Averaging principle for a class of stochastic differential equations *

Wei Liu Xiaobin Sun Yingchao Xie

School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, 221116, China.

Abstract. In this paper we mainly investigate the averaging principle for a class of stochastic differential equations with slow and fast time-scales, where the drift coefficient in slow equation only satisfies the local Lipschitz condition. We prove that the slow component strongly converges to the solution of corresponding averaged equation, and the result is applicable to some slow-fast SDE models with polynomial coefficients.

Keywords: Averaging principle; Slow-fast system; Local Lipschitz; Invariant measure; Strong convergence.

AMS Subject Classification: 60H10; 34E13; 34D23

1 Introduction

In this paper, we mainly consider the following stochastic slow-fast system

\[
\begin{align*}
\frac{dX^\epsilon}{\epsilon} &= b(X^\epsilon_t, Y^\epsilon_t)dt + \sigma(X^\epsilon_t) dW^1_t, \quad X^\epsilon_0 = x, \\
\frac{dY^\epsilon}{\epsilon} &= \frac{1}{\epsilon} f(X^\epsilon_t, Y^\epsilon_t)dt + \frac{1}{\sqrt{\epsilon}} g(X^\epsilon_t, Y^\epsilon_t) dW^2_t, \quad Y^\epsilon_0 = y,
\end{align*}
\]

(1.1)

where \( \epsilon \) is a small and positive parameter describing the ratio of time scale between the slow component \( X^\epsilon_t \in \mathbb{R}^n \) and fast component \( Y^\epsilon_t \in \mathbb{R}^m \). The drift terms \( b(x, y) \in \mathbb{R}^n \) and \( f(x, y) \in \mathbb{R}^m \), diffusion terms \( \sigma(x) \in \mathbb{R}^{n \times d_1} \) and \( g(x, y) \in \mathbb{R}^{m \times d_2} \), \( W^1_t \) and \( W^2_t \) are mutually independent \( d_1 \) and \( d_2 \) dimensional standard Brownian motion on complete probability space \((\Omega, \mathcal{F}, \mathbb{P})\) with a filtration \( \{\mathcal{F}_t, t \geq 0\} \) satisfying the usual conditions.

Under some assumptions, we aim to show that \( X^\epsilon \) converges to \( \bar{X} \) in the strong sense, i.e. for some \( p > 0 \),

\[
\lim_{\epsilon \to 0} \mathbb{E} \left( \sup_{t \in [0,T]} |X^\epsilon_t - \bar{X}_t|^p \right) = 0 \quad (1.2)
\]

here \( \bar{X} \) is the solution of the following averaged equation,

\[
\begin{align*}
\frac{d\bar{X}_t}{t} &= \bar{b}(\bar{X}_t)dt + \sigma(\bar{X}_t) dW^1_t, \\
\bar{X}_0 &= x,
\end{align*}
\]

(1.3)

*This work is supported in part by NSFC (No. 11571147, 11601196, 11771187, 11822106, 11831014), the NSF of Jiangsu Province (No. BK20160004), NSF of the Higher Education Institutions of Jiangsu Province (No. 16KJB110006), the QingLan Project and the PAPD Project of Jiangsu Higher Education Institutions.

E-mails: weiliu@jsnu.edu.cn; xbsun@jsnu.edu.cn; ycxie@jsnu.edu.cn.
where \( \bar{b}(x) = \int_{\mathbb{R}^m} b(x, y) \mu^x(dy) \) and \( \mu^x \) denotes the unique invariant measure for the corresponding frozen equation (see Eq. (2.6) below).

Now we give an intuitive way to explain how average principle occurs. If the coefficients \( f(x, y) \equiv f(y) \) and \( g(x, y) \equiv g(y) \), we can write the fast component as

\[
Y_\epsilon^t = y + \frac{1}{\epsilon} \int_0^t f(Y_{\epsilon}^s) ds + \frac{1}{\sqrt{\epsilon}} \int_0^t g(Y_{\epsilon}^s) dW^2_s.
\]

Then we have

\[
Y_{\epsilon \epsilon}^t = y + \frac{1}{\epsilon} \int_0^{\epsilon t} f(Y_{\epsilon}^s) ds + \frac{1}{\sqrt{\epsilon}} \int_0^{\epsilon t} g(Y_{\epsilon}^s) dW^2_s
\]

\( (1.4) \)

\[
= y + \int_0^t f(Y_{\epsilon}^s) ds + \int_0^t g(Y_{\epsilon}^s) d\tilde{W}_r^2,
\]

\( (1.5) \)

where \( \tilde{W}_r^2 := \frac{1}{\sqrt{r}} W^2_r \) is also a Brownian motion.

Based on the uniqueness of solutions of the following frozen equation (cf. Eq. (2.6) for the general case):

\[
Y_t = y + \int_0^t f(Y_s) ds + \int_0^t g(Y_s) d\bar{W}_s^2,
\]

which is also assumed that there exists a unique invariant measure \( \mu = \lim_{t \to \infty} \mathbb{P} \circ (Y_t)^{-1} \).

Then for any \( \epsilon > 0, t > 0 \),

\[
\mathbb{P} \circ (Y_t^\epsilon)^{-1} = \mathbb{P} \circ (Y_t)^{-1},
\]

where \( \mathbb{P} \circ X^{-1} \) means the distribution of random variable \( X \). This implies for any fixed \( t > 0 \),

\[
\lim_{\epsilon \to 0} \mathbb{P} \circ (Y_t^\epsilon)^{-1} = \lim_{s \to \infty} \mathbb{P} \circ (Y_s)^{-1} = \mu.
\]

In other words, the asymptotic behavior of the law of \( Y_t^\epsilon \) when \( \epsilon \to 0 \) equals to the asymptotic behavior of the law of \( Y_t \) when \( t \to \infty \), which explains that

\[
\mathbb{E} b(x, Y_t^\epsilon) \to \int b(x, y) \mu(dy), \quad \epsilon \to 0.
\]

The theory of averaging principle has been intensively investigated by many researchers. Bogoliubov and Mitropolsky \cite{2} first studied the averaging principle for the deterministic systems. Then the theory of averaging principle for stochastic differential equations was firstly proved by Khasminskii in \cite{15} and further developed in e.g. \cite{3, 12, 13, 14, 16, 17, 21, 22, 24}. Since then, averaging principle for stochastic reaction-diffusion systems has become an active research area which attracted much attention (see \cite{1, 3, 4, 5, 6, 7, 9, 10, 11, 19, 20}).

This paper mainly focus on the averaging principle for a class of stochastic differential equations. Usually, the Lipschitz continuous or sublinear growth condition on the coefficients in slow equation is crucial in many research papers, for example, Liu proved the convergence in \( L^2 \) sense in \cite{16} under the assumptions of \( b \) and \( \sigma \) are bounded smooth with bounded derivative of any order. If \( b \) and \( \sigma \) are Lipschitz continuous and have linear growth, Givon also proved convergence in \( L^2 \) sense in \cite{12}. Golec studied the \( L^\alpha \) convergence for some \( \alpha > 0 \) in \cite{13} when \( b \) and \( \sigma \) are continuous and sublinear growth. For more references we refer to \cite{13, 17, 21, 22}. The main contribution of this work is to weaken the Lipschitz condition of the drift coefficient \( b \) to locally Lipschitz and prove the convergence in \( L^p \) sense for some \( p \geq 2 \). We here mention a recent paper by Xu, Liu and Miao in \cite{23}, where they prove the
$L^2$ convergence for two-time-scale SDEs with non-Lipschitz coefficients, but it can not cover the case of polynomial coefficients. In our case we can apply our results to some models with polynomial coefficients.

The proof of strong convergence result (1.2) is mainly based on the Khasminskii argument introduced in [15]. More precisely, we split the interval $[0, T]$ into some subintervals of size $\delta > 0$, then on each interval $[k\delta, (k+1)\delta)$, $k \geq 0$, we construct an auxiliary process $(\hat{X}^i_t, \hat{Y}^i_t)$ associated with the system (1.1). Then result (1.2) can be obtained by following two steps. Step 1, due to the non-Lipschitz condition on coefficient $b$, we will construct some stopping times inspired by the techniques used in [7] and control the error between $X^i_t$ and $\bar{X}_t$ when time $t$ before the stopping time, which will be done by controlling $|X^i_t - \bar{X}^i_t|$ and $|\bar{X}^i_t - \bar{X}_t|$ respectively. Step 2, we can control the remaining term of time $t$ after the stopping time by the priori estimates of solution.

