \\
&\quad + \frac{p}{\epsilon} \mathbb{E} \int_{k\delta}^t \|Z_s^\epsilon\|^{2p-2} \|g(X_s^\epsilon, Y_s^\epsilon) - g(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon)\|^2 ds \\
&\quad + \frac{p\lambda_2}{\epsilon} \mathbb{E} \int_{k\delta}^t \|Z_s^\epsilon\|^{2p-2} \|h(X_s^\epsilon, Y_s^\epsilon) - h(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon)\|^2 ds \\
&\quad + \frac{2p(p-1)}{\epsilon} \mathbb{E} \int_{k\delta}^t \|Z_s^\epsilon\|^{2(p-2)} (g(X_s^\epsilon, Y_s^\epsilon) - g(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon), Z_s^\epsilon)^2 ds \\
&\quad + \frac{2p(p-1)\lambda_2}{\epsilon} \mathbb{E} \int_{k\delta}^t \|Z_s^\epsilon\|^{2(p-2)} (h(X_s^\epsilon, Y_s^\epsilon) - h(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon), Z_s^\epsilon)^2 ds.
\end{aligned}$$

It follows from **(H1)** and Young's inequality

$$(f(X_s^\epsilon, Y_s^\epsilon) - f(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon), Z_s^\epsilon) \leq C(\|X_s^\epsilon - X_{k\delta}^\epsilon\|^2 + \|Z_s^\epsilon\|^2), \quad (3.18)$$

$$(h(X_s^\epsilon, Y_s^\epsilon) - h(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon), Z_s^\epsilon) \leq C(\|X_s^\epsilon - X_{k\delta}^\epsilon\|^2 + \|Z_s^\epsilon\|^2), \quad (3.19)$$

$$\|g(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon) - g(X_s^\epsilon, Y_s^\epsilon)\|^2 \leq C(\|X_s^\epsilon - X_{k\delta}^\epsilon\|^2 + \|Z_s^\epsilon\|^2), \quad (3.20)$$

$$\|h(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon) - h(X_s^\epsilon, Y_s^\epsilon)\|^2 \leq C(\|X_s^\epsilon - X_{k\delta}^\epsilon\|^2 + \|Z_s^\epsilon\|^2), \quad (3.21)$$

$$(g(X_s^\epsilon, Y_s^\epsilon) - g(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon), Z_s^\epsilon)^2 \leq C(\|X_s^\epsilon - X_{k\delta}^\epsilon\|^2 + \|Z_s^\epsilon\|^2) \|Z_s^\epsilon\|^2 \quad (3.22)$$

and

$$(h(X_s^\epsilon, Y_s^\epsilon) - h(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon), Z_s^\epsilon)^2 \leq C(\|X_s^\epsilon - X_{k\delta}^\epsilon\|^2 + \|Z_s^\epsilon\|^2) \|Z_s^\epsilon\|^2. \quad (3.23)$$

Taking (3.18)-(3.23) into account, we obtain

$$\mathbb{E}\|Z_t^\epsilon\|^{2p} \leq \frac{C_{p,\lambda_2}}{\epsilon} \int_{k\delta}^t \mathbb{E}\|Z_s^\epsilon\|^{2p} ds + \frac{C_{p,\lambda_2}}{\epsilon} \int_{k\delta}^t \mathbb{E}\|Z_s^\epsilon\|^{2p-2} \|X_s^\epsilon - X_{k\delta}^\epsilon\|^2 ds.$$

Using Young's inequality first, then we have by (3.16)

$$\begin{aligned}\mathbb{E}\|Z_t^\epsilon\|^{2p} &\leq \frac{C_{p,\lambda_2}}{\epsilon} \int_{k\delta}^t \mathbb{E}\|Z_s^\epsilon\|^{2p} ds + \frac{C_{p,\lambda_2}}{\epsilon} \int_{k\delta}^t \mathbb{E}\|X_s^\epsilon - X_{k\delta}^\epsilon\|^{2p} ds \\ &\leq \frac{C_{p,\lambda_2}}{\epsilon} \int_{k\delta}^t \mathbb{E}\|Z_s^\epsilon\|^{2p} ds + \frac{C_{p,\lambda_2}}{\epsilon} \delta^{p+1}.\end{aligned}$$

Furthermore, we have by Gronwall's inequality

$$\mathbb{E}\|Z_t^\epsilon\|^{2p} \leq \frac{C_{p,\lambda_2}}{\epsilon} \delta^{p+1} e^{C_{p,\lambda_2} \frac{\delta}{\epsilon}}.$$

Hence, we complete the proof of this lemma. $\square$

Lemma 3.5. *For all $t \in [0, T_0]$ and $p > 1$ there are two constants C_{p,λ_2,T_0} and C_{p,λ_2} such that*

$$\mathbb{E} \sup_{0 \leq t \leq T_0} \|X_t^\epsilon - \hat{X}_t^\epsilon\|^{2p} \leq C_{p,\lambda_2,T_0} \left(\delta^p + \frac{\delta^{p+1}}{\epsilon} e^{C_{p,\lambda_2} \frac{\delta}{\epsilon}} \right). \quad (3.24)$$

