AVERAGING PRINCIPLE FOR SLOW-FAST STOCHASTIC DIFFERENTIAL EQUATIONS WITH TIME DEPENDENT LOCALLY LIPSCHITZ COEFFICIENTS

WEI LIU, MICHAEL RÖCKNER, XIAOBIN SUN, AND YINGCHAO XIE

Abstract. This paper is devoted to studying the averaging principle for stochastic differential equations with slow and fast time-scales, where the drift coefficients satisfy local Lipschitz conditions with respect to the slow and fast variables, and the coefficients in the slow equation depend on time t and ω. Making use of the techniques of time discretization and truncation, we prove that the slow component strongly converges to the solution of the corresponding averaged equation.

1. Introduction

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

$$\left\{\begin{array}{l} dX_t^\epsilon = b(t, X_t^\epsilon, Y_t^\epsilon)dt + \sigma(t, X_t^\epsilon)dW_t^1, \quad X_0^\epsilon = x \in \mathbb{R}^n, \\ dY_t^\epsilon = \frac{1}{\epsilon} f(t, X_t^\epsilon, Y_t^\epsilon)dt + \frac{1}{\sqrt{\epsilon}} g(t, X_t^\epsilon, Y_t^\epsilon)dW_t^2, \quad Y_0^\epsilon = y \in \mathbb{R}^m, \end{array}\right. \tag{1.1}$$

where ϵ is a small positive parameter describing the ratio of time scales between the slow component $X_t^\epsilon \in \mathbb{R}^n$ and fast component $Y_t^\epsilon \in \mathbb{R}^m$. Let $\{W_t^1\}_{t\geqslant 0}$ and $\{W_t^2\}_{t\geqslant 0}$ be mutually independent d_1 and d_2 dimensional standard Brownian motions on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and $\{\mathscr{F}_t, t \geqslant 0\}$ is the natural filtration generated by W_t^1 and W_t^2. Let us consider the following given maps

$$\begin{aligned} b &: [0,\infty) \times \mathbb{R}^n \times \mathbb{R}^m \times \Omega \to \mathbb{R}^n; \\ \sigma &: [0,\infty) \times \mathbb{R}^n \times \Omega \to \mathbb{R}^{n\times d_1}; \\ f &: [0,\infty) \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m; \\ g &: [0,\infty) \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{m\times d_2} \end{aligned}$$

such that b, σ, f and g are continuous in $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ for each fixed $t \in [0,\infty)$, $\omega \in \Omega$, and progressively measurable, i.e., for each t, their restrictions to $[0, t] \times \Omega$ are $\mathcal{B}([0, t]) \otimes \mathscr{F}_t$-measurable for any fixed $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$. In particular, for fixed $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$ and $t \in [0,\infty)$, $b(t, x, y)$ and $\sigma(t, x)$ are $\mathscr{F}_t$-measurable.

Under some reasonable assumptions, we intend to prove X^ϵ converges to $\bar{X}$ in the sense of $L^p(\Omega; C([0, T], \mathbb{R}^n))$, *i.e.*, for some $p > 0$,

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

Date: May 15, 2019.

2000 *Mathematics Subject Classification.* Primary 60H10, 34K33; Secondary 34D20.

Key words and phrases. Averaging principle; Local Lipschitz; Time-dependent; Strong convergence.

where $\bar{X}$ is the solution of the corresponding averaged equation

$$\left\{\begin{array}{l} d\bar{X}_t = \bar{b}(t, \bar{X}_t)dt + \sigma(t, \bar{X}_t)dW^1_t. \\ \bar{X}_0 = x, \end{array}\right. \tag{1.3}$$

Here $\bar{b}(t, x) = \int_{\mathbb{R}^m} b(t, x, y)\mu^{t,x}(dy)$ and $\mu^{t,x}$ denotes the unique invariant measure for the transition semigroup of the following frozen equation

$$\left\{\begin{array}{l} dY_s = f(t, x, Y_s)ds + g(t, x, Y_s)d\tilde{W}^2_s, \\ Y_0 = y, \end{array}\right. \tag{1.4}$$

where $\{\tilde{W}^2_s\}_{s\geqslant 0}$ is a d_2-dimensional standard Brownian motion on another complete probability space. Notice that for fixed $t \geqslant 0$ and $x \in \mathbb{R}^n$, the solution of Eq. (1.4) is a time-homogeneous Markov process, so its transition semigroup has a unique invariant measure $\mu^{t,x}$ depending on t and x under appropriate conditions. Hence, the definition of the averaged coefficient $\bar{b}$ is meaningful.

Another simple understanding about the averaged coefficient is to change the time-dependent coefficients to time-independent coefficients. If b and σ are independent of ω, then we define

$$Z^\epsilon_t = \begin{pmatrix} t \\ X^\epsilon_t \end{pmatrix}, \quad \tilde{b}(z, y) = \begin{pmatrix} 1 \\ b(z, y) \end{pmatrix}$$

and

$$\tilde{\sigma}(z) = \begin{pmatrix} 0 & 0 \\ 0 & \sigma(z, y) \end{pmatrix}, \quad \tilde{W}^1_t = \begin{pmatrix} W_t \\ W^1_t \end{pmatrix}.$$

where $z \in \mathbb{R}^{n+1}$, $\{W_t\}_{t\geqslant 0}$ is another one dimensional standard Brownian motion independent of W^1_t and W^2_t. By an easy transformation, the system (1.1) is then equivalent to the following slow-fast system

$$\left\{\begin{array}{l} dZ^\epsilon_t = \tilde{b}(Z^\epsilon_t, Y^\epsilon_t)dt + \tilde{\sigma}(Z^\epsilon_t)d\tilde{W}^1_t, \quad Z^\epsilon_0 = \begin{pmatrix} 0 \\ x \end{pmatrix}, \\ dY^\epsilon_t = \frac{1}{\epsilon}f(Z^\epsilon_t, Y^\epsilon_t)dt + \frac{1}{\sqrt{\epsilon}}g(Z^\epsilon_t, Y^\epsilon_t)dW^2_t, \quad Y^\epsilon_0 = y, \end{array}\right. \tag{1.5}$$

where $Z^\epsilon_t \in \mathbb{R}^{1+n}$ and $Y^\epsilon_t \in \mathbb{R}^m$ are the corresponding slow and fast components for the new system (1.5) respectively. Notice that the system (1.5) is a time-independent case, and it is easy to see the corresponding frozen equation should be Eq. (1.4).

Although the coefficients depend on time in our paper, it is different from the non-autonomous case in [3]. Recently, Cerrai [3] studied the averaging principle for non-autonomous slow-fast systems of stochastic reaction-diffusion equations, where the coefficients depend on time and satisfy the almost periodic in time condition. Because the corresponding frozen equation is a non-homogeneous Markov process and the invariant measure does not exist any longer, the assumption of almost periodic in time for the coefficients seems natural and it is used to define the averaged coefficient in a new way.

The theory of averaging principle has a long and rich history in multiscale problems, which arise from material sciences, chemistry, fluid dynamics, biology, climate dynamics and other application areas, see, e.g., [1, 5, 6, 10, 13, 16] and references therein. The multiscale model is very common and involved by slow and fast components in mathematical models. For instance, dynamics of chemical reaction networks often take place on notably different times scales, from the order of nanoseconds (10^{-9} s) to the order of several days. Studying the averaging principle is essential to describe the asymptotic behavior of slow component.

The averaging principle for stochastic differential equations (SDEs for short) was first studied by Khasminskii [11], see, e.g. [7, 8, 9, 12, 18] (and the references therein) for further generalizations. However, most of the known results in the literature mainly considered the cases of coefficients satisfying Lipschitz continuous or sublinear growth conditions. It seems that there are few results about the non-Lipschitz case. Veretennikov [15] established the averaging of systems of Itô stochastic equations, where the drift coefficient b is bounded and measurable *w.r.t.* the slow variable and the other coefficients satisfy Lipschitz conditions. Then convergence in probability was obtained. Xu et al. [17] proved the L^2 convergence for two-time-scales with special non-Lipschitz, but linear growth coefficients.

However, in [15, 17] it can not cover the superlinear growth case of drift coefficient b such as $b(x,y) = x + y^3$. Hence, the motivation of this paper is to weaken the conditions on the drift coefficients b and f to local Lipschitz conditions *w.r.t.* both the slow and fast variables, and to the case where the coefficients in the slow equation can depend on time t and ω, which is inspired from the models in [14, Chapter 3].

Comparing with the known results, the main difficulties here are how to deal with the local Lispchitz continuity *w.r.t.* the fast variable and the dependence on ω of the coefficients. In order to overcome these difficulties, we will continue to use the technique of stopping times very frequently. The main result is e.g. applicable to many slow-fast SDE models with polynomial drift coefficients. It is worth to mention that the approach based on time discretization will be used in the proof, so we need the local Lipschitz conditions instead of the one-sided type conditions in [14, Theorem 3.1.1].

The paper is organized as follows. In the next section, we introduce some notations and assumptions that we use throughout the paper and formulate the main result. Section 3 is devoted to proving the strong convergence result. In Section 4, we will give some examples to illustrate the applicability of our result. The final section is the Appendix, where we present the detailed proof of existence and uniqueness of solutions for system (1.1) and the corresponding averaged equation.

Please note that C and C_p denote some positive constants which may change from line to line throughout this paper, where p is one or more than one parameter and C_p is used to emphasize that the constant depends on the corresponding parameter. C_T will usually denote some nondecreasing function *w.r.t.* T.

2. Main results

Now we impose the following assumptions on the coefficients b, σ, f and g. Let $|\cdot|$ be the Euclidean norm, $\langle \cdot, \cdot \rangle$ be the Euclidean inner product and $\|\cdot\|$ be the matrix norm.

($\mathbf{H}_1$) (i) There exists $\theta_1 \geqslant 0$ such that for any $t, R \geqslant 0$, $x_i \in \mathbb{R}^n, y \in \mathbb{R}^m$ with $|x_i| \leqslant R$,

$$2|b(t,x_1,y) - b(t,x_2,y)||x_1 - x_2| + \|\sigma(t,x_1) - \sigma(t,x_2)\|^2 \leqslant K_t(R)(1 + |y|^{\theta_1})|x_1 - x_2|^2,$$

where $K_t(R)$ is an $\mathbb{R}_+$-valued $\mathscr{F}_t$-adapted process satisfying for all $R, T, p \in [0, \infty)$,

$$\alpha_T(R) := \int_0^T K_t(R)dt < \infty, \quad \text{on} \quad \Omega,$$

$$\mathbb{E}e^{p\alpha_T(1)} < \infty, \quad \sup_{t\in[0,T]} \mathbb{E}|K_t(1)|^4 < \infty.$$

Furthermore, there exists $R_0 > 0$, such that for any $R \geqslant R_0$, $T \geqslant 0$,

$$\mathbb{E}\int_0^T [K_t(R)]^4 dt < \infty.$$

(ii) There exist constants $\theta_2, \theta_3 \geqslant 1$ and $\gamma_1 \in (0,1]$ such that for any $x \in \mathbb{R}^n$, $y, y_1, y_2 \in \mathbb{R}^m$ and $T > 0$ with $t, s \in [0,T]$,

$$|b(t,x,y_1) - b(t,x,y_2)| \leqslant C_T|y_1 - y_2| \left[|y_1|^{\theta_2} + |y_2|^{\theta_2} + K_t(1) + |x|^{\theta_3}\right]$$

and

$$|b(t,x,y) - b(s,x,y)| \leqslant C_T|t-s|^{\gamma_1} \left[|y|^{\theta_2} + |x|^{\theta_3} + Z_T\right], \quad \text{on} \quad \Omega,$$

where $C_T > 0$ and Z_T is some random variable satisfying $\mathbb{E}Z_T^2 < \infty$.

(iii) There exist $\lambda_1 \geqslant 0$, $C > 0$, $\theta_4 \geqslant 2$ and $\theta_5, \theta_6 \geqslant 1$ such that for any $t > 0, x \in \mathbb{R}^n, y \in \mathbb{R}^m$,

$$2\langle x, b(t,x,y)\rangle \leqslant K_t(1)(1 + |x|^2) + \lambda_1|y|^{\theta_4}$$

and

$$|b(t,x,y)| \leqslant K_t(1) + C(|x|^{\theta_5} + |y|^{\theta_6}), \quad \|\sigma(t,x)\|^2 \leqslant K_t(1) + C|x|^2.$$

($\mathbf{H}_2$) (i) There exists $\beta > 0$ such that for any $t \geqslant 0$, $x \in \mathbb{R}^n, y_1, y_2 \in \mathbb{R}^m$,

$$2\langle f(t,x,y_1) - f(t,x,y_2), y_1 - y_2\rangle + \|g(t,x,y_1) - g(t,x,y_2)\|^2 \leqslant -\beta|y_1 - y_2|^2. \tag{2.1}$$

(ii) For any $T > 0$, there exist $\gamma_2 \in (0,1]$, $C_T > 0$, $\alpha_i \geqslant 1$, $i = 1,2,3,4$ such that for any $t, s \in [0,T]$ and $x_i \in \mathbb{R}^n, y_i \in \mathbb{R}^m$, $i = 1,2$,

$$|f(t,x_1,y_1) - f(s,x_2,y_1)| \leqslant C_T(|t-s|^{\gamma_2} + |x_1 - x_2|)(1 + |x_1|^{\alpha_1} + |x_2|^{\alpha_1} + |y_1|^{\alpha_2});$$

$$\|g(t,x_1,y_1) - g(s,x_2,y_2)\| \leqslant C_T(|t-s|^{\gamma_2} + |x_1 - x_2| + |y_1 - y_2|);$$

$$|f(t,x_1,y_1)| \leqslant C_T(1 + |x_1|^{\alpha_3} + |y_1|^{\alpha_4});$$

$$\|g(t,x_1,y_1)\| \leqslant C_T(1 + |x_1| + |y_1|).$$

($\mathbf{A}_k$) For some fixed $k \geqslant 2$ and any $T > 0$, there exist $C_{T,k}, \beta_k > 0$ such that for any $t \in [0,T]$, $x \in \mathbb{R}^n, y \in \mathbb{R}^m$,

$$2\langle y, f(t,x,y)\rangle + (k-1)\|g(t,x,y)\|^2 \leqslant -\beta_k|y|^2 - \lambda_2|y|^{\theta_4} + C_{T,k}(|x|^{\frac{4}{\theta_4}} + 1), \tag{2.2}$$

where $\lambda_2 = 0$ if $\lambda_1 = 0$, and $\lambda_2 > 0$ otherwise.

Remark 2.1. (1) Condition (2.1) is called strict monotonicity condition, which ensures that exponential ergodicity holds (see Proposition 3.9 below). Condition (2.2) is called strict coercivity condition, which is used to guarantee the existence of invariant measures for the frozen equation (see Eq. (2.7) below). Hence the uniqueness of invariant measures for the frozen equation follows (see Proposition 3.8 below).

(2) The powers θ_4 and $\frac{4}{\theta_4}$ in (2.2) are used to ensure the existence and uniqueness of solutions to the system (1.1) and the corresponding averaged equation (see Eq. (3.8) below) respectively.

(3) If $k_1 > k_2 \geqslant 2$, then ($\mathbf{A}_{k_1}$) implies ($\mathbf{A}_{k_2}$).

(4) We will give some examples in Section 4 to show the assumptions above hold for many drift coefficients of polynomial type.

The following theorem is the existence and uniqueness result for system (1.1), which can be obtained using the classical result due to Krylov (cf. [14, Chapter 3]). The detailed proof will be given in the Appendix.

Theorem 2.2. *Suppose that* $(\mathbf{H}_1)$, $(\mathbf{H}_2)$ *and* $(\mathbf{A}_2)$*hold. Let* $\epsilon_0 = \frac{\lambda_2}{\lambda_1}$ *if* $\lambda_1 > 0$*, and* $\epsilon_0 = 1$ *otherwise. Then for any* $\epsilon \in (0, \epsilon_0)$*, any given initial values* $x \in \mathbb{R}^n, y \in \mathbb{R}^m$*, there exists a unique solution* $\{(X_t^\epsilon, Y_t^\epsilon), t \geqslant 0\}$ *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 for all* $t \in [0,T]$,

$$\left\{\begin{array}{l} X_t^\epsilon = x + \int_0^t b(s, X_s^\epsilon, Y_s^\epsilon)ds + \int_0^t \sigma(s, X_s^\epsilon)dW_s^1, \\ Y_t^\epsilon = y + \frac{1}{\epsilon}\int_0^t f(s, X_s^\epsilon, Y_s^\epsilon)ds + \frac{1}{\sqrt{\epsilon}}\int_0^t g(s, X_s^\epsilon, Y_s^\epsilon)dW_s^2. \end{array}\right. \tag{2.3}$$

Now we formulate the main result of this work.