The paper is organized as follows. In the next Section, we introduce some notations and assumptions and then formulate the main results. The Section 3 is devoted to show the proofs of our main results. Throughout the paper, $C$, $C_p$, $C_{p,T}$, $C_{T,R}$, $C_{T,x,y}$ and $C_{T,R,x,y}$ will denote positive constants which may change from line to line, where:

$C$ denotes positive constants which may change from line to line, where

$C_p$ depends on $p$, $C_{p,T}$ depends on $p$, $T$, $C_{T,R}$ depends on $T$, $R$, $C_{T,x,y}$ depends on $T$, $x$, $y$ and $C_{T,R,x,y}$ depends on $T$, $R$, $|x|$, $|y|$.

## 2 Main Results

Now we impose the following assumptions on the coefficients $b, \sigma, f$ and $g$, here we use $| \cdot |$ to denote the Euclidean vector norm and $\| \cdot \|$ to denote the matrix norm.

$(H_1)$ The functions $b(x, y), \sigma(x)$ satisfy the local Lipschitz condition w.r.t. $x$ and Lipschitz condition w.r.t. $y$, i.e., for any $R > 0$ there exists a constant $K(R)$ depending on $R$ such that for $x_1, x_2 \in \mathbb{R}^n, y \in \mathbb{R}^m$ with $|x_i| \leq R, i = 1, 2,$

$$|b(x_1, y) - b(x_2, y)| + \|\sigma(x_1) - \sigma(x_2)\| \leq K(R)|x_1 - x_2|$$

and there exists a constant $C$ such that for any $x \in \mathbb{R}^n, y_1, y_2 \in \mathbb{R}^m$,

$$|b(x, y_1) - b(x, y_2)| \leq C|y_1 - y_2|.$$  

Furthermore, there exist constants $C, \lambda \geq 0$ and $\gamma \geq 1$ such that for any $x \in \mathbb{R}^n, y \in \mathbb{R}^m$,

$$2\langle x, b(x, y) \rangle + \|\sigma(x)\|^2 \leq C(1 + |x|^2) + \lambda|y|^2$$

and

$$|b(x, y)| \leq C(1 + |x|^\gamma + |y|), \quad \|\sigma(x)\| \leq C(1 + |x|).$$

$(H_2)$ The functions $f(x, y), g(x, y)$ satisfy Lipschitz condition w.r.t. $(x, y)$, i.e., there exists $C > 0$ such that for any $x_1, x_2 \in \mathbb{R}^n, y_1, y_2 \in \mathbb{R}^m$,

$$|f(x_1, y_1) - f(x_2, y_2)| + \|g(x_1, y_1) - g(x_2, y_2)\| \leq C(|x_1 - x_2| + |y_1 - y_2|).$$

$(A_k)$ For fixed $k \geq 2$, there exists $\beta_k > 0$ such that for any $x \in \mathbb{R}^n, y_1, y_2 \in \mathbb{R}^m$,

$$2\langle y_1 - y_2, f(x, y_1) - f(x, y_2) \rangle + (k - 1)||g(x, y_1) - g(x, y_2)||^2 \leq -\beta_k|y_1 - y_2|^2. \quad (2.1)$$
Remark 2.1 Here we give some comments concerning the assumptions above.

- Assumptions (H₁) and (H₂) ensure that (1.1) has a unique solution.
- The condition (A₂) is called the strong dissipative condition, which is used to guarantee that there exists a unique invariant measure for frozen equation (see Eq. (2.6) below) and the exponential ergodicity also holds (see Theorem 3.6 below).
- If \( k_1 > k_2 \geq 2 \), then (A_k₁) implies (A_k₂).
- Under the assumption of (H₂), if (A_k) holds, it is easy to obtain that there exist \( C \geq 0 \) and \( \beta > 0 \) such that for any \( x \in \mathbb{R}^n, y \in \mathbb{R}^m \)

\[
2\langle y, f(x, y) \rangle + (k - 1)\|g(x, y)\|^2 \leq -\beta|y|^2 + C|x|^2 + C.
\]

The following theorem is the existence and uniqueness result for system (1.1).

**Theorem 2.1** Suppose that the assumptions (H₁) and (H₂) hold. For any given initial value \( x \in \mathbb{R}^n, y \in \mathbb{R}^m \), there exists a unique solution \( \{(X_t^\epsilon, Y_t^\epsilon), t \geq 0\} \) to system (1.1) and for all \( T > 0, (X^\epsilon, Y^\epsilon) \in C([0, T]; \mathbb{R}^n) \times C([0, T]; \mathbb{R}^m) \), \( \mathbb{P} \)-a.s. and

\[
\begin{aligned}
X_t^\epsilon &= x + \int_0^t b(X_s^\epsilon, Y_s^\epsilon) ds + \int_0^t \sigma(X_s^\epsilon) dW_s^1, \\
Y_t^\epsilon &= y + \frac{1}{\epsilon} \int_0^t f(X_s^\epsilon, Y_s^\epsilon) ds + \frac{1}{\sqrt{\epsilon}} \int_0^t g(X_s^\epsilon, Y_s^\epsilon) dW_s^2.
\end{aligned}
\tag{2.2}
\]

Now we formulate our main result of averaging principle.

**Theorem 2.2** Suppose that (H₁) and (H₂) hold.

(i) If \( \lambda = 0 \) in (H₁) and (A₂) holds, then for any \( p > 0 \) we have

\[
\lim_{\epsilon \to 0} \mathbb{E} \left( \sup_{t \in [0, T]} |X_t^\epsilon - \bar{X}_t|^p \right) = 0.
\tag{2.3}
\]

(ii) If \( \lambda > 0 \) in (H₁) and (A_k) with \( k \geq 2\gamma \) holds, then for any \( 0 < p < k \) we have

\[
\lim_{\epsilon \to 0} \mathbb{E} \left( \sup_{t \in [0, T]} |X_t^\epsilon - \bar{X}_t|^p \right) = 0.
\tag{2.4}
\]

Here \( \bar{X} \) is the solution of the following averaged equation,

\[
\begin{aligned}
d\bar{X}_t &= \bar{b}(\bar{X}_t) dt + \sigma(\bar{X}_t) dW_t^1, \\
\bar{X}_0 &= x,
\end{aligned}
\tag{2.5}
\]

where \( \bar{b}(x) = \int_{\mathbb{R}^m} b(x, y) \mu^\tau(dy) \) and \( \mu^\tau \) denotes the unique invariant measure for the following frozen equation,

\[
\begin{aligned}
dY_t &= f(x, Y_t) dt + g(x, Y_t) d\bar{W}_t^2, \\
Y_0 &= y,
\end{aligned}
\tag{2.6}
\]

where \( \bar{W}_t^2 \) is a \( d_2 \)-dimensional Brownian motion independent of \( W_t^1 \) and \( W_t^2 \).

We can give an simple example to illustrate the application of our result.
Example 2.1 Let us consider the following slow-fast SDEs,
\[
\begin{align*}
    dX^t &= -(X^t)^3 dt + \sin(Y^t) dt + \theta Y^t dt + X^t dW^1_t, \quad X^0_t = x \in \mathbb{R} \\
    dY^t &= \frac{1}{\epsilon} (-3Y^t + X^t) dt + \frac{1}{\sqrt{\epsilon}} [3 + \sin(X^t) + \sin(Y^t)] dW^2_t, \quad Y^0_t = y \in \mathbb{R},
\end{align*}
\]
where \( \theta \geq 0, W^1_t \) and \( W^2_t \) are independent 1-dimensional Brownian motion. Put
\[
    b(x, y) = -x^3 + \sin y + \theta y, \quad \sigma(x) = x
\]
and
\[
    f(x, y) = -3y + x, \quad g(x, y) = 3 + \sin x + \sin y.
\]
It is easy to verify that \((H_1)\) with \( \gamma = 3\), \((H_2)\) and \((A_k)\) with any \( 2 \leq k < 7 \) hold.

Hence, \((H_1)\) holds with \( \lambda = 0 \) if \( \theta = 0 \), then by Theorem 2.2 for any \( p > 0 \) we have
\[
    \lim_{\epsilon \to 0} \mathbb{E} \left( \sup_{t \in [0, T]} |X^t - \bar{X}_t|^p \right) = 0.
\]
Moreover, \((H_1)\) holds with \( \lambda > 0 \) if \( \theta > 0 \), then by Theorem 2.2 for any \( 0 < p < 7 \) we have
\[
    \lim_{\epsilon \to 0} \mathbb{E} \left( \sup_{t \in [0, T]} |X^t - \bar{X}_t|^p \right) = 0.
\]

3 Proofs of Main Results
The existence and uniqueness of solutions to (1.1) can be obtained by using the classical result due to Krylov and Rozovskii (cf. [18]).