Proof. First of all, we observe that

$$X_t^\epsilon - \hat{X}_t^\epsilon = \int_0^t (a(X_s^\epsilon, Y_s^\epsilon) - a(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon)) ds. \quad (3.25)$$

By Itô's formula we have for any $t \in [0, T_0]$

$$\|X_t^\epsilon - \hat{X}_t^\epsilon\|^{2p} = 2p \int_0^t \|X_s^\epsilon - \hat{X}_s^\epsilon\|^{2p-2} (a(X_s^\epsilon, Y_s^\epsilon) - a(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon), X_s^\epsilon - \hat{X}_s^\epsilon) ds.$$

It can conclude that by **(H1)** and Young's inequality

$$\begin{aligned}&(a(X_s^\epsilon, Y_s^\epsilon) - a(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon), X_s^\epsilon - \hat{X}_s^\epsilon) \\ &\leq C \left[(a(X_s^\epsilon, Y_s^\epsilon) - a(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon))^2 + \|X_s^\epsilon - \hat{X}_s^\epsilon\|^2 \right] \\ &\leq C (\|X_s^\epsilon - X_{s(\delta)}^\epsilon\|^2 + \|Y_s^\epsilon - \hat{Y}_s^\epsilon\|^2 + \|X_s^\epsilon - \hat{X}_s^\epsilon\|^2). \quad (3.26)\end{aligned}$$

By (3.26) and Young's inequality, we have

$$\|X_t^\epsilon - \hat{X}_t^\epsilon\|^{2p} \leq C_p \int_0^t \|X_s^\epsilon - \hat{X}_s^\epsilon\|^{2p} ds + C_p \int_0^t \|X_s^\epsilon - X_{s(\delta)}^\epsilon\|^{2p} ds$$

$$+C_p \int_0^t \|Y_s^\epsilon - \hat{Y}_s^\epsilon\|^{2p} ds.$$

Thanks to Lemma 3.3 and Lemma 3.4, we obtain for any $T \in [0, T_0]$ that

$$\begin{aligned} \mathbb{E} \sup_{0 \leq t \leq T} \|X_t^\epsilon - \hat{X}_t^\epsilon\|^{2p} &\leq C_p \int_0^T \mathbb{E} \sup_{0 \leq r \leq s} \|X_r^\epsilon - \hat{X}_r^\epsilon\|^{2p} ds \\ &\quad + C_p \delta^p + \frac{C_{p,\lambda_2}}{\epsilon} \delta^{p+1} e^{C_{p,\lambda_2} \frac{\delta}{\epsilon}}, \end{aligned}$$

which, with aid of the Gronwall's inequality, yields

$$\mathbb{E} \sup_{0 \leq t \leq T_0} \|X_t^\epsilon - \hat{X}_t^\epsilon\|^{2p} \leq C_{p,\lambda_2,T_0} \left(\delta^p + \frac{\delta^{p+1}}{\epsilon} e^{C_{p,\lambda_2} \frac{\delta}{\epsilon}} \right).$$

□

4. Averaging principle

In this section we will consider the effective dynamical system

$$\begin{cases} \dot{\bar{X}}_t = \bar{a}(\bar{X}_t) + b(\bar{X}_t)\dot{B}_t + c(\bar{X}_t)\dot{P}_t, \\ \bar{X}_0 = x_0 \end{cases} \quad (4.1)$$

with

$$\bar{a}(x) = \int_{\mathbb{R}^m} a(x, y) \mu^x(dy), \quad x \in \mathbb{R}^n,$$

where μ^x denotes the unique invariant measure for system (2.1) introduced in Section 2. Moreover, due to [1], the mapping $\bar{a} : \mathbb{R}^n \mapsto \mathbb{R}^n$ is Lipschitz continuous. Our aim here is proving the main result of this paper. Namely, we are going to verify that the sequence $\{X_t^\epsilon : t \geq 0\}_{\epsilon > 0}$ $L^p(p > 2)$ -strongly converges to the solution process $\{\bar{X}_t : t \geq 0\}$ of the averaged system (4.1) as ϵ goes to zero. To this end, we shall explore the differences between the solution of the averaged equation and the auxiliary process $\hat{X}_t^\epsilon$. Next, by the construction of $\hat{Y}_t^\epsilon$ and a time shift transformation, we have for any fixed k and $s \in [0, \delta)$ the equalities

$$\hat{Y}_{s+k\delta}^\epsilon = Y_{k\delta}^\epsilon + \frac{1}{\epsilon} \int_{k\delta}^{k\delta+s} f(X_{k\delta}^\epsilon, \hat{Y}_r^\epsilon) dr + \frac{1}{\sqrt{\epsilon}} \int_{k\delta}^{k\delta+s} g(X_{k\delta}^\epsilon, \hat{Y}_r^\epsilon) dW_r$$

$$\begin{aligned}
& + \int_{k\delta}^{k\delta+s} h(X_{k\delta}^\epsilon, \hat{Y}_r^\epsilon) dN_r^\epsilon \\
& = Y_{k\delta}^\epsilon + \frac{1}{\epsilon} \int_0^s f(X_{k\delta}^\epsilon, \hat{Y}_{r+k\delta}^\epsilon) dr + \frac{1}{\sqrt{\epsilon}} \int_0^s g(X_{k\delta}^\epsilon, \hat{Y}_{r+k\delta}^\epsilon) dW_r^* \\
& + \int_0^s h(X_{k\delta}^\epsilon, \hat{Y}_{r+k\delta}^\epsilon) dN_r^{\epsilon*}, \tag{4.2}
\end{aligned}$$