Theorem 2.3. *Suppose that* $(\mathbf{H}_1)$ *and* $(\mathbf{H}_2)$ *hold.*

(i) If $\lambda_1 = 0$ *in* $(\mathbf{H}_1)$ *and* $(\mathbf{A}_{\tilde{\theta}_1})$ *holds for* $\tilde{\theta}_1 = \max\{4\theta_1, 2\theta_2+2, 2\theta_6, 4\alpha_2\}$*. Then for any* $p > 0$ *we have*

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

(ii) If $\lambda_1 > 0$ *in* $(\mathbf{H}_1)$ *and* $(\mathbf{A}_k)$ *holds for some* $k > \tilde{\theta}_2$ *with* $\tilde{\theta}_2 = \max\{4\theta_1, 2\theta_2+2, 2\theta_6, 4\alpha_2, \theta_5\theta_4, 2\alpha_1\theta_4\}$*. Then for any* $0 < p < \frac{2k}{\theta_4}$ *we have*

$$\lim_{\epsilon\to 0}\mathbb{E}\left(\sup_{t\in[0,T]}|X_t^\epsilon - \bar{X}_t|^p\right) = 0. \tag{2.5}$$

Here $\bar{X}$ *is the solution of the following averaged equation*

$$\left\{\begin{array}{l} d\bar{X}_t = \bar{b}(t, \bar{X}_t)dt + \sigma(t, \bar{X}_t)dW_t^1, \\ \bar{X}_0 = x, \end{array}\right. \tag{2.6}$$

where $\bar{b}(t,x) = \int_{\mathbb{R}^m} b(t,x,y)\mu^{t,x}(dy)$ *and* $\mu^{t,x}$ *denotes the unique invariant measure for the transition semigroup of the corresponding frozen equation*

$$\left\{\begin{array}{l} dY_s = f(t, x, Y_s)ds + g(t, x, Y_s)d\tilde{W}_s^2, \\ Y_0 = y, \end{array}\right. \tag{2.7}$$

where $\{\tilde{W}_s^2\}_{s\geqslant 0}$ *is a* d_2*-dimensional Brownian motion on another complete probability space* $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$.

3. Proof of the Main Result

This section is devoted to proving Theorem 2.3. The proof consists of the following steps. Firstly, we give some a-priori estimates for the solution $(X_t^\epsilon, Y_t^\epsilon)$ to the system (1.1). Secondly, following the discretization techniques inspired by Khasminskii in [11], we introduce an auxiliary process $(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon)$ for which we derive uniform bounds. Making use of the stopping time techniques inspired by [4], we control the (difference) process $X_t^\epsilon - \hat{X}_t^\epsilon$ before the stopping time. Thirdly, based on the ergodic property of the frozen equation, we obtain appropriate control of $\hat{X}_t^\epsilon - \bar{X}_t$ before the stopping time. Finally, we shall use the a-priori estimates of the solution to control the difference after the stopping time. Note that we always assume that $(\mathbf{H}_1)$ and $(\mathbf{H}_2)$ hold and from now on we fix some initial values $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ in this section.

3.1. **Some a-priori estimates of** $(X_t^\epsilon, Y_t^\epsilon)$**.** In this subsection, we prove some uniform bounds *w.r.t.* $\epsilon \in (0, \epsilon_0)$ for the moments of the solution $(X_t^\epsilon, Y_t^\epsilon)$ to system (1.1).

Lemma 3.1. *(i) If $\lambda_1 = 0$ in $(\mathbf{H}_1)$ and $(\mathbf{A}_{k\theta_4})$ holds for some $k \geqslant \frac{2}{\theta_4}$, then for any $T, p > 0$, there exist positive constants $C_{T,p}, C_{T,k}$ such that*

$$\sup_{\epsilon \in (0,\epsilon_0)} \mathbb{E}\left(\sup_{t\in[0,T]} |X_t^\epsilon|^p\right) \leqslant C_{T,p}(1+|x|^p)$$

and

$$\sup_{\epsilon \in (0,\epsilon_0)} \sup_{t\in[0,T]} \mathbb{E}|Y_t^\epsilon|^{k\theta_4} \leqslant C_{T,k}(1+|x|^{2k}+|y|^{k\theta_4}).$$

(ii) If $\lambda_1 > 0$ in $(\mathbf{H}_1)$ and $(\mathbf{A}_{k\theta_4})$ holds for some $k \geqslant 1$, then for any $T > 0$, $k' < k$, there exists a positive constant $C_{T,k}$ such that

$$\sup_{\epsilon \in (0,\epsilon_0)} \mathbb{E}\left(\sup_{t\in[0,T]} |X_t^\epsilon|^{2k'}\right) \leqslant C_{T,k}(|x|^{2k'}+|y|^{k'\theta_4}+1)$$

and

$$\sup_{\epsilon \in (0,\epsilon_0)} \sup_{t\in[0,T]} \mathbb{E}|Y_t^\epsilon|^{k'\theta_4} \leqslant C_{T,k}(|x|^{2k'}+|y|^{k'\theta_4}+1).$$

Proof. (i) According to Itô's formula and $(\mathbf{H}_1)$ with $\lambda_1 = 0$, we have for any $p \geqslant 2$,

$$\begin{aligned}
&e^{-\frac{p}{2}\alpha_t(1)}|X_t^\epsilon|^p\\
=&|x|^p + \frac{p}{2}\int_0^t e^{-\frac{p}{2}\alpha_s(1)}\left[-K_s(1)\right]|X_s^\epsilon|^p ds + p\int_0^t e^{-\frac{p}{2}\alpha_s(1)}|X_s^\epsilon|^{p-2}\langle X_s^\epsilon, b(s,X_s^\epsilon,Y_s^\epsilon)\rangle ds\\
&+\frac{p}{2}\int_0^t e^{-\frac{p}{2}\alpha_s(1)}|X_s^\epsilon|^{p-2}\|\sigma(s,X_s^\epsilon)\|^2 ds + \frac{p(p-2)}{2}\int_0^t e^{-\frac{p}{2}\alpha_s(1)}|X_s^\epsilon|^{p-4}\left|\langle X_s^\epsilon, \sigma(s,X_s^\epsilon)\rangle\right|^2 ds\\
&+p\int_0^t e^{-\frac{p}{2}\alpha_s(1)}|X_s^\epsilon|^{p-2}\langle X_s^\epsilon, \sigma(s,X_s^\epsilon)dW_s^1\rangle\\
\leqslant&|x|^p + \frac{p}{2}\int_0^t e^{-\frac{p}{2}\alpha_s(1)}\left[-K_s(1)\right]|X_s^\epsilon|^p ds + \frac{p}{2}\int_0^t e^{-\frac{p}{2}\alpha_s(1)}|X_s^\epsilon|^{p-2}K_s(1)(1+|X_s^\epsilon|^2)ds\\
&+\frac{p(p-1)}{2}\int_0^t e^{-\frac{p}{2}\alpha_s(1)}|X_s^\epsilon|^{p-2}[K_s(1)+C|X_s^\epsilon|^2]ds\\
&+p\int_0^t e^{-\frac{p}{2}\alpha_s(1)}|X_s^\epsilon|^{p-2}\langle X_s^\epsilon, \sigma(s,X_s^\epsilon)dW_s^1\rangle\\
\leqslant&|x|^p + \frac{p^2}{2}\int_0^t e^{-\frac{p}{2}\alpha_s(1)}|X_s^\epsilon|^{p-2}K_s(1)ds + C_p\int_0^t e^{-\frac{p}{2}\alpha_s(1)}|X_s^\epsilon|^p ds\\
&+p\int_0^t e^{-\frac{p}{2}\alpha_s(1)}|X_s^\epsilon|^{p-2}\langle X_s^\epsilon, \sigma(s,X_s^\epsilon)dW_s^1\rangle.
\end{aligned}$$

Then by the Burkholder-Davis-Gundy inequality and Young's inequality, for any $T>0$, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-\frac{p}{2}\alpha_t(1)}|X_t^\epsilon|^p\right)\right]\\
\leqslant\ & |x|^p+\frac{p^2}{2}\mathbb{E}\int_0^T e^{-\frac{p}{2}\alpha_t(1)}|X_t^\epsilon|^{p-2}K_t(1)dt+C_p\int_0^T\mathbb{E}\left(e^{-\frac{p}{2}\alpha_t(1)}|X_t^\epsilon|^p\right)dt\\
&+C_p\mathbb{E}\left[\int_0^T e^{-p\alpha_t(1)}|X_t^\epsilon|^{2p-2}(K_t(1)+C|X_t^\epsilon|^2)dt\right]^{1/2}\\
\leqslant\ & |x|^p+\frac{1}{4}\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-\frac{p}{2}\alpha_t(1)}|X_t^\epsilon|^p\right)\right]+C_p\mathbb{E}\left[\int_0^T e^{-\alpha_s(1)}K_s(1)ds\right]^{p/2}\\
&+C_p\int_0^T\mathbb{E}\left(e^{-\frac{p}{2}\alpha_t(1)}|X_t^\epsilon|^p\right)dt+\frac{1}{4}\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-\frac{p}{2}\alpha_t(1)}|X_t^\epsilon|^p\right)\right]\\
&+C_p\mathbb{E}\left[\int_0^T e^{-\alpha_t(1)}(K_t(1)+C|X_t^\epsilon|^2)dt\right]^{\frac{p}{2}},
\end{aligned}$$

which implies that

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-\frac{p}{2}\alpha_t(1)}|X_t^\epsilon|^p\right)\right]\leqslant C_p(|x|^p+1)+C_{T,p}\int_0^T\mathbb{E}\left(e^{-\frac{p}{2}\alpha_t(1)}|X_t^\epsilon|^p\right)dt.$$

Then Gronwall's inequality yields that

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-\frac{p}{2}\alpha_t(1)}|X_t^\epsilon|^p\right)\right]\leqslant C_{T,p}(|x|^p+1).$$

Hence, by Hölder inequality and since $\mathbb{E}e^{p\alpha_T(1)}<\infty$ for any $p>0$, we obtain for any $p>0$

$$\begin{aligned}
\mathbb{E}\left(\sup_{t\in[0,T]}|X_t^\epsilon|^p\right)&\leqslant\left\{\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-p\alpha_t(1)}|X_t^\epsilon|^{2p}\right)\right]\right\}^{\frac{1}{2}}\cdot\left[\mathbb{E}e^{p\alpha_T(1)}\right]^{\frac{1}{2}}\\
&\leqslant C_{T,p}(|x|^p+1).
\end{aligned}$$

By Itô's formula we also have

$$\begin{aligned}
\mathbb{E}|Y_t^\epsilon|^{k\theta_4}=&\frac{k\theta_4}{\epsilon}\int_0^t\mathbb{E}\left[|Y_s^\epsilon|^{k\theta_4-2}\langle f(s,X_s^\epsilon,Y_s^\epsilon),Y_s^\epsilon\rangle\right]ds+\frac{k\theta_4}{2\epsilon}\int_0^t\mathbb{E}\left[|Y_s^\epsilon|^{k\theta_4-2}\|g(s,X_s^\epsilon,Y_s^\epsilon)\|^2\right]ds\\
&+\frac{k\theta_4(k\theta_4-2)}{2\epsilon}\int_0^t\mathbb{E}\left[|Y_s^\epsilon|^{k\theta_4-4}\cdot|\langle Y_s^\epsilon,g(s,X_s^\epsilon,Y_s^\epsilon)\rangle|^2\right]ds.
\end{aligned}$$

If $(\mathbf{A}_{k\theta_4})$ holds for $k\geqslant\frac{2}{\theta_4}$, then there exist $\tilde\beta_k, C_{T,k}>0$ such that

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}|Y_t^\epsilon|^{k\theta_4}&\leqslant\frac{k\theta_4}{2\epsilon}\mathbb{E}\left[|Y_t^\epsilon|^{k\theta_4-2}\left(2\langle f(t,X_t^\epsilon,Y_t^\epsilon),Y_t^\epsilon\rangle+(k\theta_4-1)\|g(t,X_t^\epsilon,Y_t^\epsilon)\|^2\right)\right]\\
&\leqslant-\frac{\tilde\beta_k}{\epsilon}\mathbb{E}|Y_t^\epsilon|^{k\theta_4}+\frac{C_{T,k}}{\epsilon}\left(\mathbb{E}|X_t^\epsilon|^{2k}+1\right).
\end{aligned}$$

Hence, by the comparison theorem we obtain

$$\begin{aligned}
\mathbb{E}|Y_t^\epsilon|^{k\theta_4}&\leqslant|y|^{k\theta_4}e^{-\frac{\tilde\beta_k}{\epsilon}t}+\frac{C_{T,k}}{\epsilon}\int_0^t e^{-\frac{\tilde\beta_k}{\epsilon}(t-s)}\Big(1+\mathbb{E}|X_s^\epsilon|^{2k}\Big)ds\\
&\leqslant C_{T,k}(1+|x|^{2k}+|y|^{k\theta_4}),
\end{aligned}$$

which gives the statement (i).

(ii). Notice that since $(\mathbf{H}_1)$ holds with $\lambda_1>0$, for any $k\geqslant 1$, Itô's formula implies that

$$\begin{aligned}
&e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\\
=&|x|^{2k}+k\int_0^t e^{-k\alpha_s(1)}\left[-K_s(1)\right]|X_s^\epsilon|^{2k}ds+2k\int_0^t e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k-2}\langle X_s^\epsilon, b(s,X_s^\epsilon,Y_s^\epsilon)\rangle ds\\
&+k\int_0^t e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k-2}\|\sigma(s,X_s^\epsilon)\|^2ds+2k(k-1)\int_0^t e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k-4}\left|\langle X_s^\epsilon,\sigma(s,X_s^\epsilon)\rangle\right|^2ds\\
&+2k\int_0^t e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k-2}\langle X_s^\epsilon,\sigma(s,X_s^\epsilon)dW_s^1\rangle\\
\leqslant&|x|^{2k}+k\int_0^t e^{-k\alpha_s(1)}\left[-K_s(1)\right]|X_s^\epsilon|^{2k}ds+k\int_0^t e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k-2}K_s(1)(1+|X_s^\epsilon|^2)ds\\
&+k\lambda_1\int_0^t e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k-2}|Y_s^\epsilon|^{\theta_4}ds+k(2k-1)\int_0^t e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k-2}[K_s(1)+C|X_s^\epsilon|^2]ds\\
&+2k\int_0^t e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k-2}\langle X_s^\epsilon,\sigma(s,X_s^\epsilon)dW_s^1\rangle\\
\leqslant&|x|^{2k}+2k^2\int_0^t e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k-2}K_s(1)ds+C_k\int_0^t e^{-k\alpha_s(1)}|Y_s^\epsilon|^{k\theta_4}ds\\
&+C_k\int_0^t e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k}ds+2k\int_0^t e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k-2}\langle X_s^\epsilon,\sigma(s,X_s^\epsilon)dW_s^1\rangle.
\end{aligned}$$

Then by the Burkholder-Davis-Gundy inequality and Young's inequality, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\right)\right]\\
\leqslant&|x|^{2k}+2k^2\int_0^T\mathbb{E}\left[e^{-\frac{k}{2}\alpha_t(1)}|X_t^\epsilon|^{2k-2}K_t(1)\right]dt+C_k\int_0^T\mathbb{E}\left(e^{-k\alpha_t(1)}|Y_t^\epsilon|^{k\theta_4}\right)dt\\
&+C_k\int_0^T\mathbb{E}\left(e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\right)dt+C_k\mathbb{E}\left[\int_0^T e^{-2k\alpha_t(1)}|X_t^\epsilon|^{4k-2}(K_t(1)+C|X_t^\epsilon|^2)dt\right]^{1/2}\\
\leqslant&|x|^{2k}+\frac{1}{4}\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\right)\right]+C_k\mathbb{E}\left[\int_0^T e^{-\alpha_s(1)}K_s(1)ds\right]^k\\
&+C_k\int_0^T\mathbb{E}\left(e^{-k\alpha_t(1)}|Y_t^\epsilon|^{k\theta_4}\right)dt+C_k\int_0^T\mathbb{E}\left(e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\right)dt\\
&+\frac{1}{4}\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\right)\right]+C_k\mathbb{E}\left[\int_0^T e^{-\alpha_t(1)}(K_t(1)+C|X_t^\epsilon|^2)dt\right]^k,
\end{aligned}$$

which implies that

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\right)\right]\leqslant&\,C_{T,k}(|x|^{2k}+1)+C_k\int_0^T\mathbb{E}\left(e^{-k\alpha_t(1)}|Y_t^\epsilon|^{k\theta_4}\right)dt\\
&+C_{T,k}\int_0^T\mathbb{E}\left(e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\right)dt. \qquad (3.1)
\end{aligned}$$