3.1 Proof of Theorem 2.1
We denote
\[
    Z^\epsilon_t = \begin{pmatrix} X^\epsilon_t \\ Y^\epsilon_t \end{pmatrix}, \quad \tilde{b}^\epsilon(x, y) = \begin{pmatrix} b(x, y) \\ \frac{1}{\epsilon} f(x, y) \end{pmatrix}
\]
and
\[
    \tilde{\sigma}^\epsilon(x, y) = \begin{pmatrix} \sigma(x) \\ 0 \\ \frac{1}{\sqrt{\epsilon}} g(x, y) \end{pmatrix}, \quad W_t = \begin{pmatrix} W^1_t \\ W^2_t \end{pmatrix}.
\]
Then system (1.1) can be rewritten as the following equation:
\[
    dZ^\epsilon_t = \tilde{b}^\epsilon(Z^\epsilon_t) dt + \tilde{\sigma}^\epsilon(Z^\epsilon_t) dW_t, \quad Z^\epsilon_0 = \begin{pmatrix} x \\ y \end{pmatrix}.
\]
Under the assumptions \((H_1)\) and \((H_2)\), it is easy to prove that \( \tilde{b}^\epsilon \) and \( \tilde{\sigma}^\epsilon \) satisfy the local Lipschitz conditions and there exists a positive constant \( C \) depending on \( \epsilon \), such that for any \( z \in \mathbb{R}^{n+m}, \)
\[
    2\langle z, \tilde{b}^\epsilon(z) \rangle + \| \tilde{\sigma}^\epsilon(z) \|^2 \leq C(1 + |z|^2),
\]
which imply the local weak monotonicity and weak coercivity conditions in [18]. Hence by [18, Theorem 3.1], there exists a unique solution \( \{(X^\epsilon_t, Y^\epsilon_t), t \geq 0\} \) to equation (1.1). The proof is complete.
In the rest of this section is devoted to prove Theorem 2.2. The proof consists of the following steps. In the first step, we give some a priori estimates of the solution \((X_t^\varepsilon, Y_t^\varepsilon)\) to the system (1.1). In the second step, following the idea inspired by Khasminskii in [15], we introduce an auxiliary process \((\hat{X}_t^\varepsilon, \hat{Y}_t^\varepsilon)\) and derive some uniform bounds. Meanwhile, making use of stopping time techniques, we also deduce some estimate of the (difference) process \(X_t^\varepsilon - \hat{X}_t^\varepsilon\) when time \(t\) before the stopping time. In the last step, based on the ergodicity property of the frozen equation, we first obtain appropriate control of \(\hat{X}_t^\varepsilon - \hat{X}_t\) in case of time \(t\) before the stopping time. Then we shall use the priori estimates of the solution to control the term of time \(t\) after the stopping time. Note that we always assume \((H_1)\) and \((H_2)\) hold in this section.

### 3.2 Some priori estimates of \((X_t^\varepsilon, Y_t^\varepsilon)\)

Firstly, we prove some uniform bounds with respect to \(\varepsilon \in (0,1)\) for the moment of the solutions \(X_t^\varepsilon\) and \(Y_t^\varepsilon\) to system (1.1).

**Lemma 3.1** (i) If \(\lambda = 0\) in \((H_1)\) and \((A_2)\) holds, then for any \(p > 0\), there exist positive constants \(C_{p,T}, C_T\) such that

\[
\sup_{\varepsilon \in (0,1)} \mathbb{E}\left(\sup_{t \in [0,T]} |X_t^\varepsilon|^p\right) \leq C_{p,T}(1 + |x|^p + |y|^p)
\]

and

\[
\sup_{\varepsilon \in (0,1)} \sup_{t \in [0,T]} \mathbb{E}|Y_t^\varepsilon|^2 \leq C_T(1 + |x|^2 + |y|^2).
\]

(ii) If \(\lambda > 0\) in \((H_1)\) and \((A_k)\) with \(k \geq 2\) holds, then there exists positive constant \(C_T\) such that

\[
\sup_{\varepsilon \in (0,1)} \mathbb{E}\left(\sup_{t \in [0,T]} |X_t^\varepsilon|^k\right) \leq C_T(1 + |x|^k + |y|^k)
\]

and

\[
\sup_{\varepsilon \in (0,1)} \sup_{t \in [0,T]} \mathbb{E}|Y_t^\varepsilon|^k \leq C_T(1 + |x|^k + |y|^k).
\]

**Proof.** (i) According to Itô’s formula and \((H_1)\) holds with \(\lambda = 0\), we have

\[
|X_t^\varepsilon|^2 = |x|^2 + 2\int_0^t \langle X_s^\varepsilon, b(X_s^\varepsilon, Y_s^\varepsilon) \rangle ds + \int_0^t \|\sigma(X_s^\varepsilon)\|^2 ds + 2\int_0^t \langle X_s^\varepsilon, \sigma(X_s^\varepsilon) dW_s^1 \rangle \\
\leq |x|^2 + C(1 + |X_s^\varepsilon|^2) ds + 2\int_0^t \langle X_s^\varepsilon, \sigma(X_s^\varepsilon) dW_s^1 \rangle.
\]

Then by Burkholder-Davis-Gundy’s inequality, for any \(p > 0\) large enough we have

\[
\mathbb{E}\left(\sup_{t \in [0,T]} |X_t^\varepsilon|^p\right) \leq C_p|x|^p + C_{p,T} \int_0^T \mathbb{E}(1 + |X_s^\varepsilon|^p) ds + CE\left[\int_0^T |X_s^\varepsilon|^2 \|\sigma(X_s^\varepsilon)\|^2 ds\right]^{p/4} \\
\leq C_{p,T}(|x|^p + 1) + C_{p,T} \int_0^T \mathbb{E}\left(\sup_{r \in [0,s]} |X_r^\varepsilon|^p\right) ds.
\]
Hence, Grownall’s inequality yields that
\[ \mathbb{E} \left( \sup_{t \in [0,T]} |X'_t|^p \right) \leq C_{p,T}(|x|^p + 1). \]

By Itô’s formula again we have
\[ \mathbb{E}|Y'_t|^2 = |y|^2 + 2\varepsilon \int_0^t \langle f(X'_s, Y'_s), Y'_s \rangle ds + \frac{1}{\varepsilon} \mathbb{E} \int_0^t \|g(X'_s, Y'_s)\|^2 ds. \]
Then, by \((A_2)\) and Remark \(2.1\) there exists \(\beta > 0\),
\[ \frac{d}{dt} \mathbb{E}|Y'_t|^2 = \frac{2}{\varepsilon} \mathbb{E} \left( \langle f(X'_t, Y'_t), Y'_t \rangle \right) + \frac{1}{\varepsilon} \mathbb{E} \|g(X'_t, Y'_t)\|^2 \leq -\frac{\beta}{\varepsilon} \mathbb{E}|Y'_t|^2 + \frac{C}{\varepsilon} \mathbb{E}|X'_t|^2 + \frac{C}{\varepsilon}. \]
Hence, by comparison theorem we obtain
\[ \mathbb{E}|Y'_t|^2 \leq |y|^2 e^{-\frac{\beta t}{\varepsilon}} + \frac{C}{\varepsilon} \int_0^t e^{-\frac{\beta (t-s)}{\varepsilon}} \left( 1 + \mathbb{E}|X'_s|^2 \right) ds \leq C_T(1 + |x|^2 + |y|^2), \]
which gives the statement \((i)\).