where $W_t^* = W_{t+k\delta} - W_{k\delta}$ and $N_t^{\epsilon*} = N_{t+k\delta}^\epsilon - N_{k\delta}^\epsilon$ are the shift version of W_t and N_t^ϵ respectively, and hence they have the same distribution. Let $\bar{W}_t$ be a Wiener process and independent of B_t and W_t , $\bar{N}_t$ be a simple scalar Poisson process with intensity λ_2 and independent of P_t and N_t^ϵ . Construct a process $Y^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}$ by means of

$$\begin{aligned}
Y_{s/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon} & = Y_{k\delta}^\epsilon + \int_0^{s/\epsilon} f(X_{k\delta}^\epsilon, Y_r^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) dr + \int_0^{s/\epsilon} g(X_{k\delta}^\epsilon, Y_r^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) d\bar{W}_r \\
& + \int_0^{s/\epsilon} h(X_{k\delta}^\epsilon, Y_r^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) d\bar{N}_r \\
& = Y_{k\delta}^\epsilon + \frac{1}{\epsilon} \int_0^s f(X_{k\delta}^\epsilon, Y_{r/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) dr + \frac{1}{\sqrt{\epsilon}} \int_0^s g(X_{k\delta}^\epsilon, Y_{r/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) d\bar{W}_r^\epsilon \\
& + \int_0^s h(X_{k\delta}^\epsilon, Y_{r/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) d\bar{N}_r^\epsilon, \tag{4.3}
\end{aligned}$$

where $\bar{W}_t^\epsilon = \sqrt{\epsilon} \bar{W}_{t/\epsilon}$ and $\bar{N}_t^\epsilon = \bar{N}_{t/\epsilon}$ are the scaled version of $\bar{W}_t$ and $\bar{N}_t$ respectively. By comparison, (4.2) and (4.3) yield

$$(X_{k\delta}^\epsilon, \hat{Y}_{s+k\delta}^\epsilon) \sim (X_{k\delta}^\epsilon, Y_{s/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}), \quad s \in [0, \delta),$$

where $\sim$ denotes coincidence in distribution sense. Set

$$\mathcal{J}_k^\epsilon := \mathbb{E} \left\| \int_0^{\delta/\epsilon} (a(X_{k\delta}^\epsilon, \hat{Y}_{s\epsilon+k\delta}^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)) ds \right\|^2, \quad 0 \leq k \leq \lfloor T_0/\delta \rfloor - 1,$$

then we state the critical lemma, which will be used later.

Lemma 4.1. *Suppose that (H1)–(H5) hold, then there exists a constant $C > 0$ such that*

$$\mathcal{J}_k^\epsilon \leq C \frac{\delta}{\epsilon}, \quad 0 \leq k \leq \lfloor T_0/\delta \rfloor - 1, \tag{4.4}$$

where C is independent of (ϵ, δ) .

Proof. Let $\mathbb{Q}^y$ denote the probability law of the diffusion process $\{Y_t^x : t \geq 0\}$ which is governed by differential equation

$$dY_t^x = f(x, Y_t^x)dt + g(x, Y_t^x)d\bar{W}_t + h(x, Y_t^x)d\bar{N}_t.$$

When its initial value is $Y_0^x = y$ and we denote the solution by $Y_t^{x,y}$. The expectation with respect to $\mathbb{Q}^y$ is denoted by $\mathbb{E}^y$. Hence we have

$$\mathbb{E}^y(\psi(Y_t^x)) = \mathbb{E}(\psi(Y_t^{x,y}))$$

for all bounded function ψ . For more details on $\mathbb{Q}^y$ the readers are referred to [19]. First we note that it is easy to show that $\mathcal{J}_k^\epsilon < \infty$, $k = 0, 1, \dots, \lfloor T_0/\delta \rfloor - 1$. Then by Fubini theorem, we see that

$$\begin{aligned} \mathcal{J}_k^\epsilon &= \mathbb{E} \left\| \int_0^{\delta/\epsilon} [a(X_{k\delta}^\epsilon, \hat{Y}_{s\epsilon+k\delta}^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)] ds \right\|^2 \\ &= \frac{1}{\epsilon^2} \mathbb{E} \left\| \int_0^\delta [a(X_{k\delta}^\epsilon, \hat{Y}_{s+k\delta}^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)] ds \right\|^2 \\ &= \frac{1}{\epsilon^2} \mathbb{E} \left\| \int_0^\delta [a(X_{k\delta}^\epsilon, Y_{s/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) - \bar{a}(X_{k\delta}^\epsilon)] ds \right\|^2 \\ &= \frac{1}{\epsilon^2} \mathbb{E} \int_0^\delta \int_0^\delta [a(X_{k\delta}^\epsilon, Y_{s/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) - \bar{a}(X_{k\delta}^\epsilon)] [a(X_{k\delta}^\epsilon, Y_{\tau/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) - \bar{a}(X_{k\delta}^\epsilon)] ds d\tau \\ &= \frac{2}{\epsilon^2} \int_0^\delta \int_\tau^\delta \mathbb{E} \left[\left(a(X_{k\delta}^\epsilon, Y_{s/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) - \bar{a}(X_{k\delta}^\epsilon) \right) \right. \\ &\quad \left. \times \left(a(X_{k\delta}^\epsilon, Y_{\tau/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) - \bar{a}(X_{k\delta}^\epsilon) \right) \right] ds d\tau, \end{aligned}$$