Using Itô's formula again we have

$$\begin{aligned}
&\mathbb{E}\left(e^{-k\alpha_t(1)}|Y_t^\epsilon|^{k\theta_4}\right)\\
=&\int_0^t \mathbb{E}\left[e^{-k\alpha_s(1)}[-kK_s(1)]|Y_s^\epsilon|^{k\theta_4}\right]ds+\frac{k\theta_4}{\epsilon}\int_0^t \mathbb{E}\left[e^{-k\alpha_s(1)}|Y_s^\epsilon|^{k\theta_4-2}\langle f(s,X_s^\epsilon,Y_s^\epsilon),Y_s^\epsilon\rangle\right]ds\\
&+\frac{k\theta_4}{2\epsilon}\int_0^t \mathbb{E}\left[e^{-k\alpha_s(1)}|Y_s^\epsilon|^{k\theta_4-2}\|g(s,X_s^\epsilon,Y_s^\epsilon)\|^2\right]ds\\
&+\frac{k\theta_4(k\theta_4-2)}{2\epsilon}\int_0^t \mathbb{E}\left[e^{-k\alpha_s(1)}|Y_s^\epsilon|^{k\theta_4-4}\cdot|\langle Y_t^\epsilon,g(s,X_s^\epsilon,Y_s^\epsilon)\rangle|^2\right]ds.
\end{aligned}$$

For any $t\in[0,T]$, $(\mathbf{A}_{k\theta_4})$ implies that there exists $\tilde{\beta}_k, C_{T,k}>0$ such that

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}\left(e^{-k\alpha_t(1)}|Y_t^\epsilon|^{k\theta_4}\right)&\leqslant\frac{k\theta_4}{2\epsilon}\mathbb{E}\left[e^{-k\alpha_s(1)}|Y_t^\epsilon|^{k\theta_4-2}\left(2\langle f(t,X_t^\epsilon,Y_t^\epsilon),Y_t^\epsilon\rangle+(k\theta_4-1)\|g(t,X_t^\epsilon,Y_t^\epsilon)\|^2\right)\right]\\
&\leqslant-\frac{\tilde{\beta}_k}{\epsilon}\mathbb{E}\left(e^{-k\alpha_t(1)}|Y_t^\epsilon|^{k\theta_4}\right)+\frac{C_{T,k}}{\epsilon}\left[\mathbb{E}\left(e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\right)+1\right].
\end{aligned}$$

By the comparison theorem there exists $\tilde{\beta}_k>0$

$$\begin{aligned}
\mathbb{E}\left(e^{-k\alpha_t(1)}|Y_t^\epsilon|^{k\theta_4}\right)&\leqslant|y|^{k\theta_4}e^{-\frac{\tilde{\beta}_k t}{\epsilon}}+\frac{C_{T,k}}{\epsilon}\int_0^t e^{-\frac{\tilde{\beta}_k(t-s)}{\epsilon}}\left(\mathbb{E}\left[e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k}\right]+1\right)ds\\
&\leqslant|y|^{k\theta_4}+C_{T,k}\mathbb{E}\left[\sup_{s\in[0,t]}\left(e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k}\right)\right]+C_{T,k}. \qquad (3.2)
\end{aligned}$$

Combining this with (3.1) we obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\right)\right]\leqslant C_{T,k}(|x|^{2k}+|y|^{k\theta_4}+1)+C_{T,k}\int_0^T\mathbb{E}\left[\sup_{s\in[0,t]}\left(e^{-k\alpha_s(1)}|X_s^\epsilon|^{2k}\right)\right]dt.$$

Hence Gronwall's inequality implies that

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\right)\right]\leqslant C_{T,k}(|x|^{2k}+|y|^{k\theta_4}+1). \qquad (3.3)$$

By (3.2) and (3.3), we also have

$$\sup_{t\in[0,T]}\mathbb{E}\left(e^{-k\alpha_t(1)}|Y_t^\epsilon|^{k\theta_4}\right)\leqslant C_{T,k}(|x|^{2k}+|y|^{k\theta_4}+1).$$

Hence, by Hölder's inequality and since $\mathbb{E}e^{p\alpha_T(1)}<\infty$ for any $p>0$, we obtain for any $k'<k$,

$$\begin{aligned}
\mathbb{E}\left(\sup_{t\in[0,T]}|X_t^\epsilon|^{2k'}\right)&\leqslant\left\{\mathbb{E}\left[\sup_{t\in[0,T]}\left(e^{-k\alpha_t(1)}|X_t^\epsilon|^{2k}\right)\right]\right\}^{\frac{k'}{k}}\cdot\left[\mathbb{E}e^{\frac{kk'}{k-k'}\alpha_T(1)}\right]^{\frac{k-k'}{k}}\\
&\leqslant C_{T,k}(|x|^{2k'}+|y|^{k'\theta_4}+1).
\end{aligned}$$

Similarly, we have

$$\sup_{t\in[0,T]}\mathbb{E}|Y_t^\epsilon|^{k'\theta_4}\leqslant C_{T,k}(|x|^{2k'}+|y|^{k'\theta_4}+1).$$

Hence the proof is complete. □

Lemma 3.2. *Assume that either* $(\mathbf{H}_1)$ *with* $\lambda_1 = 0$ *and* $(\mathbf{A}_{2\theta_6})$ *hold or* $(\mathbf{H}_1)$ *with* $\lambda_1 > 0$ *and* $(\mathbf{A}_k)$ *with some* $k > \max\{2\theta_6, \theta_5\theta_4\}$ *hold. Then for any* $T > 0$, $0 \leqslant t \leqslant t+h \leqslant T$, *there exists* $C_{T,x,y} > 0$ *such that*

$$\sup_{\epsilon\in(0,1)} \mathbb{E}|X^\epsilon_{t+h} - X^\epsilon_t|^2 \leqslant C_{T,x,y}h.$$

Proof. We have

$$X^\epsilon_{t+h} - X^\epsilon_t = \int_t^{t+h} b(s, X^\epsilon_s, Y^\epsilon_s)ds + \int_t^{t+h} \sigma(s, X^\epsilon_s)dW^1_s.$$

Then by condition $(\mathbf{H}_1)$ and Lemma 3.1, we get

$$\begin{aligned}
\mathbb{E}|X^\epsilon_{t+h} - X^\epsilon_t|^2 &\leqslant C\mathbb{E}\left|\int_t^{t+h} b(s, X^\epsilon_s, Y^\epsilon_s)ds\right|^2 + C\mathbb{E}\left|\int_t^{t+h} \sigma(s, X^\epsilon_s)dW^1_s\right|^2 \\
&\leqslant C\mathbb{E}\left|\int_t^{t+h} |b(s, X^\epsilon_s, Y^\epsilon_s)|ds\right|^2 + C\int_t^{t+h} \mathbb{E}\|\sigma(s, X^\epsilon_s)\|^2 ds \\
&\leqslant Ch\mathbb{E}\int_t^{t+h} \left[(K_s(1))^2 + |X^\epsilon_s|^{2\theta_5} + |Y^\epsilon_s|^{2\theta_6}\right] ds + C\int_t^{t+h} \mathbb{E}(K_s(1) + C|X^\epsilon_s|^2)ds \\
&\leqslant C_{T,x,y}h.
\end{aligned}$$

The proof is complete. □

3.2. **Estimates of the auxiliary process** $(\hat{X}^\epsilon_t, \hat{Y}^\epsilon_t)$**.** Following the idea inspired by Khasminskii in [11], we introduce an auxiliary process $(\hat{X}^\epsilon_t, \hat{Y}^\epsilon_t) \in \mathbb{R}^n \times \mathbb{R}^m$ and divide $[0, T]$ into intervals depending of size δ, where δ is a fixed positive number depending on ϵ, which will be chosen later. We construct a process $\hat{Y}^\epsilon_t$ with initial value $\hat{Y}^\epsilon_0 = Y^\epsilon_0 = y$, and for $t \in [k\delta, \min((k+1)\delta, T)]$,

$$\hat{Y}^\epsilon_t = \hat{Y}^\epsilon_{k\delta} + \frac{1}{\epsilon}\int_{k\delta}^t f(k\delta, X^\epsilon_{k\delta}, \hat{Y}^\epsilon_s)ds + \frac{1}{\sqrt{\epsilon}}\int_{k\delta}^t g(k\delta, X^\epsilon_{k\delta}, \hat{Y}^\epsilon_s)dW^2_s,$$

i.e.,

$$\hat{Y}^\epsilon_t = y + \frac{1}{\epsilon}\int_0^t f(s(\delta), X^\epsilon_{s(\delta)}, \hat{Y}^\epsilon_s)ds + \frac{1}{\sqrt{\epsilon}}\int_0^t g(s(\delta), X^\epsilon_{s(\delta)}, \hat{Y}^\epsilon_s)dW^2_s,$$

where $s(\delta) = [s/\delta]\delta$ and $[s/\delta]$ is the integer part of s/δ. Also, we define the process $\hat{X}^\epsilon_t$ by

$$\hat{X}^\epsilon_t = x + \int_0^t b(s(\delta), X^\epsilon_{s(\delta)}, \hat{Y}^\epsilon_s)ds + \int_0^t \sigma(s, X^\epsilon_s)dW^1_s.$$

By the construction of $\hat{Y}^\epsilon_t$ and similar arguments as in Lemma 3.1, it is easy to obtain the following estimate whose proof we omit here.

Lemma 3.3. *(i) If* $\lambda_1 = 0$ *in* $(\mathbf{H}_1)$ *and* $(\mathbf{A}_{k\theta_4})$ *holds with some* $k \geqslant \frac{2}{\theta_4}$, *then for any* $T > 0$, *there exists a constant* $C_{T,k} > 0$ *such that*

$$\sup_{\epsilon\in(0,\epsilon_0)}\sup_{t\in[0,T]} \mathbb{E}|\hat{Y}^\epsilon_t|^{k\theta_4} \leqslant C_{T,k}(|x|^{2k} + |y|^{k\theta_4} + 1).$$

(ii) If $\lambda_1 > 0$ *in* $(\mathbf{H}_1)$ *and* $(\mathbf{A}_{k\theta_4})$ *holds with some* $k \geqslant 1$, *then for any* $T > 0$, $k' < k$, *there exists a constant* $C_{T,k} > 0$ *such that*

$$\sup_{\epsilon\in(0,\epsilon_0)}\sup_{t\in[0,T]} \mathbb{E}|\hat{Y}^\epsilon_t|^{k'\theta_4} \leqslant C_{T,k}(|x|^{2k'} + |y|^{k'\theta_4} + 1).$$

Now, we intend to estimate the difference process $Y_t^\epsilon - \hat{Y}_t^\epsilon$ and furthermore the difference process $X_t^\epsilon - \hat{X}_t^\epsilon$.

Lemma 3.4. *Assume that either* $(\mathbf{H}_1)$ *with* $\lambda_1 = 0$ *and* $(\mathbf{A}_{\tilde{\theta}_1})$ *hold or* $(\mathbf{H}_1)$ *with* $\lambda_1 > 0$ *and* $(\mathbf{A}_k)$ *with some* $k > \tilde{\theta}_2$ *hold, where* $\tilde{\theta}_1 = \max\{2\theta_6, 4\alpha_2\}$ *and* $\tilde{\theta}_2 = \max\{2\theta_6, \theta_5\theta_4, 4\alpha_2, 2\alpha_1\theta_4\}$. *Then for any* $T > 0$, *there exists a constant* $C_{T,x,y} > 0$ *such that*

$$\sup_{\epsilon\in(0,1)} \sup_{t\in[0,T]} \mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \leqslant C_{T,x,y}\delta^{\frac{1}{2}\wedge\gamma_2}.$$

Proof. Note that

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

For any $t \in [0, T]$, by Itô's formula we have

$$\begin{aligned} \mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 = & \frac{2}{\epsilon}\int_0^t \mathbb{E}\Big[\langle f(s, X_s^\epsilon, Y_s^\epsilon) - f(s(\delta), X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon), Y_s^\epsilon - \hat{Y}_s^\epsilon\rangle\Big] ds \\ & + \frac{1}{\epsilon}\int_0^t \mathbb{E}\|g(s, X_s^\epsilon, Y_s^\epsilon) - g(s(\delta), X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon)\|^2 ds \\ = & \frac{1}{\epsilon}\int_0^t \mathbb{E}\Big[2\langle f(s, X_s^\epsilon, Y_s^\epsilon) - f(s, X_s^\epsilon, \hat{Y}_s^\epsilon), Y_s^\epsilon - \hat{Y}_s^\epsilon\rangle + \|g(s, X_s^\epsilon, Y_s^\epsilon) - g(s, X_s^\epsilon, \hat{Y}_s^\epsilon)\|^2\Big] ds \\ & + \frac{2}{\epsilon}\int_0^t \mathbb{E}\langle f(s, X_s^\epsilon, \hat{Y}_s^\epsilon) - f(s(\delta), X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon), Y_s^\epsilon - \hat{Y}_s^\epsilon\rangle ds \\ & + \frac{2}{\epsilon}\int_0^t \mathbb{E}\langle g(s, X_s^\epsilon, Y_s^\epsilon) - g(s, X_s^\epsilon, \hat{Y}_s^\epsilon), g(s, X_s^\epsilon, \hat{Y}_s^\epsilon) - g(s(\delta), X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon)\rangle ds \\ & + \frac{1}{\epsilon}\int_0^t \mathbb{E}\|g(s, X_s^\epsilon, \hat{Y}_s^\epsilon) - g(s(\delta), X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon)\|^2 ds. \end{aligned}$$

By condition $(\mathbf{H}_2)$, we obtain

$$\begin{aligned} \frac{d}{dt}\mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \leqslant & \frac{-\beta}{\epsilon}\mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \\ & + \frac{C_T}{\epsilon}\mathbb{E}\left[(|X_t^\epsilon - X_{t(\delta)}^\epsilon| + \delta^{\gamma_2})\left(|\hat{Y}_t^\epsilon|^{\alpha_2} + |X_t^\epsilon|^{\alpha_1} + |X_{t(\delta)}^\epsilon|^{\alpha_1} + 1\right) \cdot |Y_t^\epsilon - \hat{Y}_t^\epsilon|\right] \\ & + \frac{C_T}{\epsilon}\mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|(|X_t^\epsilon - X_{t(\delta)}^\epsilon| + \delta^{\gamma_2}) + \frac{C_T}{\epsilon}\mathbb{E}|X_t^\epsilon - X_{t(\delta)}^\epsilon|^2 + \frac{C_T\delta^{2\gamma_2}}{\epsilon} \\ \leqslant & -\frac{\beta}{2\epsilon}\mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 + \frac{C_T}{\epsilon}\mathbb{E}|X_t^\epsilon - X_{t(\delta)}^\epsilon|^2 + \frac{C_T\delta^{2\gamma_2}}{\epsilon} \\ & + \frac{C_T}{\epsilon}\mathbb{E}\left[(|X_t^\epsilon - X_{t(\delta)}^\epsilon| + \delta^{\gamma_2})\left(|\hat{Y}_t^\epsilon|^{\alpha_2} + |X_t^\epsilon|^{\alpha_1} + |X_{t(\delta)}^\epsilon|^{\alpha_1} + 1\right) \cdot |Y_t^\epsilon - \hat{Y}_t^\epsilon|\right]. \end{aligned}$$

The comparison theorem implies that

$$\begin{aligned}
&\mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \\
\leqslant\ & \frac{C_T}{\epsilon}\int_0^t e^{-\frac{\beta(t-s)}{2\epsilon}}\mathbb{E}\left[(|X_s^\epsilon - X_{s(\delta)}^\epsilon| + \delta^{\gamma_2})\left(|\hat{Y}_s^\epsilon|^{\alpha_2} + |X_s^\epsilon|^{\alpha_1} + |X_{s(\delta)}^\epsilon|^{\alpha_1} + 1\right)\cdot|Y_s^\epsilon - \hat{Y}_s^\epsilon|\right]ds \\
&+\frac{C}{\epsilon}\int_0^t e^{-\frac{\beta(t-s)}{2\epsilon}}\mathbb{E}|X_s^\epsilon - X_{s(\delta)}^\epsilon|^2 ds + \frac{C_T}{\epsilon}\int_0^t e^{-\frac{\beta(t-s)}{2\epsilon}}\delta^{2\gamma_2}ds \\
\leqslant\ & \frac{C_T}{\epsilon}\int_0^t e^{-\frac{\beta(t-s)}{2\epsilon}}\left(\mathbb{E}|X_s^\epsilon - X_{s(\delta)}^\epsilon|^2 + \delta^{2\gamma_2}\right)^{1/2} \\
&\qquad\left[\mathbb{E}\left(|\hat{Y}_s^\epsilon|^{4\alpha_2} + |X_s^\epsilon|^{4\alpha_1} + |X_{s(\delta)}^\epsilon|^{4\alpha_1} + 1\right)\mathbb{E}\left(|Y_s^\epsilon - \hat{Y}_s^\epsilon|^4\right)\right]^{1/4}ds \\
&+\frac{C}{\epsilon}\int_0^t e^{-\frac{\beta(t-s)}{2\epsilon}}\mathbb{E}|X_s^\epsilon - X_{s(\delta)}^\epsilon|^2 ds + C_T\delta^{2\gamma_2}.
\end{aligned}$$

Hence, by Lemma 3.1 and 3.2, we have

$$\mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \leqslant C_{T,x,y}\delta^{\frac{1}{2}\wedge\gamma_2}.$$

The proof is complete. □

In order to estimate the difference process $X_t^\epsilon - \hat{X}_t^\epsilon$. We first construct the following stopping time, for fixed $\epsilon \in (0, \epsilon_0), R \geqslant R_0, M \geqslant 0$,

$$\begin{aligned}
\tau_{R,M}^\epsilon := \inf&\left\{t \geqslant 0 : |X_t^\epsilon| + \int_0^t |Y_s^\epsilon|^{2\theta_2}ds + \int_0^t |\hat{Y}_s^\epsilon|^{4\theta_1\vee 2\theta_2}ds + \int_0^t [K_s(1)]^2\, ds \geqslant R\right\} \\
&\wedge \inf\left\{t \geqslant 0 : \int_0^t |K_s(R)|^4 ds \geqslant M\right\},
\end{aligned}$$

and $\inf\{\emptyset\} := \infty$.