\(\text{(ii).}\) Notice that \((H_1)\) holds with \(\lambda > 0\), by Itô’s formula and Burkholder-Davis-Gundy’s inequality, we can obtain for any \(t \in [0,T]\),
\[ \mathbb{E} \left( \sup_{s \in [0,t]} |X''_s|^k \right) \leq C_T(|x|^k + 1) + C_T \int_0^t \mathbb{E} \left( \sup_{r \in [0,s]} |X''_r|^k \right) ds + C_T \int_0^t \mathbb{E}|Y''_s|^k ds. \]
Then Grownall’s inequality yields that
\[ \mathbb{E} \left( \sup_{s \in [0,t]} |X''_s|^k \right) \leq C_T(|x|^k + 1) + C_T \int_0^t \mathbb{E}|Y''_s|^k ds. \tag{3.9} \]
Using Itô formula again, we have
\[ \mathbb{E}|Y''_t|^k = |y|^k + \frac{k}{\varepsilon} \int_0^t |Y'_s|^{k-2} \langle f(X'_s, Y'_s), Y'_s \rangle ds + \frac{k}{2\varepsilon} \mathbb{E} \int_0^t |Y'_s|^{k-2} \|g(X'_s, Y'_s)\|^2 ds \]
\[ + \frac{k(k-2)}{2\varepsilon} \mathbb{E} \int_0^t |Y'_s|^{k-4} \left| \langle Y'_s, g(X'_s, Y'_s) \rangle \right|^2 ds. \]
\((A_k)\) yields that there exist \(\beta > 0\) and \(C \geq 0\) such that
\[ \frac{d}{dt} \mathbb{E}|Y''_t|^k \leq \frac{k}{2\varepsilon} \mathbb{E} \left[ |Y''_t|^{k-2} \left( 2 \langle f(X'_t, Y'_t), Y'_t \rangle + (k-1)\|g(X'_t, Y'_t)\|^2 \right) \right] \]
\[ \leq -\frac{\beta}{\varepsilon} \mathbb{E}|Y''_t|^k + \frac{C}{\varepsilon} \left( \mathbb{E}|X'_t|^k + 1 \right). \]
Hence, by comparison theorem and (3.9), for any $t \in [0, T]$ we have

$$E|Y_t^\epsilon|^k \leq |y|^k e^{-\frac{\alpha t}{c}} + C_T \int_0^t e^{-\frac{\alpha(s-t)}{c}} \left(1 + |x|^k + \int_0^s E|Y_s^\epsilon|^k\right) dr ds$$

$$\leq C_T(1 + |x|^k + |y|^k) + \int_0^t E|Y_s^\epsilon|^k ds.$$  

This implies that

$$\sup_{\epsilon \in (0, 1)} \sup_{t \in [0, T]} E|Y_t^\epsilon|^k \leq C_T(1 + |x|^k + |y|^k),$$

which also gives

$$\sup_{\epsilon \in (0, 1)} \sup_{t \in [0, T]} |X_t^\epsilon|^k \leq C_T(1 + |x|^k + |y|^k).$$

The proof is complete.

Lemma 3.2 Assume either (H1) with $\lambda = 0$ and (A2) hold or (H1) with $\lambda > 0$ and (A2\gamma) hold. Then for any $x \in \mathbb{R}^n, y \in \mathbb{R}^m, t \geq 0, 0 < t \leq t + h \leq T$, then there exists a positive constant $C_{T,x,y}$ such that

$$\sup_{\epsilon \in (0, 1)} E|X_{t+h}^\epsilon - X_t^\epsilon|^2 \leq C_{T,x,y} h.$$  

Proof. By a simple calculations, we have

$$X_{t+h}^\epsilon - X_t^\epsilon = \int_t^{t+h} b(X_s^\epsilon, Y_s^\epsilon) ds + \int_t^{t+h} \sigma(X_s^\epsilon) dW_s^1.$$  

Then by condition (H1) and Lemma 3.1

$$E|X_{t+h}^\epsilon - X_t^\epsilon|^2 \leq 2E \left( \int_t^{t+h} b(X_s^\epsilon, Y_s^\epsilon) ds \right)^2 + 2E \left( \int_t^{t+h} \sigma(X_s^\epsilon) dW_s^1 \right)^2$$

$$\leq C \int_t^{t+h} E|b(X_s^\epsilon, Y_s^\epsilon)|^2 ds + C \int_t^{t+h} E\|\sigma(X_s^\epsilon)\|^2 ds$$

$$\leq C \int_t^{t+h} E(1 + |X_s^\epsilon|^2 + |Y_s^\epsilon|^2) ds + C \int_t^{t+h} E(1 + |X_s^\epsilon|^2) ds$$

$$\leq C_{T,x,y} h.$$  

The proof is complete.

3.3 Estimates of auxiliary process ($\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon$)

Following the idea inspired by Khasminskii in [15], we introduce an auxiliary process ($\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon$) $\in \mathbb{R}^n \times \mathbb{R}^m$ and divide $[0, T]$ into intervals of size $\delta$, where $\delta$ is a fixed positive number and will be chosen by $\delta = \epsilon^{1/2}$ later. We construct a process $\hat{Y}_t^\epsilon$ with initial value $\hat{Y}_0^\epsilon = Y_0^\epsilon = y$, and for $t \in [k\delta, \min((k+1)\delta, T)]$,

$$\hat{Y}_t^\epsilon = \hat{Y}_{k\delta}^\epsilon + \frac{1}{\epsilon} \int_{k\delta}^t f(X_s^\epsilon, \hat{Y}_s^\epsilon) ds + \frac{1}{\sqrt{\epsilon}} \int_{k\delta}^t g(X_s^\epsilon, \hat{Y}_s^\epsilon) dW_s^2,$$
\[ \hat{Y}_t^\epsilon = y + \frac{1}{\epsilon} \int_0^t f(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon)ds + \frac{1}{\sqrt{\epsilon}} \int_0^t g(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon)dW_s^2. \]

where \( s(\delta) = [s/\delta]\delta \) is the nearest breakpoint proceeding \( s \). Also, we define the process \( \hat{X}_t^\epsilon \) by
\[ \hat{X}_t^\epsilon = x + \int_0^t b(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon)ds + \int_0^t \sigma(X_s^\epsilon)W_s^1. \]

We remark that on each intervals the fast component \( \hat{Y}_s^\epsilon \) does not depend on the slow component \( \hat{X}_s^\epsilon \), but only on the value of \( X_t^\epsilon \) at the initial point of underlying interval.

By the construction of \( \hat{Y}_t^\epsilon \) and Lemma 3.1, it is easy to obtain the following estimate and we omit the proof here.

**Lemma 3.3** Suppose that \((A_2)\) holds. Then there exists a constant \( C_T > 0 \) such that
\[ \sup_{\epsilon \in (0,1)} \sup_{t \in [0,T]} \mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \leq C_T(1 + |x|^2 + |y|^2). \]

Now, we will control the error of \( Y_t^\epsilon - \hat{Y}_t^\epsilon \) and furthermore the error of \( X_t^\epsilon - \hat{X}_t^\epsilon \).

**Lemma 3.4** Assume either \((H_1)\) with \( \lambda = 0 \) and \((A_2)\) hold or \((H_1)\) with \( \lambda > 0 \) and \((A_2, \gamma)\) hold. Then for any \( x \in \mathbb{R}^n, y \in \mathbb{R}^m, T > 0 \), there exists a constant \( C_{T,x,y} > 0 \) such that
\[ \sup_{\epsilon \in (0,1)} \sup_{t \in [0,T]} \mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \leq C_{T,x,y}. \]

**Proof.** Notice that
\[ Y_t^\epsilon - \hat{Y}_t^\epsilon = \frac{1}{\epsilon} \int_0^t \left[ f(X_s^\epsilon, Y_s^\epsilon) - f(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon) \right] ds + \frac{1}{\sqrt{\epsilon}} \int_0^t \left[ g(X_s^\epsilon, Y_s^\epsilon) - g(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon) \right] dW_s^2. \]