$k = 0, 1, \dots, \lfloor T_0/\delta \rfloor - 1$. Set

$$J_k(\tau, s, x, y) := \mathbb{E} \left[\left(a(x, Y_s^{x,y}) - \bar{a}(x) \right) \left(a(x, Y_\tau^{x,y}) - \bar{a}(x) \right) \right].$$

It follows from the Markov property of $Y_t^{x,y}$ that

$$J_k(\tau, s, x, y) = \mathbb{E}^y \left\{ \mathbb{E}^y \left[\left(a(x, Y_s^x) - \bar{a}(x) \right) \left(a(x, Y_\tau^x) - \bar{a}(x) \right) \middle| \mathcal{M}_\tau^x \right] \right\}$$

$$= \mathbb{E}^y \left\{ \left(a(x, Y_\tau^x) - \bar{a}(x) \right) \times \mathbb{E}^{Y_\tau^{x,y}} \left[\left(a(x, Y_{s-\tau}^x) - \bar{a}(x) \right) \right] \right\},$$

where $\mathcal{M}_t^x$ denotes the σ -field generated by $\{Y_r^x; r \leq t\}$, $\mathbb{E}^{Y_\tau^{x,y}}[(a(x, Y_{s-\tau}^x) - \bar{a}(x))]$ means the function $\mathbb{E}^{\tilde{y}}[(a(x, Y_{s-\tau}^x) - \bar{a}(x))]$ evaluated at $\tilde{y} = Y_\tau^{x,y}$. Therefore the Cauchy- Schwarz's inequality yields

$$\begin{aligned} J_k(\tau, s, x, y) &\leq \left\{ \mathbb{E}^y \left\| a(x, Y_\tau^x) - \bar{a}(x) \right\|^2 \right\}^{\frac{1}{2}} \\ &\quad \times \left\{ \mathbb{E}^y \left[\left\| \mathbb{E}^{\tilde{y}}(a(x, Y_{s-\tau}^x)) - \bar{a}(x) \right\|_{\tilde{y}=Y_\tau^{x,y}}^2 \right] \right\}^{\frac{1}{2}}, \end{aligned}$$

which, with aid of linear growth bound of the functions $(a, \bar{a})$ and (2.7), implies that

$$\begin{aligned} \mathbb{E}^y \left\| a(x, Y_\tau^x) - \bar{a}(x) \right\|^2 &\leq 2\mathbb{E}^y (\|a(x, Y_\tau^x)\|^2 + \|\bar{a}(x)\|^2) \\ &\leq C(1 + \|x\|^2 + \mathbb{E}^y \|Y_\tau^x\|^2) \\ &\leq C(1 + \|x\|^2 + \|y\|^2). \end{aligned}$$

By using (2.10), we have

$$J_k(\tau, s, x, y) \leq C(1 + \|x\|^2 + \|y\|^2) e^{\frac{-\eta(s-\tau)}{2}}. \quad (4.5)$$

Let $\mathcal{M}_{k\delta}^\epsilon$ be the σ -field generated by $X_{k\delta}^\epsilon$ and $Y_{k\delta}^\epsilon$, which is independent of $\{Y_r^{x,y} : r \geq 0\}$. By adapting the approach in [19] (Theorem 7.1.2) we can deduce from (4.5), Lemma 3.1 and Lemma 3.2 that

$$\begin{aligned} \mathcal{J}_k^\epsilon &= \frac{2}{\epsilon^2} \int_0^\delta \int_\tau^\delta \mathbb{E} \left(\mathbb{E} \left[\left(a(X_{k\delta}^\epsilon, Y_{s/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) - \bar{a}(X_{k\delta}^\epsilon) \right) \right. \right. \\ &\quad \left. \left. \times \left(a(X_{k\delta}^\epsilon, Y_{\tau/\epsilon}^{X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon}) - \bar{a}(X_{k\delta}^\epsilon) \right) \middle| \mathcal{M}_{k\delta}^\epsilon \right] \right) ds d\tau \\ &= \frac{2}{\epsilon^2} \int_0^\delta \int_\tau^\delta \mathbb{E} \left[\left(J_k(\tau/\epsilon, s/\epsilon, x, y) \right) \middle|_{(x,y)=(X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon)} \right] ds d\tau \\ &\leq \frac{C}{\epsilon^2} \int_0^\delta \int_\tau^\delta e^{-\eta(s-\tau)/2\epsilon} ds d\tau \end{aligned}$$

$$\leq \frac{C}{\epsilon} \left[\delta - \frac{2\epsilon}{\eta} (1 - e^{\frac{\delta\eta}{2\epsilon}}) \right],$$

which completes the proof. $\square$

Lemma 4.2. *Suppose that (H1)–(H5) are satisfied. Then for any $T_0 > 0$ and $p > 1$, we have*

$$\mathbb{E} \sup_{0 \leq t \leq T_0} \left\| \int_0^t (a(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_s^\epsilon)) ds \right\|^{2p} \leq C_{p,T_0} (\delta^p + \frac{\epsilon}{\delta}), \quad (4.6)$$

where C_{p,T_0} is independent of (ϵ, δ) .