Lemma 3.5. *Assume that either* $(\mathbf{H}_1)$ *with* $\lambda_1 = 0$ *and* $(\mathbf{A}_{\tilde{\theta}_1})$ *hold or* $(\mathbf{H}_1)$ *with* $\lambda_1 > 0$ *and* $(\mathbf{A}_k)$ *with some* $k > \tilde{\theta}_2$ *hold, where* $\tilde{\theta}_1 = \max\{2\theta_6, 4\alpha_2\}$ *and* $\tilde{\theta}_2 = \max\{2\theta_6, \theta_5\theta_4, 4\alpha_2, 2\alpha_1\theta_4\}$. *Then for any* $T, M > 0$ *and* $R \geqslant R_0$*, there exists a constant* $C_{T,R,M} > 0$ *such that*

$$\mathbb{E}\Big(\sup_{t\in[0,T\wedge\tau_{R,M}^\epsilon]}|X_t^\epsilon - \hat{X}_t^\epsilon|^2\Big) \leqslant C_{T,R,M}\delta^\gamma,$$

where $\gamma = \min\{2\gamma_1, \gamma_2, 1/2\}$.

Proof. Recall that

$$X_t^\epsilon = x + \int_0^t b(s, X_s^\epsilon, Y_s^\epsilon)ds + \int_0^t \sigma(s, X_s^\epsilon)dW_s^1$$

and

$$\hat{X}_t^\epsilon = x + \int_0^t b(s(\delta), X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon)ds + \int_0^t \sigma(s, X_s^\epsilon)dW_s^1.$$

Then we have

$$X_t^\epsilon - \hat{X}_t^\epsilon = \int_0^t \left[b(s, X_s^\epsilon, Y_s^\epsilon) - b(s(\delta), X_{s(\delta)}^\epsilon, \hat{Y}_s^\epsilon)\right]ds.$$

By Lemma 3.2 and 3.4 we have

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{t\in[0,T\wedge\tau^{\epsilon}_{R,M}]}|X^{\epsilon}_t-\hat{X}^{\epsilon}_t|^2\right)\\
\leqslant\ &\mathbb{E}\left[\int_0^{T\wedge\tau^{\epsilon}_{R,M}}\left|b(s,X^{\epsilon}_s,Y^{\epsilon}_s)-b(s(\delta),X^{\epsilon}_{s(\delta)},\hat{Y}^{\epsilon}_s)\right|ds\right]^2\\
\leqslant\ &C\mathbb{E}\left[\int_0^{T\wedge\tau^{\epsilon}_{R,M}}\left|b(s,X^{\epsilon}_s,Y^{\epsilon}_s)-b(s,X^{\epsilon}_s,\hat{Y}^{\epsilon}_s)\right|ds\right]^2+C\mathbb{E}\left[\int_0^{T\wedge\tau^{\epsilon}_{R,M}}\left|b(s,X^{\epsilon}_s,\hat{Y}^{\epsilon}_s)-b(s,X^{\epsilon}_{s(\delta)},\hat{Y}^{\epsilon}_s)\right|ds\right]^2\\
&+C\mathbb{E}\left[\int_0^{T\wedge\tau^{\epsilon}_{R,M}}\left|b(s,X^{\epsilon}_{s(\delta)},\hat{Y}^{\epsilon}_s)-b(s(\delta),X^{\epsilon}_{s(\delta)},\hat{Y}^{\epsilon}_s)\right|ds\right]^2\\
\leqslant\ &C\mathbb{E}\left[\int_0^{T\wedge\tau^{\epsilon}_{R,M}}|Y^{\epsilon}_s-\hat{Y}^{\epsilon}_s|^2ds\cdot\int_0^{T\wedge\tau^{\epsilon}_{R,M}}\left(|Y^{\epsilon}_s|^{2\theta_2}+|\hat{Y}^{\epsilon}_s|^{2\theta_2}+|X^{\epsilon}_s|^{2\theta_3}+[K_s(1)]^2\right)ds\right]\\
&+C\mathbb{E}\left[\int_0^{T\wedge\tau^{\epsilon}_{R,M}}|X^{\epsilon}_s-X^{\epsilon}_{s(\delta)}|^2ds\left(\int_0^{T\wedge\tau^{\epsilon}_{R,M}}[K_s(R)]^4ds\right)^{1/2}\left(\int_0^{T\wedge\tau^{\epsilon}_{R,M}}\left(1+|\hat{Y}^{\epsilon}_s|^{4\theta_1}\right)ds\right)^{1/2}\right]\\
&+\delta^{2\gamma_1}C_T\mathbb{E}\int_0^{T\wedge\tau^{\epsilon}_{R,M}}\left(|X^{\epsilon}_{s(\delta)}|^{2\theta_3}+|\hat{Y}^{\epsilon}_s|^{2\theta_2}\right)ds+\delta^{2\gamma_1}C_T\mathbb{E}Z^2_T\\
\leqslant\ &C_{T,R}\mathbb{E}\int_0^{T\wedge\tau^{\epsilon}_{R,M}}|Y^{\epsilon}_s-\hat{Y}^{\epsilon}_s|^2ds+C_{R,M}\mathbb{E}\int_0^{T\wedge\tau^{\epsilon}_{R,M}}|X^{\epsilon}_s-X^{\epsilon}_{s(\delta)}|^2ds+C_{T,R,M}\delta^{2\gamma_1}\\
\leqslant\ &C_{T,R,M}\delta^{\min\{2\gamma_1,\gamma_2,1/2\}}.
\end{aligned}
$$

The proof is complete. □

3.3. **The frozen equation.** We first introduce the frozen equation associated to the fast motion for fixed $t>0$ and fixed slow component $x\in\mathbb{R}^n$.

$$
\left\{\begin{array}{l} dY_s=f(t,x,Y_s)ds+g(t,x,Y_s)d\tilde{W}^2_s,\\ Y_0=y,\end{array}\right. \tag{3.4}
$$

where $\{\tilde{W}^2_s\}_{s\geqslant 0}$ is a d_2-dimensional Brownian motion on another complete probability space $(\tilde{\Omega},\tilde{\mathscr{F}},\tilde{\mathbb{P}})$ and $\{\tilde{\mathscr{F}}_s,s\geqslant 0\}$ is the natural filtration generated by $\{\tilde{W}^2_s\}_{s\geqslant 0}$. If $(\mathbf{H}_2)$ and $(\mathbf{A}_2)$ hold, then it is easy to prove for any fixed $t>0$, $x\in\mathbb{R}^n$ and any initial data $y\in\mathbb{R}^m$, Eq. (3.4) has a unique strong solution $\{Y^{t,x,y}_s\}_{s\geqslant 0}$, which is a time homogeneous Markov process. Let $\{P^{t,x}_s\}_{s\geqslant 0}$ be the transition semigroup of $\{Y^{t,x,y}_s\}_{s\geqslant 0}$, *i.e.* for any bounded measurable function $\varphi:\mathbb{R}^m\to\mathbb{R}$,

$$
P^{t,x}_s\varphi(y):=\tilde{\mathbb{E}}\varphi(Y^{t,x,y}_s),\quad y\in\mathbb{R}^m,s\geqslant 0,
$$

where $\tilde{\mathbb{E}}$ is the expectation on $(\tilde{\Omega},\tilde{\mathscr{F}},\tilde{\mathbb{P}})$.

Lemma 3.6. *Suppose that $(\mathbf{A}_k)$ holds for some $k\geqslant 2$. Then there exists $\tilde{\beta}_k>0$ such that for any $x\in\mathbb{R}^n,y\in\mathbb{R}^m$, $s\geqslant 0$ and $T>0$ with $t\in[0,T]$,*

$$
\tilde{\mathbb{E}}|Y^{t,x,y}_s|^k\leqslant e^{-\tilde{\beta}_k s}|y|^k+C_{T,k}(1+|x|^{\frac{2k}{\theta_4}}). \tag{3.5}
$$

Proof. Note that

$$
Y^{t,x,y}_s=y+\int_0^s f(t,x,Y^{t,x,y}_r)dr+\int_0^s g(t,x,Y^{t,x,y}_r)d\tilde{W}^2_r.
$$

By the Itô's formula we have

$$\tilde{\mathbb{E}}|Y_s^{t,x,y}|^k = k\int_0^s \tilde{\mathbb{E}}\left[|Y_r^{t,x,y}|^{k-2}\langle f(t,x,Y_r^{t,x,y}), Y_r^{t,x,y}\rangle\right]dr + \frac{k}{2}\int_0^s \tilde{\mathbb{E}}\left[|Y_r^{t,x,y}|^{k-2}\|g(t,x,Y_r^{t,x,y})\|^2\right]dr$$
$$+\frac{k(k-2)}{2}\int_0^s \tilde{\mathbb{E}}\left[|Y_r^{t,x,y}|^{k-4}\cdot|\langle Y_r^{t,x,y}, g(t,x,Y_r^{t,x,y})\rangle|^2\right]dr.$$

Then assumption $(\mathbf{A}_k)$ yields that there exists $\tilde{\beta}_k > 0$ such that for any $t\in[0,T]$

$$\frac{d}{ds}\tilde{\mathbb{E}}|Y_s^{t,x,y}|^k \leqslant \frac{k}{2}\tilde{\mathbb{E}}\left[|Y_s^{t,x,y}|^{k-2}\left(2\langle f(t,x,Y_s^{t,x,y}), Y_s^{t,x,y}\rangle + (k-1)\|g(t,x,Y_s^{t,x,y})\|^2\right)\right]$$
$$\leqslant -\tilde{\beta}_k\tilde{\mathbb{E}}|Y_s^{t,x,y}|^k + C_{T,k}\left(|x|^{\frac{2k}{\theta_4}}+1\right).$$

Hence, by the comparison theorem we have

$$\tilde{\mathbb{E}}|Y_s^{t,x,y}|^k \leqslant |y|^k e^{-\tilde{\beta}_k s} + C_{T,k}(1+|x|^{\frac{2k}{\theta_4}})\int_0^s e^{-\tilde{\beta}_k(s-r)}dr$$
$$\leqslant |y|^k e^{-\tilde{\beta}_k s} + C_{T,k}(1+|x|^{\frac{2k}{\theta_4}}).$$

The proof is complete. □

Lemma 3.7. *There exists $\beta>0$ such that for any $t,s\geqslant 0$, $x\in\mathbb{R}^n, y_1,y_2\in\mathbb{R}^m$,*

$$\tilde{\mathbb{E}}|Y_s^{t,x,y_1} - Y_s^{t,x,y_2}|^2 \leqslant e^{-\beta s}|y_1-y_2|^2.$$

Proof. Note that

$$Y_s^{t,x,y_1} - Y_s^{t,x,y_2} = y_1 - y_2 + \int_0^s f(t,x,Y_r^{t,x,y_1}) - f(t,x,Y_r^{t,x,y_2})dr$$
$$+\int_0^s \left[g(t,x,Y_r^{t,x,y_1}) - g(t,x,Y_r^{t,x,y_2})\right]d\tilde{W}_r^2.$$

By Itô's formula we obtain

$$\tilde{\mathbb{E}}|Y_s^{t,x,y_1} - Y_s^{t,x,y_2}|^2 = \int_0^s \tilde{\mathbb{E}}\left[2\langle f(t,x,Y_r^{t,x,y_1}) - f(t,x,Y_r^{t,x,y_2}), Y_r^{t,x,y_1} - Y_r^{t,x,y_2}\rangle\right]dr$$
$$+\int_0^s \tilde{\mathbb{E}}\|g(t,x,Y_r^{t,x,y_1}) - g(t,x,Y_r^{t,x,y_2})\|^2 dr.$$

Then condition (2.1) in $(\mathbf{H}_2)$ yields that there exist $\beta>0$ and $C\geqslant 0$ such that

$$\frac{d}{ds}\tilde{\mathbb{E}}|Y_s^{t,x,y_1} - Y_s^{t,x,y_2}|^2 \leqslant -\beta\tilde{\mathbb{E}}|Y_s^{t,x,y_1} - Y_s^{t,x,y_2}|^2.$$

The comparison theorem implies that

$$\tilde{\mathbb{E}}|Y_s^{t,x,y_1} - Y_s^{t,x,y_2}|^2 \leqslant e^{-\beta s}|y_1-y_2|^2.$$

The proof is complete. □

Proposition 3.8. *Suppose that $(\mathbf{A}_k)$ holds for some $k\geqslant 2$. For any $t\in[0,T]$, $x\in\mathbb{R}^n$, $\{P_s^{t,x}\}_{s\geqslant 0}$ has a unique invariant measure $\mu^{t,x}$. Moreover,*

$$\int_{\mathbb{R}^m}|z|^k\mu^{t,x}(dz) \leqslant C_{T,k}(1+|x|^{\frac{2k}{\theta_4}}). \tag{3.6}$$

Proof. We first check (3.6). If $\mu^{t,x}$ is an invariant measure of $\{P_s^{t,x}\}_{s\geqslant 0}$, it follows from Lemma 3.6 that for all $s>0$

$$\begin{aligned}\int_{\mathbb{R}^m}|z|^k\mu^{t,x}(dz)&=\int_{\mathbb{R}^m}\tilde{\mathbb{E}}|Y_s^{t,x,z}|^k\mu^{t,x}(dz)\\&\leqslant\int_{\mathbb{R}^m}\left[e^{-\tilde{\beta}_k s}|z|^k+C_{T,k}(1+|x|^{\frac{2k}{\theta_4}})\right]\mu^{t,x}(dz)\\&=e^{-\tilde{\beta}_k s}\int_{\mathbb{R}^m}|z|^k\mu^{t,x}(dz)+C_{T,k}(1+|x|^{\frac{2k}{\theta_4}}).\end{aligned}$$

Taking s large enough such that $e^{-\tilde{\beta}_k s}\leqslant\frac{1}{2}$, we obtain (3.6).