For any \( t \in [0, T] \), by Itô’s formula we have
\[
\frac{d}{dt} \mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 = 2 \mathbb{E} \left[ \langle f(X_t^\epsilon, Y_t^\epsilon) - f(X_{t(\delta)}^\epsilon, \hat{Y}_t^\epsilon), Y_t^\epsilon - \hat{Y}_t^\epsilon \rangle \right] \\
+ \frac{1}{\epsilon} \mathbb{E} \| g(X_t^\epsilon, Y_t^\epsilon) - g(X_{t(\delta)}^\epsilon, \hat{Y}_t^\epsilon) \|^2 \\
= \frac{1}{\epsilon} \mathbb{E} \left[ 2 \langle f(X_t^\epsilon, Y_t^\epsilon) - f(X_{t(\delta)}^\epsilon, \hat{Y}_t^\epsilon), Y_t^\epsilon - \hat{Y}_t^\epsilon \rangle \right] \\
+ \frac{1}{\epsilon} \mathbb{E} \langle f(X_t^\epsilon, \hat{Y}_t^\epsilon) - f(X_{t(\delta)}^\epsilon, \hat{Y}_t^\epsilon), Y_t^\epsilon - \hat{Y}_t^\epsilon \rangle \\
+ \frac{2}{\epsilon} \mathbb{E} \langle g(X_t^\epsilon, Y_t^\epsilon) - g(X_{t(\delta)}^\epsilon, \hat{Y}_t^\epsilon), g(X_t^\epsilon, \hat{Y}_t^\epsilon) - g(X_{t(\delta)}^\epsilon, \hat{Y}_t^\epsilon) \rangle \\
+ \frac{1}{\epsilon} \mathbb{E} \| g(X_t^\epsilon, \hat{Y}_t^\epsilon) - g(X_{t(\delta)}^\epsilon, \hat{Y}_t^\epsilon) \|^2.
\]
By condition \((A_2)\), we obtain
\[
\frac{d}{dt} \mathbb{E}|Y_t^\varepsilon - \hat{Y}_t^\varepsilon|^2 \leq -\frac{\beta_2}{\varepsilon} \mathbb{E}|Y_t^\varepsilon - \hat{Y}_t^\varepsilon|^2 + \frac{C}{\varepsilon} \mathbb{E}\left(|X_t^\varepsilon - X_{t(\delta)}^\varepsilon| : |Y_t^\varepsilon - \hat{Y}_t^\varepsilon|\right)
+ \frac{C}{\varepsilon} \mathbb{E}|X_t^\varepsilon - X_{t(\delta)}^\varepsilon|^2
\leq -\frac{\beta_2}{2\varepsilon} \mathbb{E}|Y_t^\varepsilon - \hat{Y}_t^\varepsilon|^2 + \frac{C}{\varepsilon} \mathbb{E}|X_t^\varepsilon - X_{t(\delta)}^\varepsilon|^2.
\]

Then by Lemma 3.2 and compare theorem,
\[
\mathbb{E}|Y_t^\varepsilon - \hat{Y}_t^\varepsilon|^2 \leq \frac{C_{T,x,y}}{\varepsilon} \int_0^t e^{-\frac{s(t-s)}{2\varepsilon}} ds
\leq C_{T,x,y},
\]
which complete the proof.

**Lemma 3.5** Suppose that \((A_2)\) holds. Then for any \(x \in \mathbb{R}^n, y \in \mathbb{R}^m, T > 0\) and \(R > 0\), there exists a constant \(C_{T,R,x,y} > 0\) such that
\[
\mathbb{E}\left(\sup_{t \in [0,T \wedge \tau_{R}]} |X_t^\varepsilon - X_t^\varepsilon| \right) \leq C_{T,R,x,y} \delta.
\]

Proof. Recall that
\[
X_t^\varepsilon = x + \int_0^t b(X_s^\varepsilon, Y_s^\varepsilon) ds + \int_0^t \sigma(X_s^\varepsilon) dW_s^1
\]
and
\[
\hat{X}_t^\varepsilon = x + \int_0^t b(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) ds + \int_0^t \sigma(X_s^\varepsilon) dW_s^1.
\]
Then we have
\[
X_t^\varepsilon - X_t^\varepsilon = \int_0^t [b(X_s^\varepsilon, Y_s^\varepsilon) - b(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon)] ds.
\]
By Lemmas 3.2 and 3.4 we obtain
\[
\mathbb{E}\left(\sup_{t \in [0,T \wedge \tau_{R}]} |X_t^\varepsilon - \hat{X}_t^\varepsilon| \right) \leq C_T \mathbb{E} \int_0^{T \wedge \tau_{R}} \left|b(X_s^\varepsilon, Y_s^\varepsilon) - b(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon)\right|^2 ds
\leq C_T \mathbb{E} \int_0^{T \wedge \tau_{R}} \left[K^2(R)|X_s^\varepsilon - X_{s(\delta)}^\varepsilon| + C|Y_s^\varepsilon - \hat{Y}_s^\varepsilon|^2\right] ds
\leq C_{T,R,x,y} \delta.
\]
The proof is complete.
3.4 The Frozen and averaged equation

We first introduce the frozen equation associate to the fast motion for fixed slow component $x \in \mathbb{R}^n$.

\[
\begin{align*}
  dY_t &= f(x, Y_t)dt + g(x, Y_t)d\tilde{W}_t^2, \\
  Y_0 &= y,
\end{align*}
\]

(3.10)

where $\tilde{W}_t^2$ is a $d_2$-dimensional Brownian motion independent of $W_t^1$ and $W_t^2$. If $(H_2)$ holds, then it is easy to prove for any fixed slow component $x \in \mathbb{R}^n$ and any initial data $y \in \mathbb{R}^m$, Eq. (3.10) has a unique strong solution $Y_{t,x,y}$, which is a Markov process. Let $P_t^x$ be the transition semigroup of $Y_{t,x,y}$. Under the assumption $(A_2)$, it is easy to prove that $\sup_{t\geq 0}E|Y_{t,x,y}|^2 \leq C(1 + |x|^2 + |y|^2)$ and that $P_t^x$ has unique invariant measure $\mu^x$ satisfying $\int_{\mathbb{R}^m}|y|\mu^x(dy) \leq C(|x| + 1)$ and the following exponential ergodicity also holds (see [16 Lemma A.3]).

**Lemma 3.6** Suppose that $(A_2)$ holds. For any given value $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, there exist $C > 0$ and $\beta > 0$ such that for any Lipschitz function $\varphi : \mathbb{R}^m \to \mathbb{R}$,

\[
|P_t^x \varphi(y) - \int_{\mathbb{R}^m} \varphi(z) \mu^x(dz)| \leq C(1 + |x| + |y|)e^{-\beta t} |\varphi|_{\text{Lip}},
\]

where $|\varphi|_{\text{Lip}} = \sup_{x,y \in \mathbb{R}^m} \frac{|\varphi(x) - \varphi(y)|}{|x-y|}$.

**Lemma 3.7** Suppose that $(A_2)$ holds. Then for any $x_1, x_2 \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $t > 0$, there exists a constant $C > 0$ such that

\[
E|Y_{t,x_1,y} - Y_{t,x_2,y}|^2 \leq C|x_1 - x_2|^2.
\]

Proof. Notice that

\[
Y_{t,x_1,y} - Y_{t,x_2,y} = \int_0^t f(x_1, Y_{s,x_1,y}^1) - f(x_2, Y_{s,x_2,y}^1)ds + \int_0^t g(x_1, Y_{s,x_1,y}^2) - g(x_2, Y_{s,x_2,y}^2)d\tilde{W}_s^2.
\]

By Itô’s formula and Young’s inequality, there exists $\beta > 0$ such that

\[
\begin{align*}
\frac{d}{dt}E|Y_{t,x_1,y} - Y_{t,x_2,y}|^2 &= E\left[2\langle f(x_1, Y_{t,x_1,y}^1) - f(x_2, Y_{t,x_2,y}^1), Y_{t,x_1,y}^1 - Y_{t,x_2,y}^1 \rangle + \|g(x_1, Y_{t,x_1,y}^1) - g(x_2, Y_{t,x_2,y}^2)\|^2 \right] \\
&= E\left[2\langle f(x_1, Y_{t,x_1,y}^1) - f(x_1, Y_{t,x_2,y}^1), Y_{t,x_1,y}^1 - Y_{t,x_1,y}^1 \rangle + \|g(x_1, Y_{t,x_1,y}^1) - g(x_1, Y_{t,x_2,y})\|^2 \right] \\
&\quad + E\left[2\langle f(x_1, Y_{t,x_2,y}^1) - f(x_2, Y_{t,x_2,y}^1), Y_{t,x_1,y}^2 - Y_{t,x_2,y}^2 \rangle \right] + E\|g(x_1, Y_{t,x_2,y}^2) - g(x_2, Y_{t,x_2,y}^2)\|^2 \\
&\leq -\beta E|Y_{t,x_1,y} - Y_{t,x_2,y}|^2 + C|x_1 - x_2|^2.
\end{align*}
\]

Hence, comparison theorem yields that

\[
E|Y_{t,x_1,y} - Y_{t,x_2,y}|^2 \leq C \int_0^t e^{-\beta(t-s)} ds |x_1 - x_2|^2 \leq C|x_1 - x_2|^2.
\]

The proof is complete.
Now, we can introduce the averaged equation as follows,
\[
\begin{aligned}
  \begin{cases}
    d\bar{X}_t = \bar{b}(\bar{X}_t)dt + \sigma(\bar{X}_t)dW_t^1, \\
    \bar{X}_0 = x \in \mathbb{R}^n,
  \end{cases}
\end{aligned}
\] (3.11)
where
\[
\bar{b}(x) = \int_{\mathbb{R}^m} b(x, y)\mu^x(dy),
\]
where \(\mu^x\) is the unique invariant measure for Eq. (3.10).