Proof. Notice that for any $t \in [0, T_0]$, there exists an $n_t = \lfloor t/\delta \rfloor$ such that $t \in [n_t\delta, (n_t + 1)\delta \wedge T_0]$. Therefore, we have representation in the form

$$\int_0^t (a(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_s^\epsilon)) ds := \mathcal{U}_1(t) + \mathcal{U}_2(t) + \mathcal{U}_3(t), \quad (4.7)$$

where

$$\begin{aligned} \mathcal{U}_1(t) &= \sum_{k=0}^{n_t-1} \int_{k\delta}^{(k+1)\delta} (a(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)) ds, \\ \mathcal{U}_2(t) &= \sum_{k=0}^{n_t-1} \int_{k\delta}^{(k+1)\delta} (\bar{a}(X_{k\delta}^\epsilon) - \bar{a}(X_s^\epsilon)) ds = \int_0^{n_t\delta} (\bar{a}(X_{s(\delta)}^\epsilon) - \bar{a}(X_s^\epsilon)) ds, \\ \mathcal{U}_3(t) &= \int_{n_t\delta}^t (a(X_{n_t\delta}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_s^\epsilon)) ds. \end{aligned}$$

For $\mathcal{U}_2(t)$, we have by Lipschitz continuity of $\bar{a}$ and Hölder's inequality

$$\begin{aligned} \mathbb{E} \sup_{0 \leq t \leq T_0} \|\mathcal{U}_2(t)\|^{2p} &= \mathbb{E} \sup_{0 \leq t \leq T_0} \left\| \int_0^{n_t\delta} (\bar{a}(X_{s(\delta)}^\epsilon) - \bar{a}(X_s^\epsilon)) ds \right\|^{2p} \\ &\leq C_{p,T_0} \mathbb{E} \sup_{0 \leq t \leq T_0} \int_0^{n_t\delta} \|X_{s(\delta)}^\epsilon - X_s^\epsilon\|^{2p} ds \\ &\leq C_{p,T_0} \int_0^{T_0} \mathbb{E} \|X_{s(\delta)}^\epsilon - X_s^\epsilon\|^{2p} ds \\ &\leq C_{p,T_0} \delta^p \end{aligned} \quad (4.8)$$

for $T_0 > 0$, where the last inequality is due to (3.16).

We proceed next to the estimate of $I_3(t, \epsilon)$. For all $T_0 > 0$, we have by the boundedness of the function a

$$\begin{aligned}
\|\mathcal{U}_3(t)\|^{2p} &= \left\| \int_{n_t\delta}^t (a(X_{n_t\delta}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_s^\epsilon)) ds \right\|^{2p} \\
&\leq \left[\int_{n_t\delta}^t \|a(X_{n_t\delta}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_s^\epsilon)\| ds \right]^{2p} \\
&\leq \left[\int_{n_t\delta}^t \|a(X_{n_t\delta}^\epsilon, \hat{Y}_s^\epsilon)\| + \|\bar{a}(X_s^\epsilon)\| ds \right]^{2p} \\
&\leq C\delta^{2p}.
\end{aligned} \tag{4.9}$$

Concerning $\mathcal{U}_1(t)$, we can deduce

$$\begin{aligned}
&\mathbb{E} \sup_{0 \leq t \leq T_0} \|\mathcal{U}_1(t)\|^{2p} \\
&= \mathbb{E} \sup_{0 \leq t \leq T_0} \left\| \sum_{k=0}^{\lfloor t/\delta \rfloor - 1} \int_{k\delta}^{(k+1)\delta} (a(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)) ds \right\|^{2p} \\
&\leq \mathbb{E} \sup_{0 \leq t \leq T_0} \left[\lfloor t/\delta \rfloor^{2p-1} \sum_{k=0}^{\lfloor t/\delta \rfloor - 1} \left\| \int_{k\delta}^{(k+1)\delta} (a(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)) ds \right\|^{2p} \right] \\
&\leq \left(\frac{T_0}{\delta} \right)^{2p-1} \sum_{k=0}^{\lfloor T_0/\delta \rfloor - 1} \mathbb{E} \left\| \int_{k\delta}^{(k+1)\delta} (a(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)) ds \right\|^{2p} \\
&\leq \left(\frac{T_0}{\delta} \right)^{2p} \max_{0 \leq k \leq \lfloor T_0/\delta \rfloor - 1} \mathbb{E} \left\| \int_{k\delta}^{(k+1)\delta} (a(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)) ds \right\|^{2p} \\
&= \left(\frac{T_0}{\delta} \right)^{2p} \epsilon^{2p} \max_{0 \leq k \leq \lfloor T_0/\delta \rfloor - 1} \mathbb{E} \left\| \int_0^{\delta/\epsilon} (a(X_{k\delta}^\epsilon, \hat{Y}_{s\epsilon+k\delta}^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)) ds \right\|^{2p}.
\end{aligned}$$