The estimate (3.5) and the classical Bogoliubov-Krylov argument imply the existence of invariant measures. For the uniqueness, it is sufficient to prove that for any Lipschitz function $\varphi(x):\mathbb{R}^m\to\mathbb{R}$ and any invariant measure $\mu^{t,x}$ we have

$$\left|P_s^{t,x}\varphi(y)-\int_{\mathbb{R}^m}\varphi(z)\mu^{t,x}(dz)\right|\leqslant C_T Lip(\varphi)e^{-\frac{\beta s}{2}}(1+|x|^{\frac{2}{\theta_4}}+|y|),\quad s\geqslant 0,$$

where $Lip(\varphi)=\sup_{x\neq y}\frac{|\varphi(x)-\varphi(y)|}{|x-y|}$.

In fact, by Lemma 3.7 and (3.6), we have

$$\begin{aligned}\left|P_s^{t,x}\varphi(y)-\int_{\mathbb{R}^m}\varphi(z)\mu^{t,x}(dz)\right|&\leqslant\int_{\mathbb{R}^m}\left|\tilde{\mathbb{E}}\varphi(Y_s^{t,x,y})-\tilde{\mathbb{E}}\varphi(Y_s^{t,x,z})\right|\mu^{t,x}(dz)\\&\leqslant Lip(\varphi)\int_{\mathbb{R}^m}\tilde{\mathbb{E}}\left|Y_s^{t,x,y}-Y_s^{t,x,z}\right|\mu^{t,x}(dz)\\&\leqslant Lip(\varphi)\int_{\mathbb{R}^m}e^{-\frac{\beta s}{2}}|y-z|\mu^{t,x}(dz)\\&\leqslant C_T Lip(\varphi)e^{-\frac{\beta s}{2}}(1+|x|^{\frac{2}{\theta_4}}+|y|).\end{aligned}$$

Hence the proof is complete. □

Proposition 3.9. *Suppose that* $(\mathbf{A}_{2\theta_2})$ *holds. Then for any* $T>0$*, there exists* $C_T>0$ *such that any* $x\in\mathbb{R}^n, y\in\mathbb{R}^m$*,* $t\in[0,T]$ *and* $s\geqslant 0$ *,*

$$\left|\tilde{\mathbb{E}}b(t,x,Y_s^{t,x,y})-\int_{\mathbb{R}^m}b(t,x,z)\mu^{t,x}(dz)\right|\leqslant C_T e^{-\frac{\beta s}{2}}\left[(K_t(1))^2+1+|x|^{\theta}+|y|^{\theta_2+1}\right],\quad(3.7)$$

where $\theta=\max\{\frac{2\theta_2+2}{\theta_4},\frac{\theta_3\theta_4+2}{\theta_4},\frac{\theta_3(\theta_2+1)}{\theta_2}\}$.

Proof. By Lemma 3.6 and 3.7 and Proposition 3.8, for any $s \geqslant 0$ we have

$$\begin{aligned}
&\left|\tilde{\mathbb{E}} b(t,x,Y_s^{t,x,y}) - \int_{\mathbb{R}^m} b(t,x,z)\mu^{t,x}(dz)\right| \\
=& \left|\int_{\mathbb{R}^m} \tilde{\mathbb{E}} b(t,x,Y_s^{t,x,y}) - \tilde{\mathbb{E}} b(t,x,Y_s^{t,x,z})\mu^{t,x}(dz)\right| \\
\leqslant& C_T \int_{\mathbb{R}^m} \tilde{\mathbb{E}}\left[\left|Y_s^{t,x,y} - Y_s^{t,x,z}\right| (|Y_s^{t,x,y}|^{\theta_2} + |Y_s^{t,x,z}|^{\theta_2} + |x|^{\theta_3} + K_t(1))\right] \mu^{t,x}(dz) \\
\leqslant& C \int_{\mathbb{R}^m} \left[\tilde{\mathbb{E}}\left(\left|Y_s^{t,x,y} - Y_s^{t,x,z}\right|^2\right)\right]^{1/2} \left[\tilde{\mathbb{E}}\left(|Y_s^{t,x,y}|^{2\theta_2} + |Y_s^{t,x,z}|^{2\theta_2} + |x|^{2\theta_3} + [K_t(1)]^2\right)\right]^{1/2} \mu^{t,x}(dz) \\
\leqslant& C e^{-\frac{\beta s}{2}} \int_{\mathbb{R}^m} |z-y| \left[|z|^{\theta_2} + |y|^{\theta_2} + |x|^{\frac{2\theta_2}{\theta_4}} + |x|^{\theta_3} + K_t(1) + 1\right] \mu^{t,x}(dz) \\
\leqslant& C_T e^{-\frac{\beta s}{2}} \left[K_t(1)(|x|^{2/\theta_4} + |y| + 1) + |x|^{\theta} + |y|^{\theta_2+1}\right] \\
\leqslant& C_T e^{-\frac{\beta s}{2}} \left[(K_t(1))^2 + 1 + |x|^{\theta} + |y|^{\theta_2+1}\right],
\end{aligned}$$

where $\theta = \max\{\frac{2\theta_2+2}{\theta_4}, \frac{\theta_3\theta_4+2}{\theta_4}, \frac{\theta_3(\theta_2+1)}{\theta_2}\}$. The proof is complete. □

Lemma 3.10. *Suppose that* $(\mathbf{A}_{2\alpha_2})$ *holds. Then for any* $T > 0$*, there exists a constant* $C_T > 0$ *such that for all* $x_1, x_2 \in \mathbb{R}^n$*,* $y \in \mathbb{R}^m$*,* $t_1, t_2 \in [0,T]$ *and* $s \geqslant 0$*,*

$$\tilde{\mathbb{E}}|Y_s^{t_1,x_1,y} - Y_s^{t_2,x_2,y}|^2 \leqslant C_T(1 + |x_1|^{2\alpha_1} + |x_2|^{\max\{\frac{4\alpha_2}{\theta_4}, 2\alpha_1\}} + |y|^{2\alpha_2})\left(|x_1 - x_2|^2 + |t_1 - t_2|^{2\gamma_2}\right).$$

Proof. Note that

$$\begin{aligned}
Y_s^{t_1,x_1,y} - Y_s^{t_2,x_2,y} =& \int_0^s f(t_1,x_1,Y_r^{t_1,x_1,y}) - f(t_2,x_2,Y_r^{t_2,x_2,y})dr \\
&+ \int_0^s g(t_1,x_1,Y_r^{t_1,x_1,y}) - g(t_2,x_2,Y_r^{t_2,x_2,y})d\tilde{W}_r^2.
\end{aligned}$$

By Itô's formula we have

$$\begin{aligned}
&\tilde{\mathbb{E}}|Y_s^{t_1,x_1,y} - Y_s^{t_2,x_2,y}|^2 \\
=& \int_0^s \tilde{\mathbb{E}}\Big[2\langle f(t_1,x_1,Y_r^{t_1,x_1,y}) - f(t_2,x_2,Y_r^{t_2,x_2,y}), Y_r^{t_1,x_1,y} - Y_r^{t_2,x_2,y}\rangle \\
&+ \|g(t_1,x_1,Y_r^{t_1,x_1,y}) - g(t_2,x_2,Y_r^{t_2,x_2,y})\|^2\Big] dr \\
=& \int_0^s \tilde{\mathbb{E}}\Big[2\left\langle f(t_1,x_1,Y_r^{t_1,x_1,y}) - f(t_1,x_1,Y_r^{t_2,x_2,y}), Y_r^{t_1,x_1,y} - Y_r^{t_2,x_2,y}\right\rangle \\
&+ \left\|g(t_1,x_1,Y_r^{t_1,x_1,y}) - g(t_1,x_1,Y_r^{t_2,x_2,y})\right\|^2\Big] dr \\
&+ \int_0^s \tilde{\mathbb{E}}\left[2\left\langle f(t_1,x_1,Y_r^{t_2,x_2,y}) - f(t_2,x_2,Y_r^{t_2,x_2,y}), Y_r^{t_1,x_1,y} - Y_r^{t_2,x_2,y}\right\rangle\right] dr \\
&+ \int_0^s \tilde{\mathbb{E}}\left\|g(t_1,x_1,Y_r^{t_2,x_2,y}) - g(t_2,x_2,Y_r^{t_2,x_2,y})\right\|^2 dr \\
&+ \int_0^s \tilde{\mathbb{E}}\left[2\left\langle g(t_1,x_1,Y_r^{t_1,x_1,y}) - g(t_1,x_1,Y_r^{t_2,x_2,y}), g(t_1,x_1,Y_r^{t_2,x_2,y}) - g(t_2,x_2,Y_r^{t_2,x_2,y})\right\rangle\right] dr.
\end{aligned}$$

Then by Young's inequality and (2.1), there exists $\beta > 0$ such that

$$
\begin{aligned}
&\frac{d}{ds}\tilde{\mathbb{E}}|Y_s^{t_1,x_1,y} - Y_s^{t_2,x_2,y}|^2 \\
\leqslant& -\beta\tilde{\mathbb{E}}\left|Y_s^{t_1,x_1,y} - Y_s^{t_2,x_2,y}\right|^2 + C_T(|x_1-x_2|^2 + |t_1-t_2|^{2\gamma_2}) \\
&+C_T\tilde{\mathbb{E}}\left[\left|Y_s^{t_1,x_1,y} - Y_s^{t_2,x_2,y}\right|(|x_1-x_2| + |t_1-t_2|^{\gamma_2})\right] \\
&+C_T\tilde{\mathbb{E}}\left[(1+|Y_s^{t_2,x_2,y}|^{\alpha_2} + |x_1|^{\alpha_1} + |x_2|^{\alpha_1})\left|Y_s^{t_1,x_1,y} - Y_s^{t_2,x_2,y}\right|\right](|x_1-x_2| + |t_1-t_2|^{\gamma_2}) \\
\leqslant& -\frac{\beta}{2}\tilde{\mathbb{E}}\left|Y_s^{t_1,x_1,y} - Y_s^{t_2,x_2,y}\right|^2 + C_T\tilde{\mathbb{E}}\left(1+|Y_s^{t_2,x_2,y}|^{2\alpha_2} + |x_1|^{2\alpha_1} + |x_2|^{2\alpha_1}\right)\left(|x_1-x_2|^2 + |t_1-t_2|^{2\gamma_2}\right) \\
\leqslant& -\frac{\beta}{2}\tilde{\mathbb{E}}\left|Y_s^{t_1,x_1,y} - Y_s^{t_2,x_2,y}\right|^2 + C_T(1+|x_1|^{2\alpha_1} + |x_2|^{\max\{\frac{4\alpha_2}{\theta_4},2\alpha_1\}} + |y|^{2\alpha_2})\left(|x_1-x_2|^2 + |t_1-t_2|^{2\gamma_2}\right).
\end{aligned}
$$

Hence, the comparison theorem yields that

$$
\tilde{\mathbb{E}}|Y_t^{x_1,y} - Y_t^{x_2,y}|^2 \leqslant C_T(1+|x_1|^{2\alpha_1} + |x_2|^{\max\{\frac{4\alpha_2}{\theta_4},2\alpha_1\}} + |y|^{2\alpha_2})\left(|x_1-x_2|^2 + |t_1-t_2|^{2\gamma_2}\right).
$$

The proof is complete. □

3.4. The averaged equation.

Now we introduce the following averaged equation

$$
\left\{\begin{array}{l} d\bar{X}_t = \bar{b}(t,\bar{X}_t)dt + \sigma(t,\bar{X}_t)dW_t^1, \\ \bar{X}_0 = x \in \mathbb{R}^n. \end{array}\right. \tag{3.8}
$$

Here

$$
\bar{b}(t,x) = \int_{\mathbb{R}^m} b(t,x,y)\mu^{t,x}(dy),
$$

where $\mu^{t,x}$ is the unique invariant measure for Eq.(3.4).

The following lemma gives the existence, uniqueness and uniformly estimates of solutions for Eq. (3.8). The proof will be presented in the Appendix.

Lemma 3.11. *Suppose that* $(\mathbf{A}_{\tilde{\theta}})$ *holds with* $\tilde{\theta} = \max\{2\theta_2, \theta_1, \theta_4, 2\alpha_2\}$. *Then Eq.(3.8) has a unique solution. Furthermore, for any* $x \in \mathbb{R}^n$, $p \geqslant 2$ *and* $T > 0$,

$$
\mathbb{E}\left(\sup_{t\in[0,T]}|\bar{X}_t|^p\right) \leqslant C_{T,p}(1+|x|^p), \tag{3.9}
$$

where $C_{T,p}$ *is some positive constant.*

3.5. The Proof of the main result.

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

Lemma 3.12. *Assume either* $(\mathbf{H}_1)$ *with* $\lambda_1 = 0$ *and* $(\mathbf{A}_{\tilde{\theta}_1})$ *hold or* $(\mathbf{H}_1)$ *with* $\lambda_1 > 0$ *and* $(\mathbf{A}_k)$ *with some* $k > \tilde{\theta}_2$ *hold, where* $\tilde{\theta}_1 = \max\{\theta_1, 2\theta_2+2, 2\theta_6, 4\alpha_2\}$ *and* $\tilde{\theta}_2 = \max\{\theta_1, 2\theta_2+2, 2\theta_6, \theta_5\theta_4, 4\alpha_2, 2\alpha_1\theta_4\}$. *Then for any* $T > 0$, $R \geqslant R_0$ *and* $M > 0$, *there exists a constant* $C_{T,R,M,x,y} > 0$ *such that*

$$
\mathbb{E}\left(\sup_{t\in[0,T\wedge\tilde{\tau}_{R,M}^\epsilon]}|\hat{X}_t^\epsilon - \bar{X}_t|^2\right) \leqslant C_{T,R,M,x,y}\left(\frac{\epsilon}{\delta} + \delta^\gamma\right),
$$

where $\tilde{\tau}^{\epsilon}_{R,M} := \inf\{t \geqslant 0 : |\bar{X}_t| \geqslant R\} \wedge \tau^{\epsilon}_{R,M}$ *and* $\gamma = \min\{2\gamma_1, \gamma_2, 1/2\}$.

Proof. Recall that

$$
\begin{aligned}
\hat{X}^{\epsilon}_t - \bar{X}_t &= \int_0^t \left[b(s(\delta), X^{\epsilon}_{s(\delta)}, \hat{Y}^{\epsilon}_s) - \bar{b}(s, \bar{X}_s)\right] ds + \int_0^t \left[\sigma(s, X^{\epsilon}_s) - \sigma(s, \bar{X}_s)\right] dW^1_s \\
&= \int_0^t \left[b(s(\delta), X^{\epsilon}_{s(\delta)}, \hat{Y}^{\epsilon}_s) - \bar{b}(s(\delta), X^{\epsilon}_{s(\delta)})\right] ds + \int_0^t \left[\bar{b}(s(\delta), X^{\epsilon}_{s(\delta)}) - \bar{b}(s, X^{\epsilon}_{s(\delta)})\right] ds \\
&\quad + \int_0^t \left[\bar{b}(s, X^{\epsilon}_{s(\delta)}) - \bar{b}(s, X^{\epsilon}_s)\right] ds + \int_0^t \left[\bar{b}(s, X^{\epsilon}_s) - \bar{b}(s, \bar{X}_s)\right] ds \\
&\quad + \int_0^t \left[\sigma(s, X^{\epsilon}_s) - \sigma(s, \bar{X}_s)\right] dW^1_s.
\end{aligned}
$$

Then it is easy to see that

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{t\in[0,T\wedge\tilde{\tau}^{\epsilon}_{R,M}]} |\hat{X}^{\epsilon}_t - \bar{X}_t|^2\right) \\
\leqslant\ & C\mathbb{E}\left[\sup_{t\in[0,T\wedge\tilde{\tau}^{\epsilon}_{R,M}]} \left|\int_0^t b(s(\delta), X^{\epsilon}_{s(\delta)}, \hat{Y}^{\epsilon}_s) - \bar{b}(s(\delta), X^{\epsilon}_{s(\delta)}) ds\right|^2\right] \\
&+\mathbb{E}\left[\int_0^{T\wedge\tilde{\tau}^{\epsilon}_{R,M}} \left|\bar{b}(s(\delta), X^{\epsilon}_{s(\delta)}) - \bar{b}(s, X^{\epsilon}_{s(\delta)})\right| ds\right]^2 \\
&+\mathbb{E}\left[\int_0^{T\wedge\tilde{\tau}^{\epsilon}_{R,M}} \left|\bar{b}(s, X^{\epsilon}_{s(\delta)}) - \bar{b}(s, X^{\epsilon}_s)\right| ds\right]^2 + \mathbb{E}\left[\int_0^{T\wedge\tilde{\tau}^{\epsilon}_{R,M}} \left|\bar{b}(s, X^{\epsilon}_s) - \bar{b}(s, \bar{X}_s)\right| ds\right]^2 \\
&+C\mathbb{E}\int_0^{T\wedge\tilde{\tau}^{\epsilon}_{R,M}} \|\sigma(s, X^{\epsilon}_s) - \sigma(s, \bar{X}_s)\|^2 ds \\
&:= \sum_{i=1}^5 I_i(T).
\end{aligned}
\tag{3.10}
$$