**Lemma 3.8** Suppose that (A2) holds. Eq. (3.11) has a unique solution. Moreover, for any \(x \in \mathbb{R}^n\), \(p \geq 2\) and \(T > 0\), there exists a constant \(C_{p,T} > 0\) such that
\[
\mathbb{E} \left( \sup_{t \in [0, T]} |\bar{X}_t|^p \right) \leq C_{p,T}(1 + |x|^p). \] (3.12)

Proof. It is sufficient to check that the coefficients of (3.11) satisfy the following conditions:

For any \(x_1, x_2 \in \mathbb{R}^n\), \(R > 0\) with \(|x_i| \leq R\), there exist positive constant \(K_1(R) < \infty\) such that
\[
|\bar{b}(x_1) - \bar{b}(x_2)| + \|\sigma(x_1) - \sigma(x_2)\| \leq K_1(R)|x_1 - x_2| \] (3.13)
and for any \(x \in \mathbb{R}^n\),
\[
2\langle \bar{b}(x), x \rangle + \|\sigma(x)\|^2 \leq C(1 + |x|^2). \] (3.14)

Then Eq. (3.11) has a unique solution and (3.12) can be easily obtained by following the same argument Lemma 3.1(i).

In fact, by Lemmas 3.6 and 3.7
\[
|\bar{b}(x_1) - \bar{b}(x_2)| + \|\sigma(x_1) - \sigma(x_2)\| \\
\leq \int_{\mathbb{R}^m} b(x_1, z)\mu^{x_1}(dz) - \int_{\mathbb{R}^m} b(x_2, z)\mu^{x_2}(dz) \bigg| + K(R)|x_1 - x_2| \\
\leq \int_{\mathbb{R}^m} b(x_1, z)\mu^{x_1}(dz) - \mathbb{E}b(x_1, Y_t^{x_1,y}) \bigg| + \mathbb{E}b(x_2, Y_t^{x_2,y}) - \int_{\mathbb{R}^m} b(x_2, z)\mu^{x_2}(dz) \bigg| + 2|\mathbb{E}[b(x_1, Y_t^{x_1,y}) - b(x_2, Y_t^{x_2,y})]| + K(R)|x_1 - x_2| \\
\leq C(1 + |x_1| + |x_2| + |y|)e^{-\beta t} + C(K(R) + 1)|x_1 - x_2|. 
\]

Then letting \(t \to \infty\), we obtain that (3.13) holds.

Moreover, for any \(x \in \mathbb{R}^n\),
\[
2\langle \bar{b}(x), x \rangle + \|\sigma(x)\|^2 \\
= 2 \left\langle \int_{\mathbb{R}^m} b(x, z)\mu^{x}(dz), x \right\rangle + \|\sigma(x)\|^2 \\
= 2 \left\langle \int_{\mathbb{R}^m} b(x, z)\mu^{x}(dz) - \mathbb{E}b(x, Y_t^{x,0}), x \right\rangle + 2\langle \mathbb{E}b(x, Y_t^{x,0}), x \rangle + \|\sigma(x)\|^2 \\
\leq C(1 + |x|^2)e^{-\beta t} + C(1 + |x|^2). 
\]

Finally, letting \(t \to \infty\), we obtain that (3.14) holds.
### 3.5 The Proof of Main Result

In this part, we intend to give a complete proof for our main result, i.e. the slow component process $X^t$ converges strongly to the solution $\bar X_t$ of the averaged equation. We first estimate the error between auxiliary process $\hat X_t$ and the solution $\bar X_t$ of averaged equation when time $t$ before the stopping time.

**Lemma 3.9** Suppose that either (H$_1$) with $\lambda = 0$ and (A$_2$) hold or (H$_1$) with $\lambda > 0$ and (A$_{2\gamma}$) hold. Then for any $x \in \mathbb{R}^n, y \in \mathbb{R}^m$, $R > 0, T > 0$ and $\epsilon \in (0, 1)$, there exists a positive constant $C_{T,R,x,y}$ such that

$$E \left( \sup_{t \in [0,T \land \tau^*_R]} |\hat X^t_t - \bar X_t|^2 \right) \leq C_{T,R,x,y} \left( \frac{\epsilon}{\delta} + \delta \right),$$

where $\tau^*_R := \inf \{ t \geq 0 : |X^t_t| + |\bar X_t| \geq R \}$.

Proof. Recall that

$$\hat X^t_t - \bar X_t = \int_0^t \left[ b(X^\epsilon_s(\delta), \dot Y^\epsilon_s) - \bar b(X^\epsilon_s) \right] ds + \int_0^t \left[ \sigma(X^\epsilon_s) - \sigma(X^\epsilon_s) \right] dW^1_s$$

$$= \int_0^t \left[ b(X^\epsilon_s(\delta), \dot Y^\epsilon_s) - \bar b(X^\epsilon_s(\delta)) \right] ds + \int_0^t \left[ \bar b(X^\epsilon_s) - \bar b(X^\epsilon_s) \right] ds$$

$$+ \int_0^t \left[ b(X^\epsilon_s) - \bar b(X^\epsilon_s) \right] ds + \int_0^t \left[ \sigma(X^\epsilon_s) - \sigma(X^\epsilon_s) \right] dW^1_s.$$

Then notice that $\tau^*_R \leq \tau^*_R$ and the local Lipschitz condition of $\bar b$, by Lemmas 3.2 and 3.5 there exists $C_{T,R} > 0$ such that

$$E \left( \sup_{t \in [0,T \land \tau^*_R]} |\hat X^t_t - \bar X_t|^2 \right)$$

$$\leq C_{T,R} \left[ \int_0^t b(X^\epsilon_s(\delta), \dot Y^\epsilon_s) - \bar b(X^\epsilon_s(\delta)) ds \right] + C_{T,R} \left[ \int_0^t \left| \bar b(X^\epsilon_s) - \bar b(X^\epsilon_s) \right|^2 ds \right]$$

$$+ C_{T,R} \left[ \int_0^t \left| b(X^\epsilon_s) - \bar b(X^\epsilon_s) \right|^2 ds \right] + C_{T,R} \left[ \int_0^t \left| \sigma(X^\epsilon_s) - \sigma(X^\epsilon_s) \right|^2 ds \right]$$

$$\leq C_{T,R} \left[ \int_0^t b(X^\epsilon_s(\delta), \dot Y^\epsilon_s) - \bar b(X^\epsilon_s(\delta)) ds \right] + C_{T,R} \left[ \int_0^t \left| X^\epsilon_s(\delta) - X^\epsilon_s(\delta) \right|^2 ds \right]$$

$$+ C_{T,R} \left[ \int_0^t \left| X^\epsilon_s - X^\epsilon_s \right|^2 ds \right] + C_{T,R} \left[ \int_0^t \left| X^\epsilon_s(\delta) - X^\epsilon_s(\delta) \right|^2 ds \right]$$

$$+ C_{T,R} \left[ \int_0^t \left( \sup_{s \in [0,t \land \tau^*_R]} |\hat X^t_s - \bar X_s|^2 \right) ds \right] + C_{T,R} \left[ \int_0^t \left( \sup_{s \in [0,t \land \tau^*_R]} |X^t_s - \bar X^t_s|^2 \right) ds \right]$$

$$\leq C_{T,R} \left[ \int_0^t b(X^\epsilon_s(\delta), \dot Y^\epsilon_s) - \bar b(X^\epsilon_s(\delta)) ds \right] + C_{T,R} \left[ \int_0^t \left| X^\epsilon_s(\delta) - X^\epsilon_s(\delta) \right|^2 ds \right]$$

$$+ C_{T,R} \left[ \int_0^t \left( \sup_{s \in [0,t \land \tau^*_R]} |\hat X^t_s - \bar X_s|^2 \right) ds \right] + C_{T,R} \left[ \int_0^t \left( \sup_{s \in [0,t \land \tau^*_R]} |X^t_s - \bar X^t_s|^2 \right) ds \right]$$