Moreover, we have by **(H2)**,

$$\mathbb{E} \sup_{0 \leq t \leq T_0} \|\mathcal{U}_1(t)\|^{2p}$$

$$\begin{aligned}
&\leq \left(\frac{T_0}{\delta}\right)^{2p} \epsilon^{2p} \max_{0 \leq k \leq \lfloor T_0/\delta \rfloor - 1} \mathbb{E} \left\| \int_0^{\delta/\epsilon} (a(X_{k\delta}^\epsilon, \hat{Y}_{s\epsilon+k\delta}^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)) ds \right\|^{2p} \\
&= C \left(\frac{\epsilon}{\delta}\right)^{2p} \max_{0 \leq k \leq \lfloor T_0/\delta \rfloor - 1} \mathbb{E} \left[\left\| \int_0^{\delta/\epsilon} (a(X_{k\delta}^\epsilon, \hat{Y}_{s\epsilon+k\delta}^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)) ds \right\|^{2p-2} \right. \\
&\quad \left. \times \left\| \int_0^{\delta/\epsilon} (a(X_{k\delta}^\epsilon, \hat{Y}_{s\epsilon+k\delta}^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)) ds \right\|^2 \right] \\
&\leq C \left(\frac{\epsilon}{\delta}\right)^{2p} \left(\frac{\delta}{\epsilon}\right)^{2p-2} \max_{0 \leq k \leq \lfloor T_0/\delta \rfloor - 1} \mathbb{E} \left\| \int_0^{\delta/\epsilon} (a(X_{k\delta}^\epsilon, \hat{Y}_{s\epsilon+k\delta}^\epsilon) - \bar{a}(X_{k\delta}^\epsilon)) ds \right\|^2 \\
&= C \frac{\epsilon^2}{\delta^2} \max_{0 \leq k \leq \lfloor T_0/\delta \rfloor - 1} \mathcal{J}_k^\epsilon.
\end{aligned}$$

Taking (4.4) into account, we can deduce that

$$\mathbb{E} \sup_{0 \leq t \leq T_0} \|\mathcal{U}_1(t)\|^{2p} \leq C \frac{\epsilon}{\delta},$$

which, combining (4.7), (4.8) and (4.9), gives

$$\mathbb{E} \sup_{0 \leq t \leq T_0} \left\| \int_0^t (a(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_s^\epsilon)) ds \right\|^{2p} \leq C_{p,T_0} (\delta^p + \frac{\epsilon}{\delta}).$$

The proof is completed. $\square$

Remark 4.1. *The procedure of the proof of Lemma 4.2 shows that if the function a is a global bounded, we can transform $L^p(p > 2)$ convergence into L^2 convergence problem. Comparing to [15], if a is not a global bounded function, it seems that it is very hard to obtain $L^p(p > 2)$ convergence in our opinions.*

Lemma 4.3. *Suppose that **(H1)**–**(H5)** hold, then there exists a constant $C_{p,\lambda_1,\lambda_2,T_0}$ such that for any $T_0 > 0$ and $p > 1$*

$$\mathbb{E} \sup_{0 \leq t \leq T_0} \|\hat{X}_t^\epsilon - \bar{X}_t\|^{2p} \leq C_{p,\lambda_1,\lambda_2,T_0} \left(\delta^p + \frac{\epsilon}{\delta} + \frac{\delta^{p+1}}{\epsilon} e^{C_{p,\lambda_2} \frac{\delta}{\epsilon}} \right). \quad (4.10)$$

Proof. For any $t \in [0, T_0]$ we write

$$\hat{X}_t^\epsilon - \bar{X}_t = \sum_{i=1}^9 \mathcal{V}_i(t), \quad (4.11)$$

where

$$\begin{aligned} \mathcal{V}_1(t) &= \int_0^t (a(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon) - \bar{a}(X_s^\epsilon)) ds, \\ \mathcal{V}_2(t) &= \int_0^t (\bar{a}(X_s^\epsilon) - \bar{a}(\hat{X}_s^\epsilon)) ds, \\ \mathcal{V}_3(t) &= \int_0^t (\bar{a}(\hat{X}_s^\epsilon) - \bar{a}(\bar{X}_s)) ds, \\ \mathcal{V}_4(t) &= \int_0^t (b(X_s^\epsilon) - b(\hat{X}_s^\epsilon)) dB_s, \\ \mathcal{V}_5(t) &= \int_0^t (b(\hat{X}_s^\epsilon) - b(\bar{X}_s)) dB_s, \\ \mathcal{V}_6(t) &= \int_0^t (c(X_s^\epsilon) - c(\hat{X}_s^\epsilon)) d\hat{P}_s, \\ \mathcal{V}_7(t) &= \int_0^t (c(\hat{X}_s^\epsilon) - c(\bar{X}_s)) d\hat{P}_s, \\ \mathcal{V}_8(t) &= \lambda_1 \int_0^t (c(X_s^\epsilon) - c(\hat{X}_s^\epsilon)) ds, \\ \mathcal{V}_9(t) &= \lambda_1 \int_0^t (c(\hat{X}_s^\epsilon) - c(\bar{X}_s)) ds. \end{aligned}$$