For $I_2(T)$, for $t_1, t_2 \in [0,T]$ and $x \in \mathbb{R}^n$, we have

$$
\begin{aligned}
|\bar{b}(t_1, x) - \bar{b}(t_2, x)| &= \left|\int_{\mathbb{R}^m} b(t_1, x, z)\mu^{t_1,x}(dz) - \int_{\mathbb{R}^m} b(t_2, x, z)\mu^{t_2,x}(dz)\right| \\
&= \left|\int_{\mathbb{R}^m} b(t_1, x, z)\mu^{t_1,x}(dz) - \tilde{\mathbb{E}} b(t_1, x, Y^{t_1,x,0}_s)\right| \\
&\quad + \left|\tilde{\mathbb{E}} b(t_2, x, Y^{t_2,x,0}_s) - \int_{\mathbb{R}^m} b(t_2, x, z)\mu^{t_2,x}(dz)\right| \\
&\quad + \left|\tilde{\mathbb{E}} b(t_1, x, Y^{t_1,x,0}_s) - \tilde{\mathbb{E}} b(t_2, x, Y^{t_2,x,0}_s)\right|.
\end{aligned}
$$

Then Proposition 3.9 and Lemma 3.10 imply that

$$
\begin{aligned}
|\bar{b}(t_1,x)-\bar{b}(t_2,x)| &\leqslant C_T e^{-\frac{\beta s}{2}}\left[(K_{t_1}(1))^2+(K_{t_2}(1))^2+|x|^\theta+1\right]\\
&\quad +\tilde{\mathbb{E}}\left[|Y_s^{t_1,x,0}-Y_s^{t_2,x,0}|(K_{t_1}(1)+|x|^{\theta_3}+|Y_s^{t_1,x,0}|^{\theta_2}+|Y_s^{t_2,x,0}|^{\theta_2})\right]\\
&\quad +|t_1-t_2|^{\gamma_1}\tilde{\mathbb{E}}\left(|x|^{\theta_3}+|Y_s^{t_2,x,0}|^{\theta_2}+Z_T\right)\\
&\leqslant C_T e^{-\frac{\beta s}{2}}\left[(K_{t_1}(1))^2+(K_{t_2}(1))^2+|x|^\theta+1\right]\\
&\quad +\tilde{\mathbb{E}}\left[|Y_s^{t_1,x,0}-Y_s^{t_2,x,0}|(K_{t_1}(1)+|x|^{\theta_3}+|Y_s^{t_1,x,0}|^{\theta_2}+|Y_s^{t_2,x,0}|^{\theta_2})\right]\\
&\quad +|t_1-t_2|^{\gamma_1}\tilde{\mathbb{E}}\left(|x|^{\theta_3}+|Y_s^{t_2,x,0}|^{\theta_2}+Z_T\right)\\
&\leqslant C_T e^{-\frac{\beta s}{2}}\left[(K_{t_1}(1))^2+(K_{t_2}(1))^2+|x|^\theta+1\right]\\
&\quad +C_T|t_1-t_2|^{\gamma_2}\left[(1+|x|^{\frac{2\alpha_2}{\theta_4}\vee\alpha_1})(K_{t_1}(1)+|x|^{\theta_3\vee\frac{2\theta_2}{\theta_4}})\right]\\
&\quad +C_T|t_1-t_2|^{\gamma_1}\left(|x|^{\theta_3\vee\frac{2\theta_2}{\theta_4}}+Z_T\right).
\end{aligned}
$$

Then letting $s\to\infty$, there exits $\tilde{\theta}>0$ such that

$$
|\bar{b}(t_1,x)-\bar{b}(t_2,x)|\leqslant C_T\left[(K_{t_1}(1))^2+|x|^{\tilde{\theta}}+Z_T\right]|t_1-t_2|^{\gamma_1\wedge\gamma_2},
$$

which implies that

$$
\begin{aligned}
I_2(T) &\leqslant C\delta^{2(\gamma_1\wedge\gamma_2)}\mathbb{E}\left[\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}\left((K_s(1))^2+|X^\epsilon_{s(\delta)}|^{\tilde{\theta}}+Z_T\right)ds\right]^2\\
&\leqslant C_{T,R}\delta^{2(\gamma_1\wedge\gamma_2)}.
\end{aligned}\tag{3.11}
$$

For $I_3(T)$, note that for any $|x_i|\leqslant R$, $i=1,2$,

$$
|\bar{b}(t,x_1)-\bar{b}(t,x_2)|\leqslant\bar{K}_t(R)|x_1-x_2|^2,
$$

where $\bar{K}_t(R)=C_{t,R}\left[K_t(R)+K_t(1)+1\right]$ (see (5.2) below for a detailed proof). Then we have

$$
\begin{aligned}
I_3(T) &\leqslant \mathbb{E}\left[\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}[\bar{K}_s(R)]^2ds\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}|X^\epsilon_{s(\delta)}-X^\epsilon_s|^2ds\right]\\
&\leqslant C_{T,R,M}\mathbb{E}\left[\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}\left|X^\epsilon_{s(\delta)}-X^\epsilon_s\right|^2ds\right].
\end{aligned}\tag{3.12}
$$

For $I_4(T)$, we have

$$
\begin{aligned}
I_4(T) &\leqslant \mathbb{E}\left[\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}[\bar{K}_s(R)]^2ds\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}|X^\epsilon_s-\bar{X}_s|^2ds\right]\\
&\leqslant C_{T,R,M}\mathbb{E}\left[\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}\left|X^\epsilon_s-\bar{X}_s\right|^2ds\right]\\
&\leqslant C_{T,R,M}\mathbb{E}\left(\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]}|X^\epsilon_t-\hat{X}^\epsilon_t|^2\right)+C_{T,R,M}\mathbb{E}\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}|\hat{X}^\epsilon_t-\bar{X}_t|^2dt.
\end{aligned}\tag{3.13}
$$

For $I_5(T)$, it follows that

$$\begin{aligned}
I_5(T) &\leqslant \mathbb{E}\left\{\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}[\bar{K}_s(R)]^2ds\left[\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}|X^\epsilon_s-\bar{X}_s|^4ds\right]^{1/2}\right\}\\
&\leqslant C_{T,R,M}\mathbb{E}\left(\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]}|X^\epsilon_t-\hat{X}^\epsilon_t|^2\right)+\frac{1}{2}\mathbb{E}\left(\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]}|\hat{X}^\epsilon_t-\bar{X}_t|^2\right)\\
&\quad +C_{T,R,M}\mathbb{E}\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}|\hat{X}^\epsilon_t-\bar{X}_t|^2dt. \qquad (3.14)
\end{aligned}$$

By (3.10)-(3.14), we get

$$\begin{aligned}
&\mathbb{E}\left(\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]}|\hat{X}^\epsilon_t-\bar{X}_t|^2\right)\\
\leqslant\ & C_{T,R,M}\mathbb{E}\left(\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]}|X^\epsilon_t-\hat{X}^\epsilon_t|^2\right)+C_{T,R,M}\mathbb{E}\left[\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}\left|X^\epsilon_{s(\delta)}-X^\epsilon_s\right|^2ds\right]\\
&+C_{T,R,M}\mathbb{E}\int_0^{T\wedge\tilde{\tau}^\epsilon_{R,M}}|\hat{X}^\epsilon_t-\bar{X}_t|^2dt+C_{T,R,M}\delta^{2(\gamma_1\wedge\gamma_2)}+I_1(T). \qquad (3.15)
\end{aligned}$$

Next, we intend to estimate the term $I_1(T)$. Note that

$$\begin{aligned}
&\left|\int_0^t\left[b(s(\delta),X^\epsilon_{s(\delta)},\hat{Y}^\epsilon_s)-\bar{b}(s(\delta),X^\epsilon_{s(\delta)})\right]ds\right|^2\\
=&\left|\sum_{k=0}^{[t/\delta]-1}\int_{k\delta}^{(k+1)\delta}\left[b(k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_s)-\bar{b}(k\delta,X^\epsilon_{k\delta})\right]ds+\int_{t(\delta)}^t\left[b(t(\delta),X^\epsilon_{t(\delta)},\hat{Y}^\epsilon_s)-\bar{b}(t(\delta),X^\epsilon_{t(\delta)})\right]ds\right|^2\\
\leqslant&\ 2[t/\delta]\sum_{k=0}^{[t/\delta]-1}\left|\int_{k\delta}^{(k+1)\delta}\left[b(k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_s)-\bar{b}(k\delta,X^\epsilon_{k\delta})\right]ds\right|^2\\
&+2\left|\int_{t(\delta)}^t\left[b(t(\delta),X^\epsilon_{t(\delta)},\hat{Y}^\epsilon_s)-\bar{b}(t(\delta),X^\epsilon_{t(\delta)})\right]ds\right|^2\\
:=&\ I_{11}(t)+I_{12}(t). \qquad (3.16)
\end{aligned}$$

For $I_{12}(t)$, by Lemma 3.3, we easily deduce that

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]}I_{12}(t)\right]&\leqslant\delta\mathbb{E}\left[\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]}\int_{t(\delta)}^t[K_{t(\delta)}(1)]^2+|X^\epsilon_{t(\delta)}|^{2\theta_5}+|\hat{Y}^\epsilon_s|^{2\theta_6})ds\right]\\
&\leqslant\delta\left[\sup_{t\in[0,T]}\mathbb{E}[K_t(1)]^2+R^{2\theta_5}+\int_0^T\mathbb{E}|\hat{Y}^\epsilon_s|^{2\theta_6}ds\right]\\
&\leqslant C_{T,R,M}(|x|^{\frac{4\theta_6}{\theta_4}}+|y|^{2\theta_6}+1)\delta. \qquad (3.17)
\end{aligned}$$

Now, we estimate the term $I_{11}(t)$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]} I_{11}(t)\right]\\
\leqslant\ & C[T/\delta]\mathbb{E}\sum_{k=0}^{[T/\delta]-1}\left[\left|\int_{k\delta}^{(k+1)\delta}\left[b(k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_s)-\bar{b}(k\delta,X^\epsilon_{k\delta})\right]ds\right|^2 1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}\right]\\
\leqslant\ & \frac{C_T}{\delta^2}\max_{0\leqslant k\leqslant[T/\delta]-1}\mathbb{E}\left[\left|\int_{k\delta}^{(k+1)\delta}\left[b(k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_s)-\bar{b}(k\delta,X^\epsilon_{k\delta})\right]ds\right|^2 1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}\right]\\
=\ & C_T\frac{\epsilon^2}{\delta^2}\max_{0\leqslant k\leqslant[T/\delta]-1}\mathbb{E}\left[\left|\int_0^{\frac{\delta}{\epsilon}}\left[b(k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_{s\epsilon+k\delta})-\bar{b}(k\delta,X^\epsilon_{k\delta})\right]ds\right|^2 1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}\right]\\
=\ & C_T\frac{\epsilon^2}{\delta^2}\max_{0\leqslant k\leqslant[T/\delta]-1}\int_0^{\frac{\delta}{\epsilon}}\int_r^{\frac{\delta}{\epsilon}}\Psi_k(s,r)dsdr,
\end{aligned}$$

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

$$\Psi_k(s,r):=\mathbb{E}\left[\langle b(k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_{s\epsilon+k\delta})-\bar{b}(k\delta,X^\epsilon_{k\delta}),b(k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_{r\epsilon+k\delta})-\bar{b}(k\delta,X^\epsilon_{k\delta})\rangle 1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}\right].$$

For any $\epsilon,s>0$, and $\mathscr{F}_s$-measurable $\mathbb{R}^n$- resp. $\mathbb{R}^m$-valued maps X and Y, we consider the following equation

$$\tilde{Y}^{\epsilon,s,X,Y}_t=Y+\frac{1}{\epsilon}\int_s^t f(s,X,\tilde{Y}^{\epsilon,s,X,Y}_r)dr+\frac{1}{\sqrt{\epsilon}}\int_s^t g(s,X,\tilde{Y}^{\epsilon,s,X,Y}_r)dW^2_r,\quad t\geqslant s.$$

Then by the construction of $\hat{Y}^\epsilon_t$, for any $k\in\mathbb{N}_*$ and $t\in[k\delta,(k+1)\delta]$ we have

$$\hat{Y}^\epsilon_t=\tilde{Y}^{\epsilon,k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_{k\delta}}_t,$$

which implies that

$$\begin{aligned}
\Psi_k(s,r)=\mathbb{E}\Big[\Big\langle & b\left(k\delta,X^\epsilon_{k\delta},\tilde{Y}^{\epsilon,k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_{k\delta}}_{s\epsilon+k\delta}\right)-\bar{b}(k\delta,X^\epsilon_{k\delta}),\\
& b\left(k\delta,X^\epsilon_{k\delta},\tilde{Y}^{\epsilon,k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_{k\delta}}_{r\epsilon+k\delta}\right)-\bar{b}(k\delta,X^\epsilon_{k\delta})\Big\rangle 1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}\Big].
\end{aligned}$$

By approximating by functions of type $(x,y)\to H_1(x)H_2(y)$, one sees that for any measurable functions $H:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^{m\times d_2}$, $\phi:\mathbb{R}^m\to\mathbb{R}^n$, and for any $\mathscr{F}_s$-measurable $\mathbb{R}^n$-valued map X and $\mathscr{F}_t$-adapted $\mathbb{R}^m$-valued process $\{Z_t\}_{t\geqslant s}$, we have for any $t>s$,

$$\mathbb{E}\left[\phi\left(\int_s^t H(X,Z_r)dW^2_r\right)|\mathscr{F}_s\right](\omega)=\mathbb{E}\left[\phi\left(\int_s^t H(X(\omega),Z_r)dW^2_r\right)|\mathscr{F}_s\right](\omega),\ \mathbb{P}-a.s.. \quad (3.18)$$

Note that for any fixed $(x,y)\in\mathbb{R}^n\times\mathbb{R}^m$, $X^\epsilon_{k\delta}$, $\hat{Y}^\epsilon_{k\delta}$, $b(k\delta,x,y)$ and $1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}$ are $\mathscr{F}_{k\delta}$-measurable, $\{\tilde{Y}^{\epsilon,k\delta,x,y}_{s\epsilon+k\delta}\}_{s\geqslant 0}$ is independent of $\mathscr{F}_{k\delta}$, and by statement (3.18), we have

$$
\begin{aligned}
\Psi_k(s,r) = &\int_\Omega \mathbb{E}\Big[\Big\langle b\left(k\delta, X^\epsilon_{k\delta}, \tilde{Y}^{\epsilon,k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_{k\delta}}_{s\epsilon+k\delta}\right) - \bar{b}(k\delta, X^\epsilon_{k\delta}),\\
&\qquad b\left(k\delta, X^\epsilon_{k\delta}, \tilde{Y}^{\epsilon,k\delta,X^\epsilon_{k\delta},\hat{Y}^\epsilon_{k\delta}}_{r\epsilon+k\delta}\right) - \bar{b}(k\delta, X^\epsilon_{k\delta})\Big\rangle 1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}|\mathscr{F}_{k\delta}\Big](\omega)\mathbb{P}(d\omega)\\
= &\int_\Omega \Big[\mathbb{E}\Big\langle b\left(k\delta, X^\epsilon_{k\delta}(\omega), \tilde{Y}^{\epsilon,k\delta,X^\epsilon_{k\delta}(\omega),\hat{Y}^\epsilon_{k\delta}(\omega)}_{s\epsilon+k\delta}\right) - \bar{b}(k\delta, X^\epsilon_{k\delta}(\omega)),\\
&\qquad b\left(k\delta, X^\epsilon_{k\delta}(\omega), \tilde{Y}^{\epsilon,k\delta,X^\epsilon_{k\delta}(\omega),\hat{Y}^\epsilon_{k\delta}(\omega)}_{r\epsilon+k\delta}\right) - \bar{b}(k\delta, X^\epsilon_{k\delta}(\omega))\Big\rangle 1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}(\omega)\Big]\mathbb{P}(d\omega). \quad (3.19)
\end{aligned}
$$

For any given $x\in\mathbb{R}^n, y\in\mathbb{R}^m$, by the definition of process $\tilde{Y}^{\epsilon,s,x,y}_t$, it is easy to see

$$
\begin{aligned}
\tilde{Y}^{\epsilon,k\delta,x,y}_{s\epsilon+k\delta} &= y + \frac{1}{\epsilon}\int_0^{s\epsilon} f(k\delta, x, \tilde{Y}^{\epsilon,k\delta,x,y}_{r+k\delta})dr + \frac{1}{\sqrt{\epsilon}}\int_0^{s\epsilon} g(k\delta, x, \tilde{Y}^{\epsilon,k\delta,x,y}_{r+k\delta})dW^{2,k\delta}_r\\
&= y + \int_0^s f(k\delta, x, \tilde{Y}^{\epsilon,k\delta,x,y}_{r\epsilon+k\delta})dr + \int_0^s g(k\delta, x, \tilde{Y}^{\epsilon,k\delta,x,y}_{r\epsilon+k\delta})d\hat{W}^{2,k\delta}_r, \qquad (3.20)
\end{aligned}
$$

where $\{W^{2,k\delta}_r := W^2_{r+k\delta} - W^2_{k\delta}\}_{r\geqslant 0}$ and $\{\hat{W}^{2,k\delta}_t := \frac{1}{\sqrt{\epsilon}}W^{2,k\delta}_{t\epsilon}\}_{t\geqslant 0}$. Recall the solution of the frozen equation satisfies

$$
Y^{k\delta,x,y}_s = y + \int_0^s f(k\delta, x, Y^{k\delta,x,y}_r)dr + \int_0^s g(k\delta, x, Y^{k\delta,x,y}_r)d\tilde{W}^2_r. \qquad (3.21)
$$

The uniqueness of solutions of Eq. (3.20) and Eq. (3.21) implies that the distribution of $(\tilde{Y}^{\epsilon,k\delta,x,y}_{s\epsilon+k\delta})_{0\leqslant s\leqslant\delta/\epsilon}$ coincides with the distribution of $(Y^{k\delta,x,y}_s)_{0\leqslant s\leqslant\delta/\epsilon}$.