(3.15)
Next, we intend to estimate the first term above, notice that

\[
\left| \int_0^t \left[ b(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon) - \hat{b}(X_{s(\delta)}^\epsilon) \right] \, ds \right|^2 \\
= \left| \sum_{k=0}^{[t/\delta]-1} \int_{k\delta}^{(k+1)\delta} \left[ b(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon) - \hat{b}(X_{k\delta}^\epsilon) \right] \, ds + \int_{[t/\delta]}^t \left[ b(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon) - \hat{b}(X_{s(\delta)}^\epsilon) \right] \, ds \right|^2 \\
= 2[t/\delta] \sum_{k=0}^{[t/\delta]-1} \left| \int_{k\delta}^{(k+1)\delta} \left[ b(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon) - \hat{b}(X_{k\delta}^\epsilon) \right] \, ds \right|^2 + 2 \left| \int_{[t/\delta]}^t \left[ b(X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon) - \hat{b}(X_{s(\delta)}^\epsilon) \right] \, ds \right|^2 \\
:= I_1(t) + I_2(t). \quad (3.16)
\]

For $I_2(t)$, by Lemma 3.3, it is easy to prove

\[
\mathbb{E} \left[ \sup_{t \in [0,T]} I_2(t) \right] \leq C\delta \mathbb{E} \int_0^{T \wedge \hat{T}_R^\epsilon} \left( 1 + |X_{s(\delta)}^\epsilon|^2 + |\hat{Y}_s^\epsilon|^2 \right) \, ds \leq C_{T,R,x,y,\delta}. \quad (3.17)
\]

Now, we estimate the term $I_1(t)$ and have

\[
\mathbb{E} \left[ \sup_{t \in [0,T \wedge \hat{T}_R^\epsilon]} I_1(t) \right] \leq C[T/\delta] \mathbb{E} \sum_{k=0}^{[T/\delta]-1} \int_{k\delta}^{(k+1)\delta} \left| b(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon) - \hat{b}(X_{k\delta}^\epsilon) \right| \, ds \leq C_T \frac{\epsilon^2}{\delta^2} \max_{0 \leq k \leq [T/\delta]-1} \mathbb{E} \left| \int_{k\delta}^{(k+1)\delta} \left[ b(X_{k\delta}^\epsilon, \hat{Y}_s^\epsilon) - \hat{b}(X_{k\delta}^\epsilon) \right] \, ds \right|^2 \\
= C_T \frac{\epsilon^2}{\delta^2} \max_{0 \leq k \leq [T/\delta]-1} \mathbb{E} \left| \int_{0}^{\frac{T}{\delta}} \left[ b(X_{k\delta}^\epsilon, \hat{Y}_{s+k\delta}^\epsilon) - \hat{b}(X_{k\delta}^\epsilon) \right] \, ds \right|^2 \\
= C_T \frac{\epsilon^2}{\delta^2} \max_{0 \leq k \leq [T/\delta]-1} \int_{0}^{\frac{T}{\delta}} \int_{r}^{\frac{T}{\delta}} \Psi_k(s, r) \, ds \, dr,
\]

where for any $0 \leq r \leq s \leq \frac{\delta}{\epsilon},$

\[
\Psi_k(s, r) = \mathbb{E} \left[ \left| b(X_{k\delta}^\epsilon, \hat{Y}_{s+k\delta}^\epsilon) - \hat{b}(X_{k\delta}^\epsilon) \right| \right].
\]

By the construction of $\hat{Y}_s^\epsilon$, for any $k \in \mathbb{N}$ and $s \in [0, \delta)$ we obtain that

\[
\hat{Y}_{s+k\delta}^\epsilon = \hat{Y}_{k\delta}^\epsilon + \frac{1}{\epsilon} \int_{k\delta}^{k\delta+s} f(X_{k\delta}^\epsilon, \hat{Y}_{k\delta}^\epsilon) \, dr + \frac{1}{\sqrt{\epsilon}} \int_{k\delta}^{k\delta+s} g(X_{k\delta}^\epsilon, \hat{Y}_{k\delta}^\epsilon) \, dW_r^2 \\
= \hat{Y}_{k\delta}^\epsilon + \frac{1}{\epsilon} \int_{0}^{s/\epsilon} f(X_{k\delta}^\epsilon, \hat{Y}_{r+k\delta}^\epsilon) \, dr + \frac{1}{\sqrt{\epsilon}} \int_{0}^{s/\epsilon} g(X_{k\delta}^\epsilon, \hat{Y}_{r+k\delta}^\epsilon) \, dW_r^{k\delta}, \quad (3.18)
\]

where $W_r^{k\delta} := W_{r+k\delta}^2 - W_{k\delta}^2$ is the shift of $W_r^2.$

Recall that $W_t$ is a $d$-dimensional Brownian motion, which is independent of $(X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon).$ Denote by $\hat{W}_t = \sqrt{\epsilon} \hat{W}_{t/\epsilon}.$ We construct a process $Y_t^{X_{k\delta}^\epsilon, \hat{Y}_{k\delta}^\epsilon}$ by $Y_{t-}^{x,y} |_{(x,y)=(X_{k\delta}^\epsilon, \hat{Y}_{k\delta}^\epsilon)},$ i.e.,

\[
Y_{s/\epsilon}^{X_{k\delta}^\epsilon, \hat{Y}_{k\delta}^\epsilon} = \hat{Y}_{k\delta}^\epsilon + \int_{0}^{s/\epsilon} f(X_{k\delta}^\epsilon, Y_{r/\epsilon}^{X_{k\delta}^\epsilon, \hat{Y}_{k\delta}^\epsilon}) \, dr + \int_{0}^{s/\epsilon} g(X_{k\delta}^\epsilon, Y_{r/\epsilon}^{X_{k\delta}^\epsilon, \hat{Y}_{k\delta}^\epsilon}) \, d\hat{W}_r \\
= \hat{Y}_{k\delta}^\epsilon + \frac{1}{\epsilon} \int_{0}^{s} f(X_{k\delta}^\epsilon, Y_{r/\epsilon}^{X_{k\delta}^\epsilon, \hat{Y}_{k\delta}^\epsilon}) \, dr + \frac{1}{\sqrt{\epsilon}} \int_{0}^{s} g(X_{k\delta}^\epsilon, Y_{r/\epsilon}^{X_{k\delta}^\epsilon, \hat{Y}_{k\delta}^\epsilon}) \, d\hat{W}_r. \quad (3.19)
\]
The uniqueness of the solution of Eq. (3.18) and Eq. (3.19) implies that the distribution of \((X_{k\delta}^\epsilon, \hat{Y}_{s+k\delta}^\epsilon)\) coincides with the distribution of \((X_{k\delta}^\epsilon, \hat{Y}_{s+k\delta}^\epsilon)\).

Denote
\[ \hat{\mathcal{F}}_s := \sigma\{Y_u^{x,y}, u \leq s\}. \]

Then by Lemma 3.6 and Markov property, we have
\[
\Psi_k(s, r) = E\left[ b(X_{k\delta}^\epsilon, Y_s, X_{k\delta}^\epsilon, \hat{Y}_{s+k\delta}^\epsilon) - \tilde{b}(X_{k\delta}^\epsilon, b(X_{k\delta}^\epsilon, Y_{k\delta}^\epsilon, X_{k\delta}^\epsilon, \hat{Y}_{s+k\delta}^\epsilon) - \tilde{b}(X_{k\delta}^\epsilon) \right]
\]
\[
= E\left[ b(x, Y_s^{x,y}) - \tilde{b}(x) \right]_{x = X_{k\delta}^\epsilon, y = \hat{Y}_{s+k\delta}^\epsilon}
\]
\[
= E\left[ E\left[ b(x, Y_s^{x,y}) | \hat{\mathcal{F}}_r \right] - \tilde{b}(x) \right]_{x = X_{k\delta}^\epsilon, y = \hat{Y}_{s+k\delta}^\epsilon}
\]
\[
= E\left[ \left| E\left[ b(x, Y_s^{x,y}) \right] \right|_{x = X_{k\delta}^\epsilon, y = \hat{Y}_{s+k\delta}^\epsilon} \right] \leq C_T E\left[ (|x|^2 + |r|^2)^{1/2} \right] e^{-\delta r^2} \leq C_T e^{-\delta r^2},
\]
where the last inequality comes from Lemmas 3.1 and 3.3. Hence we have
\[
E\left[ \sup_{t \in [0, T \wedge \tau_R^\epsilon]} I(t) \right] \leq C_T, x, y \epsilon^2 \delta^2 \int_0^\delta \int_0^{\tau} e^{-\epsilon \delta r^2} dr d\tau
\]
\[
= C_T, x, y \epsilon^2 \delta^2 \left( \frac{\delta}{\eta} - \frac{1}{\eta^2} + \frac{1}{\eta^2} e^{-\frac{\eta^2}{\delta}} \right)
\]
\[
\leq C_T, x, y \epsilon^2 \delta. \tag{3.20}
\]

According to estimates (3.15) to (3.20), we obtain that
\[
E\left( \sup_{t \in [0, T \wedge \tau_R^\epsilon]} |\hat{X}_t^\epsilon - \hat{X}_t| \right) \leq C_{T, R} \int_0^T E\left( \sup_{s \in [0, t \wedge \tau_R^\epsilon]} |\hat{X}_s^\epsilon - \hat{X}_s| \right) dt + C_{T, x, y} \epsilon \delta + C_{T, x, y} \delta.
\]
By Gronwall’s inequality we get
\[
E\left( \sup_{t \leq T \wedge \tau_R^\epsilon} |\hat{X}_t^\epsilon - \hat{X}_t| \right) \leq C_{T, R, x, y} \left( \frac{\epsilon}{\delta} + \delta \right).
\]
The proof is complete.