By Hölder's inequality, the Burkholder's inequality (cf.[7]) and (3.24) we obtain

$$\begin{aligned} \mathbb{E} \left(\sum_{i=2,4,6,8} \sup_{0 \leq t \leq T_0} \|\mathcal{V}_i(t)\|^{2p} \right) &\leq C_{p,\lambda_1,T_0} \mathbb{E} \sup_{0 \leq t \leq T_0} \int_0^t \|X_s^\epsilon - \hat{X}_s^\epsilon\|^{2p} ds \\ &\leq C_{p,\lambda_1,T_0} \mathbb{E} \sup_{0 \leq t \leq T_0} \|X_t^\epsilon - \hat{X}_t^\epsilon\|^{2p} \\ &\leq C_{p,\lambda_1,\lambda_2,T_0} \left(\delta^p + \frac{\delta^{p+1}}{\epsilon} e^{C_{p,\lambda_2} \frac{\delta}{\epsilon}} \right). \end{aligned} \quad (4.12)$$

Similarly, we have for any $u \in [0, T_0]$

$$\begin{aligned} \mathbb{E} \left(\sum_{i=3,5,7,9} \sup_{0 \leq t \leq u} \|\mathcal{V}_i(t)\|^{2p} \right) &\leq C_{p,\lambda_1,T_0} \mathbb{E} \sup_{0 \leq t \leq u} \int_0^t \|\hat{X}_s^\epsilon - \bar{X}_s\|^{2p} ds \\ &\leq C_{p,\lambda_1,T_0} \int_0^u \mathbb{E} \sup_{0 \leq r \leq s} \|\hat{X}_r^\epsilon - \bar{X}_r\|^{2p} ds. \end{aligned} \quad (4.13)$$

From (4.6), (4.11), (4.12) and (4.13), we have for all $u \in [0, T_0]$,

$$\begin{aligned} \mathbb{E} \sup_{0 \leq t \leq u} \|\hat{X}_t^\epsilon - \bar{X}_t\|^{2p} &\leq C_{p,\lambda_1,\lambda_2,T_0} \left(\delta^p + \frac{\epsilon}{\delta} + \frac{\delta^{p+1}}{\epsilon} e^{C_{p,\lambda_2} \frac{\delta}{\epsilon}} \right) \\ &\quad + C_{p,\lambda_1,T_0} \int_0^u \mathbb{E} \sup_{0 \leq r \leq s} \|\hat{X}_r^\epsilon - \bar{X}_r\|^{2p} ds, \end{aligned}$$

which, with aid of Gronwall's inequality, yields

$$\mathbb{E} \sup_{0 \leq t \leq T_0} \|\hat{X}_t^\epsilon - \bar{X}_t\|^{2p} \leq C_{p,\lambda_1,\lambda_2,T_0} \left(\delta^p + \frac{\epsilon}{\delta} + \frac{\delta^{p+1}}{\epsilon} e^{C_{p,\lambda_2} \frac{\delta}{\epsilon}} \right).$$

The proof is completed. $\square$

With Lemma 3.5 and Lemma 4.3 in hand, we now formulate the main results in this paper.

Theorem 4.1. *Suppose that **(H1)** – **(H5)** hold, then we have for any $T_0 > 0$ and $p > 1$,*

$$\lim_{\epsilon \rightarrow 0} \mathbb{E} \sup_{0 \leq t \leq T_0} \|X_t^\epsilon - \bar{X}_t\|^{2p} = 0. \quad (4.14)$$

Proof. Thanks to (3.24) and (4.10) we obtain

$$\begin{aligned} \mathbb{E} \sup_{0 \leq t \leq T_0} \|X_t^\epsilon - \bar{X}_t\|^{2p} &\leq 2\mathbb{E} \sup_{0 \leq t \leq T_0} \|X_t^\epsilon - \hat{X}_t^\epsilon\|^{2p} + 2\mathbb{E} \sup_{0 \leq t \leq T_0} \|\hat{X}_t^\epsilon - \bar{X}_t\|^{2p} \\ &\leq C_{\gamma,T_0} \left(\delta^p + \frac{\epsilon}{\delta} + \frac{\delta^{p+1}}{\epsilon} e^{C_{p,T_0} \frac{\delta}{\epsilon}} \right). \end{aligned}$$

Taking $\delta = \epsilon(-\ln \epsilon)^{\frac{1}{2}}$ in above inequality, we get

$$\lim_{\epsilon \rightarrow 0} \mathbb{E} \sup_{0 \leq t \leq T_0} \|X_t^\epsilon - \bar{X}_t\|^{2p} = 0,$$

which completes the proof. $\square$

Remark 4.2. *It should be noted that we have confined ourselves to the case the diffusion coefficient of the slow variable do not depend on the fast variable, i.e., $b(x, y) = b(x)$ and $c(x, y) = c(x)$. In fact, a simple example (see[15]) can show that for slow-fast system, strong convergence does not hold where the noise coefficient of the slow variable depends on fast variable.*

Remark 4.3. *From the proof of Theorem 4.1, the rate of convergence scale like $(-\ln \epsilon)^{-\frac{1}{2}}$ is determined as a byproduct.*

By using the procedure similar to that of Theorem 4.1, we have the following result.