By a similar argument in Proposition 3.9 and condition (ii), we can obtain

$$
\begin{aligned}
&|b(k\delta,x,y) - \bar{b}(k\delta,x)|\\
=&\left|\int_{\mathbb{R}^m} b(k\delta,x,y) - \tilde{\mathbb{E}}b(k\delta,x,Y^{k\delta,x,z}_s)\mu^{k\delta,x}(dz)\right|\\
\leqslant& C_T\int_{\mathbb{R}^m}\tilde{\mathbb{E}}\left[\left|y - Y^{k\delta,x,z}_s\right|(|y|^{\theta_2} + |Y^{k\delta,x,z}_s|^{\theta_2} + |x|^{\theta_3} + K_{k\delta}(1))\right]\mu^{k\delta,x}(dz)\\
\leqslant& C_T\int_{\mathbb{R}^m}\left[\tilde{\mathbb{E}}\left(\left|y - Y^{k\delta,x,z}_s\right|^2\right)\right]^{1/2}\left[\tilde{\mathbb{E}}\left(|Y^{k\delta,x,y}_s|^{2\theta_2} + |Y^{k\delta,x,z}_s|^{2\theta_2} + |x|^{2\theta_3} + [K_t(1)]^2\right)\right]^{1/2}\mu^{k\delta,x}(dz)\\
\leqslant& C_T\int_{\mathbb{R}^m}(|y| + |z| + |x|^{2/\theta_4})\left[|z|^{\theta_2} + |y|^{\theta_2} + |x|^{\frac{2\theta_2}{\theta_4}} + |x|^{\theta_3} + K_t(1) + 1\right]\mu^{k\delta,x}(dz)\\
\leqslant& C_T\left[(K_{k\delta}(1))^2 + 1 + |x|^{\frac{4\theta_2}{\theta_4}\vee(2\theta_3)} + |y|^{\theta_2+1}\right]. \qquad (3.22)
\end{aligned}
$$

Then by (3.19), (3.22) and Proposition 3.9, we have

$$\begin{aligned}
\Psi_k(s,r) = &\int_\Omega \Big[\tilde{\mathbb{E}}\Big\langle b\left(k\delta, X^\epsilon_{k\delta}(\omega), Y_s^{k\delta,X^\epsilon_{k\delta}(\omega),\hat{Y}^\epsilon_{k\delta}(\omega)}\right) - \bar{b}(k\delta, X^\epsilon_{k\delta}(\omega)),\\
&\qquad b\left(k\delta, X^\epsilon_{k\delta}(\omega), Y_r^{k\delta,X^\epsilon_{k\delta}(\omega),\hat{Y}^\epsilon_{k\delta}(\omega)}\right) - \bar{b}(k\delta, X^\epsilon_{k\delta}(\omega))\Big\rangle 1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}(\omega)\Big]\mathbb{P}(d\omega)\\
= &\int_\Omega\int_{\tilde{\Omega}} \Big\langle \tilde{\mathbb{E}}\Big[b\left(k\delta, X^\epsilon_{k\delta}(\omega), Y_{s-r}^{k\delta,X^\epsilon_{k\delta}(\omega),Y_r^{k\delta,X^\epsilon_{k\delta}(\omega),\hat{Y}^\epsilon_{k\delta}(\omega)}(\tilde{\omega})}\right) - \bar{b}(k\delta, X^\epsilon_{k\delta}(\omega))\Big],\\
&\qquad b\left(k\delta, X^\epsilon_{k\delta}(\omega), Y_r^{k\delta,X^\epsilon_{k\delta}(\omega),\hat{Y}^\epsilon_{k\delta}(\omega)}(\tilde{\omega})\right) - \bar{b}(k\delta, X^\epsilon_{k\delta}(\omega))\Big\rangle 1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}(\omega)\tilde{\mathbb{P}}(d\tilde{\omega})\mathbb{P}(d\omega)\\
\leqslant &\int_\Omega\int_{\tilde{\Omega}} \Big[(K_{k\delta}(1))^2 + 1 + |X^\epsilon_{k\delta}(\omega)|^{(\frac{2\theta_2+2}{\theta_4})\vee(\frac{\theta_3\theta_4+2}{\theta_4})\vee(\frac{\theta_3(\theta_2+1)}{\theta_2})} + |Y_r^{k\delta,X^\epsilon_{k\delta}(\omega),\hat{Y}^\epsilon_{k\delta}(\omega)}(\tilde{\omega})|^{\theta_2+1}\Big] e^{-\frac{(s-r)\beta}{2}}\\
&\quad\cdot\Big[(K_{k\delta}(1))^2 + 1 + |X^\epsilon_{k\delta}(\omega)|^{\frac{4\theta_2}{\theta_4}\vee(2\theta_3)} + |Y_r^{k\delta,X^\epsilon_{k\delta}(\omega),\hat{Y}^\epsilon_{k\delta}(\omega)}(\tilde{\omega})|^{\theta_2+1}\Big] 1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}(\omega)\tilde{\mathbb{P}}(d\tilde{\omega})\mathbb{P}(d\omega)\\
\leqslant &\, C_T\int_\Omega \Big[(K_{k\delta}(1))^4 + |X^\epsilon_{k\delta}(\omega)|^{\frac{8\theta_2}{\theta_4}\vee(4\theta_3)} + |\hat{Y}^\epsilon_{k\delta}(\omega)|^{2\theta_2+2} + 1)1_{\{k\delta\leqslant\tilde{\tau}^\epsilon_{R,M}\}}(\omega)\Big]\mathbb{P}(d\omega)e^{-\frac{(s-r)\beta}{2}}\\
\leqslant &\, C_{T,R}(|x|^{\frac{4(\theta_2+1)}{\theta_4}} + |y|^{2(\theta_2+1)} + 1)e^{-\frac{(s-r)\beta}{2}},
\end{aligned}$$

where the last inequality comes from the definition of stopping time, Lemmas 3.1 and 3.3.

Hence we have

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]} I_1(t)\right] &\leqslant C_{T,R,x,y}\frac{\epsilon^2}{\delta^2}\int_0^{\frac{\delta}{\epsilon}}\int_r^{\frac{\delta}{\epsilon}} e^{-\frac{(s-r)\beta}{2}}dsdr + C_{T,R,M,x,y}\delta\\
&= C_{T,R,x,y}\frac{\epsilon^2}{\delta^2}\Big(\frac{\delta}{\beta\epsilon} - \frac{1}{\beta^2} + \frac{1}{\beta^2}e^{-\frac{\beta\delta}{\epsilon}}\Big) + C_{T,R,M,x,y}\delta\\
&\leqslant C_{T,R,x,y}\frac{\epsilon}{\delta} + C_{T,R,M,x,y}\delta. \qquad (3.23)
\end{aligned}$$

According to estimates (3.15) and (3.23), we obtain that

$$\begin{aligned}
\mathbb{E}\left(\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]} |\hat{X}^\epsilon_t - \bar{X}_t|^2\right) &\leqslant C_{T,R,M,x,y}\left(\frac{\epsilon}{\delta} + \delta^\gamma\right)\\
&\quad + C_{T,R,M}\int_0^T \mathbb{E}\left(\sup_{s\in[0,t\wedge\tilde{\tau}^\epsilon_{R,M}]}\left|\hat{X}^\epsilon_s - \bar{X}_s\right|^2\right)dt,
\end{aligned}$$

where $\gamma = \min\{2\gamma_1, \gamma_2, 1/2\}$. By Gronwall's inequality, we get

$$\mathbb{E}\Big(\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]} |\hat{X}^\epsilon_t - \bar{X}_t|^2\Big) \leqslant C_{T,R,M,x,y}\left(\frac{\epsilon}{\delta} + \delta^\gamma\right).$$

Hence the proof is complete. □

Now we can finish the proof of our main result.

Proof of Theorem 2.3 Taking $\delta = \epsilon^{\tilde{\gamma}}$ with $\tilde{\gamma} = (1 + \min\{2\gamma_1, \gamma_2, 1/2\})^{-1}$, Lemmas 3.5 and 3.12 imply that

$$\begin{aligned}\mathbb{E}\left(\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]}|X^\epsilon_t - \bar{X}_t|\right) &\leqslant \mathbb{E}\left(\sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]}|X^\epsilon_t - \hat{X}^\epsilon_t| + \sup_{t\in[0,T\wedge\tilde{\tau}^\epsilon_{R,M}]}|\hat{X}^\epsilon_t - \bar{X}_t|\right)\\ &\leqslant C_{T,R,M,x,y}\left(\sqrt{\epsilon\delta^{-1}} + \delta^{\frac{1}{2}\min\{2\gamma_1,\gamma_2,1/2\}}\right)\\ &\leqslant C_{T,R,M,x,y}\epsilon^{\frac{1-\tilde{\gamma}}{2}}. \end{aligned}\tag{3.24}$$

By Chebyshev's inequality, Lemmas 3.3 and 3.11, we have

$$\begin{aligned}&\mathbb{E}\left(\sup_{t\in[0,T]}|X^\epsilon_t - \bar{X}_t|1_{\{T>\tilde{\tau}^\epsilon_{R,M}\}}\right)\\ \leqslant&\left[\mathbb{E}\left(\sup_{t\in[0,T]}|X^\epsilon_t - \bar{X}_t|^2\right)\right]^{\frac{1}{2}}\cdot\left[\mathbb{P}(T>\tilde{\tau}^\epsilon_{R,M})\right]^{\frac{1}{2}}\\ \leqslant&\frac{C_{T,x,y}}{R^{1/2}}\left[\mathbb{E}\left(\sup_{t\in[0,T]}|X^\epsilon_t| + \sup_{t\in[0,T]}|\bar{X}_t| + \int_0^T|Y^\epsilon_s|^{2\theta_2}ds + \int_0^T|\hat{Y}^\epsilon_s|^{4\theta_1\vee 2\theta_2}ds\right)\right]^{1/2}\\ &+\frac{C_{T,x,y}}{M^{1/2}}\left[\mathbb{E}\int_0^T[K_s(R)]^4ds\right]^{1/2}\\ \leqslant&\frac{C_{T,x,y}}{R^{1/2}} + \frac{C_{T,R,x,y}}{M^{1/2}}. \end{aligned}\tag{3.25}$$

Hence, by (3.24) and (3.25), we obtain that

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X^\epsilon_t - \bar{X}_t|\right) \leqslant C_{T,R,M,x,y}\epsilon^{\frac{1-\tilde{\gamma}}{2}} + \frac{C_{T,x,y}}{R^{1/2}} + \frac{C_{T,R,x,y}}{M^{1/2}}.$$

Now, letting $\epsilon \to 0$ firstly, $M \to \infty$ secondly, and $R \to \infty$ finally, we obtain that

$$\lim_{\epsilon\to 0}\mathbb{E}\left(\sup_{t\in[0,T]}|X^\epsilon_t - \bar{X}_t|\right) = 0. \tag{3.26}$$

On the one hand, if $\lambda_1 = 0$ in $(\mathbf{H}_1)$ and $(\mathbf{A}_{\tilde{\theta}_1})$ holds with $\tilde{\theta}_1 = \max\{4\theta_1, 2\theta_2 + 2, 2\theta_6, 4\alpha_2\}$, then by Lemma 3.1 and 3.11 we have

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

On the other hand, if $\lambda_1 > 0$ in $(\mathbf{H}_1)$ and $(\mathbf{A}_k)$ with some $k > \tilde{\theta}_2$ holds, where $\tilde{\theta}_2 = \max\{4\theta_1, (2\theta_2 + 2), 2\theta_6, 4\alpha_2, \theta_5\theta_4, 2\alpha_1\theta_4\}$. By Lemmas 3.1 and 3.11, for any $k' < k$ we have

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X^\epsilon_t - \bar{X}_t|^{2k'/\theta_4}\right) \leqslant C_k\mathbb{E}\left(\sup_{t\in[0,T]}|X^\epsilon_t|^{2k'/\theta_4} + \sup_{t\in[0,T]}|\bar{X}_t|^{2k'/\theta_4}\right) < \infty.$$

Hence by Hölder's inequality and (3.26), it is easy to prove that (2.4) and (2.5) hold. Therefore, the proof is complete. □

4. Examples

In this section, we give several concrete examples to illustrate the applicability of our main result. We concentrate on cases which are not covered by previous papers in the literature. For simplicity, we only consider the 1-dimensional case, but one can easily extend to the multi-dimensional case.

Example 4.1. Let us consider the following slow-fast SDEs,

$$\left\{\begin{array}{l} dX^{\epsilon}_t = \left[-(X^{\epsilon}_t)^3 + X^{\epsilon}_t + (Y^{\epsilon}_t)^3\right]dt + X^{\epsilon}_t dW^1_t, \quad X^{\epsilon}_0 = x \in \mathbb{R}, \\ dY^{\epsilon}_t = \dfrac{1}{\epsilon}\left[-(X^{\epsilon}_t)^2(Y^{\epsilon}_t)^3 - 3Y^{\epsilon}_t - (Y^{\epsilon}_t)^5\right]dt + \dfrac{1}{\sqrt{\epsilon}}\left[\sin(X^{\epsilon}_t) + \sin(Y^{\epsilon}_t)\right]dW^2_t, \quad Y^{\epsilon}_0 = y \in \mathbb{R}, \end{array}\right. \tag{4.1}$$

where $\{W^1_t\}_{t\geqslant 0}$ and $\{W^2_t\}_{t\geqslant 0}$ are independent 1-dimensional Brownian motions.

Let

$$b(x,y) = -x^3 + x + y^3, \quad \sigma(x) = x$$

and

$$f(x,y) = -x^2y^3 - 3y - y^5, \quad g(x,y) = \sin x + \sin y.$$

It is easy to verify that $(\mathbf{H}_1)$ with $\theta_4 = 6$, $(\mathbf{H}_2)$ and $(\mathbf{A}_k)$ with any $k \geqslant 2$ hold.

Hence, by Theorem 2.3 for any $p > 0$ we have

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

where $\bar{X}_t$ is the solution of the corresponding averaged equation.