Now we can finish the proof of our main result.

**Proof of Theorem 2.2** Taking \(\delta = \epsilon^{1/2}\), Lemmas 3.5 and 3.9 imply that
\[
E\left( \sup_{t \in [0, T \wedge \tau_R^\epsilon]} |X_t^\epsilon - \hat{X}_t| \right) \leq E\left( \sup_{t \in [0, T \wedge \tau_R^\epsilon]} |X_t^\epsilon - \hat{X}_t| + \sup_{t \in [0, T \wedge \tau_R^\epsilon]} |\hat{X}_t^\epsilon - \hat{X}_t| \right)
\]
\[
\leq C_{T, R, x, y} \left( \sqrt{\epsilon \delta^{-1}} + \sqrt{\delta} \right)
\]
\[
\leq C_{T, R, x, y} \epsilon^{1/4}. \tag{3.21}
\]
By Chebyshev’s inequality, Lemmas 3.3 and 3.8 we have

\[
E \left( \sup_{t \in [0, T]} |X_t^\epsilon - \bar{X}_t| I_{(T > \tilde{T}_R)} \right) \leq \left[ E \left( \sup_{t \in [0, T]} |X_t^\epsilon - \bar{X}_t|^2 \right) \right]^{1/2} \cdot [P(T > \tilde{T}_R)]^{1/2} \\
\leq \frac{C}{R} E \left( \sup_{t \in [0, T]} |X_t^\epsilon|^2 + \sup_{t \in [0, T]} |\bar{X}_t|^2 \right) \\
\leq \frac{C_{T,x,y}}{R}.
\]

(3.22)

Hence, by (3.21) and (3.22), we obtain that

\[
E \left( \sup_{t \in [0, T]} |X_t^\epsilon - \bar{X}_t| \right) \leq C_{T,R,x,y} \epsilon^{1/4} + \frac{C_{T,x,y}}{R}.
\]

Now, letting \( \epsilon \to 0 \) firstly and then \( R \to \infty \), we have

\[
\lim_{\epsilon \to 0} E \left( \sup_{t \in [0, T]} |X_t^\epsilon - \bar{X}_t| \right) = 0.
\]

(3.23)

On one hand, if \( \lambda = 0 \) in \( (H_1) \) and \( (A_2) \) holds, by Lemma 3.1 and 3.8 we have

\[
E \left( \sup_{t \in [0, T]} |X_t^\epsilon - \bar{X}_t|^p \right) \leq C_p E \left( \sup_{t \in [0, T]} |X_t^\epsilon|^p + \sup_{t \in [0, T]} |\bar{X}_t|^p \right) < \infty, \quad \forall p > 0.
\]

On the other hand, if \( \lambda > 0 \) in \( (H_1) \) and \( (A_k) \) with \( k \geq \gamma + 1 \) holds, by Lemmas 3.1 and 3.8 we have

\[
E \left( \sup_{t \in [0, T]} |X_t^\epsilon - \bar{X}_t|^k \right) \leq C_k E \left( \sup_{t \in [0, T]} |X_t^\epsilon|^k + \sup_{t \in [0, T]} |\bar{X}_t|^k \right) < \infty.
\]

Hence by Hölder’s inequality and (3.23), it is easy to prove that (2.3) and (2.4) hold, which completes the proof.

References

[1] Bréhier, C.E., Strong and weak orders in averaging for SPDEs, *Stochastic Process. Appl*. 122 (2012) 2553-2593.

[2] Bogoliubov, N.N., Mitropolsky, Y.A., Asymptotic methods in the theory of Non-linear Oscillations, *Gordon and Breach Science Publishers*, New York (1961).

[3] Cerrai S., A Khasminskii type averaging principle for stochastic reaction-diffusion equations, *Ann. Appl. Probab*. 19 (2009) 899-948.

[4] Cerrai S., Freidli M., Averaging principle for stochastic reaction-diffusion equations, *Probab. Theory Related Fields* 144 (2009) 137-177.

[5] Cerrai S., Freidlin M., Fast flow asymptotics for stochastic incompressible viscous fluids in \( \mathbb{R}^2 \) and SPDEs on graphs, to appear on *Probab. Theory Related Fields* (2018).
[6] Cerrai S., Lunardi A., Averaging principle for nonautonomous slow-fast systems of stochastic reaction-diffusion equations: the almost periodic case, *SIAM J. Math. Anal.* 49 (2017) 2843-2884.

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

[8] E W., Liu D., Vanden-Eijnden E., Analysis of multiscale methods for stochastic differential equations, *Comm. Pure Appl. Math* 58 (2005) 1544-1585.

[9] Fu H., Liu J., Strong convergence in stochastic averaging principle for two time-scales stochastic partial differential equations, *J. Math. Anal. Appl.* 384 (2011) 70-86.

[10] Fu H., Wan L., Liu J., Strong convergence in averaging principle for stochastic hyperbolic-parabolic equations with two time-scales, *Stochastic Process. Appl.* 125 (2015) 3255-3279.

[11] Fu H., Wan L., Wang Y., Liu J., Strong convergence rate in averaging principle for stochastic FitzHugh-Nagumo system with two time-scales, *J. Math. Anal. Appl.* 416 (2014) 609-628.

[12] Givon D., Kevrekidis I. G. and Kupferman R., Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems, *Comm. Math. Sci.* 4 (2006) 707-729.

[13] Golec J., Stochastic averaging principle for systems with pathwise uniqueness, *Stochastic Anal. Appl.* 13 (1995) 307-322.

[14] Golec J., Ladde G., Averaging principle and systems of singularly perturbed stochastic differential equations, *J. Math. Phys.* 31 (1990) 1116-1123.

[15] Khasminskii R.Z., On an averaging principle for Itô stochastic differential equations, *Kibernetica* (4) (1968) 260-279.

[16] Liu D., Strong convergence of principle of averaging for multiscale stochastic dynamical systems, *Commun. Math. Sci.* 8 (2010) 999-1020.

[17] Liu D., Strong convergence rate of principle of averaging for jump-diffusion processes, *Front. Math. China* 7 (2012) 305-320.

[18] Liu W. and Röckner M., Stochastic Partial Differential Equations: An Introduction, Universitext, Springer 2015.

[19] Wang W., Roberts A.J., Average and deviation for slow-fast stochastic partial differential equations, *J. Differential Equations* 253 (2012) 1265-1286.

[20] Wang W., Roberts A.J., Duan J., Large deviations and approximations for slow-fast stochastic reaction-diffusion equations, *J. Differential Equations* 253 (2012) 3501-3522.

[21] Veretennikov A.Y., On the averaging principle for systems of stochastic differential equations, *Math. USSR Sborn.* 69 (1991) 271-284.
[22] Xu J., Miao Y., $L^p$ ($p > 2$)-strong convergence of an averaging principle for two-time-scales jump-diffusion stochastic differential equations, *Nonlinear Anal. Hybrid Syst.* 18 (2015) 33-47.

[23] Xu J., Liu J., Miao Y., Strong averaging principle for two-time-scale SDEs with non-Lipschitz coefficients, to appear in *J. Math. Anal. Appl.* 2018.

[24] Xu Y., Pei, B., Wu, J.-L., Stochastic averaging principle for differential equations with non-Lipschitz coefficients driven by fractional Brownian motion, *Stoch. Dyn.* 17 (2017), no. 2, 1750013, 16 pp.