Theorem 4.2. *If we replace conditions (H3)–(H4) in Theorem 4.1 by the following:*

$$\begin{aligned} & (y_1 - y_2, f(x, y_1) - f(x, y_2) + \lambda_2(g(x, y_1) - g(x, y_2))) \\ & \quad + \|g(x, y_1) - g(x, y_2)\|^2 + \lambda_2\|h(x, y_1) - h(x, y_2)\|^2 \\ & \leq -\beta\|y_1 - y_2\|^2, \end{aligned} \tag{4.15}$$

for any (x, y_1, y_2) , where β is a positive constant. Then Eq.(4.14) still holds, and the rate of convergence is $O(\epsilon^{1/2})$.

Remark 4.4. *Theorem 4.2 shows that if conditions in Theorem 4.1 can be strengthened, we may get faster rate of convergence, which is shown to be optimal in L^2 (see [21]).*

Acknowledgments

The first author is very grateful to Professor Jinqiao Duan and Professor Yunmin Zhu for helpful discussions and comments. The authors would also like to thank the anonymous reviewers for their careful reading of manuscript, correcting errors, detailed comments and valuable suggestions, which improve the quality of this paper.

References

References

- [1] S. Cerrai, M.I. Freidlin, Averaging principle for a class of stochastic reaction-diffusion equations, *Proba. Theory Related Fields.* 144 (2009) 137-177.

- [2] S. Cerrai, A Khasminkii type averaging principle for stochastic reaction-diffusion equations, *Ann. Appl. Probab.* 19(3) (2009) 899-948.
- [3] S. Cerrai, Stochastic reaction-diffusion systems with multiplicative noise and non-Lipschitz reaction term, *Probab. Theory Related Fields.* 125 (2003) 271-304.
- [4] M.I. Freidlin, A.D. Wentzell, *Random Perturbation of Dynamical Systems*, Second ed., Springer-Verlag, New York, 1998.
- [5] M.I. Freidlin, A.D. Wentzell, Long-time behavior of weakly coupled oscillators, *J. Stat. Phys.* 123 (2006) 1311-1337.
- [6] H. Fu, J. Liu, Strong convergence in stochastic averaging for two time-scales stochastic partial differential equations, *J. Math. Anal. Appl.* 384 (2011) 70-86.
- [7] P. Protter, *Stochastic Integration and Differential Equations*, 2nd ed., Springer-Verlag, Berlin, 2004.
- [8] H. J. Kushner, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, 3, *Systems & Control: Foundations & Applications*. Boston: Birkhäuser, 1990.
- [9] Y. Kifer, Stochastic versions of Anosov and Neistadt theorems on averaging, *Stoch. Dyn.* 1(1) (2001) 1-21.
- [10] D. Givon, I.G. Kevrekidis, R. Kupferman, Strong convergence of projective integration schemes for singular perturbed stochastic differential systems, *Comm. Math. Sci.* 4 (2006) 707-729.
- [11] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *J. Comput. Phys.* 22 (1976) 403-434.
- [12] Y. Cao, D. T. Gillespie, and L. R. Petzold, The slow-scale stochastic simulation algorithm, *J. Chem. Phys.* 122 (2005) 014116.
- [13] E. Vanden-Eijnden, Numerical techniques for multi-scale dynamical systems with stochastic effects, *Commun. Math. Sci.* 1 (2003) 377-384.

- [14] W. E, D. Liu and E. Vanden-Eijnden, Analysis of multiscale methods for stochastic differential equations, *Commun. Pure Appl. Math.* 58(1) (2005) 1544-1585.
- [15] D. Givon, Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems, *SIAM J. Multi. Model. Simul.* 6 (2) (2007) 577-594.
- [16] J. Golec, G. Ladde, Averaging principle and systems of singularly perturbed stochastic differential equations, *J. Math. Phys.* 31 (1990) 1116-1123.
- [17] J. Golec, Stochastic averaging principle for systems with pathwise uniqueness, *Stoch. Anal. Appl.* 13 (3) (1995) 307-322.
- [18] R. Z. Khasminskii, On the principle of averaging the Itô's stochastic differential equations, *kibernetika.* 4 (1968) 260-279 (in Russian).
- [19] B. Øksendal, *Stochastic Differential Equations*, sixth ed., Springer, Berlin, 2003.
- [20] D. Liu, Strong convergence of principle of averaging for multiscale stochastic dynamical systems, *Commun. Math. Sci.* 8 (4) (2010) 999-1020.
- [21] D. Liu, Strong convergence rate of principle of averaging for jump-diffusion processes, *Front. Math. China.* 7 (2) (2012) 305-320.
- [22] J. S. Nagumo, S. Arimoto, S. Yoshizawa, An active pulse transmission line simulating nerve axon, *Proc. IRE.* 50 (1962) 2061-2071.
- [23] A.Y. Veretennikov, On the Averaging principle for systems of stochastic differential equations, *Math. USSR-Sb.* 69 (1991) 271-284.
- [24] A.Y. Veretennikov, On large deviations in the averaging principle for SDEs with full dependence, *Ann. Probab.* 27 (1999) 284-296.
- [25] G. Wainrib, Double averaging principle of periodically forced slow-fast stochastic systems, *Electron. Commun. Probab.* 18 (51) (2013) 1-12. 5299-5304.