Example 4.2. Let us consider the following slow-fast SDEs,

$$\left\{\begin{array}{l} dX^{\epsilon}_t = \left[t^2 X^{\epsilon}_t - (X^{\epsilon}_t)^3(Y^{\epsilon}_t)^2 + \lambda_1 Y^{\epsilon}_t\right]dt + \left(t^2 + X^{\epsilon}_t\right)dW^1_t, \quad X^{\epsilon}_0 = x \in \mathbb{R}, \\ dY^{\epsilon}_t = \dfrac{1}{\epsilon}\left(\sqrt{t}X^{\epsilon}_t - 8Y^{\epsilon}_t\right)dt + \dfrac{1}{\sqrt{\epsilon}}(t + X^{\epsilon}_t + Y^{\epsilon}_t)dW^2_t, \quad Y^{\epsilon}_0 = y \in \mathbb{R}, \end{array}\right. \tag{4.2}$$

where $\lambda_1 \geqslant 0$, $\{W^1_t\}_{t\geqslant 0}$ and $\{W^2_t\}_{t\geqslant 0}$ are independent 1-dimensional Brownian motion.

Let

$$b(t,x,y) = t^2x - x^3y^2 + \lambda_1 y, \quad \sigma(t,x) = t^2 + x$$

and

$$f(t,x,y) = \sqrt{t}x - 8y, \quad g(t,x,y) = t + x + y.$$

It is easy to verify that $(\mathbf{H}_1)$ holds with $\theta_1 = 2, \theta_2 = 1, \theta_3 = 3, \theta_4 = 2, \theta_5 = 6, \theta_6 = 4$, $\gamma_1 = 1$, $Z_T = 0$ and $K_t(R) = 6R^2 + 2t^4 + 2$; $(\mathbf{H}_2)$ holds with $\alpha_i = 1, i = 1,2,3,4$, and $\gamma_2 = 1/2$; $(\mathbf{A}_k)$ holds with any $2 \leqslant k < 17$.

Hence, by Theorem 2.3, if $\lambda_1 = 0$, for any $p > 0$ we have

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

Moreover, if $\lambda_1 > 0$, for any $0 < p < 17$ we have

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

where $\bar{X}_t$ is the solution of the corresponding averaged equation.

Example 4.3. Let us consider the following slow-fast SDEs,

$$\left\{\begin{array}{l} dX^\epsilon_t=\left[-|\sin(W^1_t)|(X^\epsilon_t)^3+Y^\epsilon_t\right]dt+X^\epsilon_t dW^1_t,\quad X^\epsilon_0=x\in\mathbb{R},\\ dY^\epsilon_t=\frac{1}{\epsilon}\left[X^\epsilon_t-8Y^\epsilon_t\right]dt+\frac{1}{\sqrt{\epsilon}}Y^\epsilon_t dW^2_t,\quad Y^\epsilon_0=y\in\mathbb{R}, \end{array}\right. \tag{4.3}$$

where $\{W^1_t\}_{t\geqslant 0}$ and $\{W^2_t\}_{t\geqslant 0}$ are independent 1-dimensional Brownian motions.

Assume that

$$b(t,x,y,\omega)=-|\sin(W^1_t(\omega))|x^3+y,\quad \sigma(x)=x$$

and

$$f(x,y)=x-8y,\quad g(x,y)=y.$$

It is easy to verify that $(\mathbf{H}_1)$ holds with $\theta_1=0,\theta_2=1,\theta_3=3,\theta_4=2,\theta_5=3,\theta_6=1$, $\gamma_1<1/2$, $K_t(R)=6R^2+1$ and $Z_T=\sup_{0\leqslant s<t\leqslant T}\frac{|W^1_t-W^1_s|}{|t-s|^{\gamma_1}}$ with $\mathbb{E}Z^2_T<\infty$ by Kolmogorov's continuity theorem ; $(\mathbf{H}_2)$ holds with $\alpha_i=1, i=1,2,3,4$; $(\mathbf{A}_k)$ holds with any $2\leqslant k<17$.

Hence, by Theorem 2.3, for any $0<p<17$ we have

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

where $\bar{X}_t$ is the solution of corresponding averaged equation.

5. Appendix

In this section, using the classical result of Krylov (cf. [14, Theorem 3.1.1]), we prove the existence and uniqueness of solutions to system (1.1) and the corresponding averaged equation.

5.1. **Proof of Theorem 2.2.**

Proof. We denote

$$Z^\epsilon_t=\left(\begin{array}{c} X^\epsilon_t\\ Y^\epsilon_t\end{array}\right),\quad \tilde{b}^\epsilon(t,x,y)=\left(\begin{array}{c} b(t,x,y)\\ \frac{1}{\epsilon}f(t,x,y)\end{array}\right)$$

and

$$\tilde{\sigma}^\epsilon(t,x,y)=\left(\begin{array}{cc}\sigma(t,x) & 0\\ 0 & \frac{1}{\sqrt{\epsilon}}g(t,x,y)\end{array}\right),\quad W_t=\left(\begin{array}{c} W^1_t\\ W^2_t\end{array}\right).$$

Then system (1.1) can be rewritten as the following equation

$$dZ^\epsilon_t=\tilde{b}^\epsilon(t,Z^\epsilon_t)dt+\tilde{\sigma}^\epsilon(t,Z^\epsilon_t)dW_t,\quad Z^\epsilon_0=\left(\begin{array}{c} x\\ y\end{array}\right). \tag{5.1}$$

Under the assumptions $(\mathbf{H}_1)$ and $(\mathbf{H}_2)$, we intend to prove the coefficients in Eq. (5.1) satisfy the local weak monotonicity and weak coercivity conditions in [14, Theorem 3.1.1].

In fact, for any $t, R > 0$, $z_i = (x_i, y_i) \in \mathbb{R}^n \times \mathbb{R}^m$ with $|z_i| \leqslant R$, $i = 1, 2$,

$$
\begin{aligned}
& 2\langle\tilde{b}^\epsilon(t,z_1) - \tilde{b}^\epsilon(t,z_2), z_1 - z_2\rangle + \|\tilde{\sigma}^\epsilon(t,z_1) - \tilde{\sigma}^\epsilon(t,z_2)\|^2 \\
\leqslant\ & 2\langle b(t,x_1,y_1) - b(t,x_2,y_2), x_1 - x_2\rangle + \|\sigma(t,x_1) - \sigma(t,x_2)\|^2 \\
& + \frac{2}{\epsilon}\langle f(t,x_1,y_1) - f(t,x_2,y_2), y_1 - y_2\rangle + \frac{1}{\epsilon}\|g(t,x_1,y_1) - g(t,x_2,y_2)\|^2 \\
\leqslant\ & 2|b(t,x_1,y_1) - b(t,x_2,y_1)| \cdot |x_1 - x_2| + \|\sigma(t,x_1) - \sigma(t,x_2)\|^2 \\
& + 2|b(t,x_2,y_1) - b(t,x_2,y_2)||x_1 - x_2| \\
& + \frac{2}{\epsilon}\langle f(t,x_1,y_1) - f(t,x_1,y_2), y_1 - y_2\rangle + \frac{1}{\epsilon}\|g(t,x_1,y_1) - g(t,x_1,y_2)\|^2 \\
& + \frac{1}{\epsilon}\|g(t,x_1,y_2) - g(t,x_2,y_2)\|^2 + \frac{2}{\epsilon}\|g(t,x_1,y_1) - g(t,x_1,y_2)\|\|g(t,x_1,y_2) - g(t,x_2,y_2)\| \\
& + \frac{2}{\epsilon}|f(t,x_1,y_2) - f(t,x_2,y_2)||y_1 - y_2| \\
\leqslant\ & K_t(R)(1 + R^{\theta_1})|x_1 - x_2|^2 + 2(2R^{\theta_2} + K_t(1) + R^{\theta_3})|y_1 - y_2| \cdot |x_1 - x_2| \\
& + \frac{C_t}{\epsilon}(1 + 2R^{\alpha_1} + R^{\alpha_2})|x_1 - x_2||y_1 - y_2| + \frac{C_t}{\epsilon}|z_1 - z_2|^2 \\
\leqslant\ & C_R\left[K_t(R) + K_t(1) + \frac{C_t}{\epsilon}\right]|z_1 - z_2|^2.
\end{aligned}
$$

Furthermore, let $\epsilon_0 = \frac{\lambda_2}{\lambda_1}$ if $\lambda_1 > 0$, and $\epsilon_0 = 1$ otherwise. Then for any $\epsilon \in (0, \epsilon_0)$

$$
\begin{aligned}
& 2\langle\tilde{b}^\epsilon(t,z_1), z_1\rangle + \|\tilde{\sigma}^\epsilon(t,z_1)\|^2 \\
\leqslant\ & 2\langle b(t,x_1,y_1), x_1\rangle + \|\sigma(t,x_1)\|^2 + \frac{2}{\epsilon}\langle f(t,x_1,y_1), y_1\rangle + \frac{1}{\epsilon}\|g(t,x_1,y_1)\|^2 \\
\leqslant\ & K_t(1)(1 + |x_1|^2) + \lambda_1|y_1|^{\theta_4} + K_t(1) + C|x_1|^2 - \frac{\lambda_2|y_1|^{\theta_4}}{\epsilon} + \frac{C_t}{\epsilon}(1 + |x_1|^{\frac{4}{\theta_4}}) \\
\leqslant\ & C\left[2K_t(1) + \frac{C_t}{\epsilon}\right](1 + |z_1|^2).
\end{aligned}
$$

Let

$$K_t^\epsilon(R) := C_R\left[K_t(R) + K_t(1) + \frac{C_t}{\epsilon}\right].$$

Then by the definition of $K_t(R)$, it is easy to see that $K_t^\epsilon(R)$ is an $\mathbb{R}_+$-valued adapted process and for all R, T, $\epsilon \in (0, \epsilon_0)$,

$$\int_0^T K_t^\epsilon(R)dt < \infty.$$

Hence by [14, Theorem 3.1.1], there exists a unique solution $\{(X_t^\epsilon, Y_t^\epsilon), t \geqslant 0\}$ to system (1.1). The proof is complete. □

5.2. **Proof of Lemma 3.11.**

Proof. It is sufficient to check that the coefficients of Eq. (3.8) satisfy the following conditions:

For any $t \geqslant 0, x_1, x_2 \in \mathbb{R}^n, R > 0$ with $|x_i| \leqslant R$,

$$2|\bar{b}(t,x_1) - \bar{b}(t,x_2)||x_1 - x_2| + \|\sigma(t,x_1) - \sigma(t,x_2)\|^2 \leqslant \bar{K}_t(R)|x_1 - x_2|^2 \tag{5.2}$$

and

$$2\langle x_1, \bar{b}(t,x_1)\rangle + \|\sigma(t,x_1)\|^2 \leqslant \bar{K}_t(1)(1+|x_1|^2), \tag{5.3}$$

where $\bar{K}_t(R)$ is an $\mathbb{R}_+$-valued adapted process and for all $R, T>0$,

$$\int_0^T \bar{K}_t(R)dt < \infty.$$

Then Eq.(3.8) has a unique solution and (3.9) can be easily obtained by following the same arguments as in Lemma 3.1(i).

In fact, by Proposition 3.9 and Lemma 3.10 we have

$$\begin{aligned}
&2|\bar{b}(t,x_1)-\bar{b}(t,x_2)||x_1-x_2| + \|\sigma(t,x_1)-\sigma(t,x_2)\|^2\\
\leqslant\ & 2\left|\int_{\mathbb{R}^m} b(t,x_1,z)\mu^{t,x_1}(dz) - \int_{\mathbb{R}^m} b(t,x_2,z)\mu^{t,x_2}(dz)\right| |x_1-x_2| + \|\sigma(t,x_1)-\sigma(t,x_2)\|^2\\
\leqslant\ & 2\left[\left|\int_{\mathbb{R}^m} b(t,x_1,z)\mu^{t,x_1}(dz) - \tilde{\mathbb{E}}b(t,x_1,Y_s^{x_1,0})\right| + \left|\tilde{\mathbb{E}}b(t,x_2,Y_s^{x_2,0}) - \int_{\mathbb{R}^m} b(t,x_2,z)\mu^{t,x_2}(dz)\right|\right] |x_1-x_2|\\
&+2\tilde{\mathbb{E}}\left|b(t,x_1,Y_s^{t,x_1,0}) - b(t,x_2,Y_s^{t,x_1,0})\right| |x_1-x_2| + \|\sigma(t,x_1)-\sigma(t,x_2)\|^2\\
&+2\tilde{\mathbb{E}}\left|b(t,x_2,Y_s^{t,x_1,0}) - b(t,x_2,Y_s^{t,x_2,0})\right| |x_1-x_2|\\
\leqslant\ & C_t e^{-\frac{\beta s}{2}}\left[(K_t(1))^2 + 1 + |x_1|^{(\frac{2\theta_2+2}{\theta_4})\vee(\frac{\theta_3\theta_4+2}{\theta_4})} + |x_2|^{(\frac{2\theta_2+2}{\theta_4})\vee(\frac{\theta_3\theta_4+2}{\theta_4})}\right]\\
&+C|x_1-x_2|^2 K_t(R)\tilde{\mathbb{E}}(1+|Y_s^{t,x_1,0}|^{\theta_1})\\
&+|x_1-x_2|\tilde{\mathbb{E}}\left[(|Y_s^{t,x_1,0}|^{\theta_2} + |Y_s^{t,x_2,0}|^{\theta_2} + |x_2|^{\theta_3} + K_t(1))|Y_s^{t,x_1,0} - Y_s^{t,x_2,0}|\right]\\
\leqslant\ & C_t e^{-\frac{\beta s}{2}}\left[(K_t(1))^2 + 1 + |x_1|^{(\frac{2\theta_2+2}{\theta_4})\vee(\frac{\theta_3\theta_4+2}{\theta_4})} + |x_2|^{(\frac{2\theta_2+2}{\theta_4})\vee(\frac{\theta_3\theta_4+2}{\theta_4})}\right]\\
&+C_{t,R}|x_1-x_2|^2 K_t(R) + C_{t,R}(1+K_t(1))|x_1-x_2|^2.
\end{aligned}$$

Then letting $s\to\infty$, we obtain

$$2|\bar{b}(t,x_1)-\bar{b}(t,x_2)||x_1-x_2| + \|\sigma(t,x_1)-\sigma(t,x_2)\|^2 \leqslant C_{t,R}\left[K_t(R)+K_t(1)+1\right]|x_1-x_2|^2.$$

Moreover, by (3.6) we have

$$\begin{aligned}
&2\langle\bar{b}(t,x_1),x_1\rangle + \|\sigma(t,x_1)\|^2\\
=\ & \int_{\mathbb{R}^m}\left[\langle 2b(t,x_1,z),x_1\rangle + \|\sigma(t,x_1)\|^2\right]\mu^{t,x_1}(dz)\\
=\ & \int_{\mathbb{R}^m} K_t(1)\left(1+|x_1|^2+\lambda_1|z|^{\theta_4}\right)\mu^{t,x_1}(dz)\\
\leqslant\ & C_t K_t(1)(1+|x_1|^2).
\end{aligned}$$

Then (5.2) and (5.3) hold by taking

$$\bar{K}_t(R) := C_{t,R}\left[K_t(R)+K_t(1)+1\right].$$

By the definition of $K_t(R)$, it is easy to see that $\bar{K}_t(R)$ is an $\mathbb{R}_+$-valued adapted process and for all $R, T>0$,

$$\int_0^T \bar{K}_t(R)dt < \infty.$$

Hence by [14, Theorem 3.1], there exists a unique solution $\{\bar{X}_t, t \geqslant 0\}$ to Eq. (3.8). The estimate (3.9) can be proved by the same arguments as in Lemma 3.1. Therefore, the proof is complete. □

Acknowledgement This work was 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. Financial support of the DFG through CRC 1283 is also gratefully acknowledged.

References

[1] Bertram R., Rubin J. E., Multi-timescale systems and fast-slow analysis. *Math. Biosci.* 287 (2017), 105-121.

[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., 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.

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

[5] E W., Engquist B. Multiscale modeling and computations. *Notice of AMS*, 50 (2003), 1062-1070.

[6] E W., Liu D., Vanden-Eijnden.E.: Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.* 58 (11) (2005) 1544-1585.

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

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

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

[10] Harvey E., Kirk V., Wechselberger M., Sneyd J., Multiple timescales, mixed mode oscillations and canards in models of intracellular calcium dynamics, *J. Nonlinear Sci.* 21 (2011) 639-683.

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

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

[13] Pavliotis G. A. and Stuart A. M., Multiscale methods: averaging and homogenization, volume 53 of Texts in Applied Mathematics. Springer, New York, 2008.

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

[15] Veretennikov A.Y., On the averaging principle for systems of stochastic differential equations, Math. USSR Sborn. 69 (1991) 271-284.

[16] Wu F., Tian T., Rawlings J. B., Yin,G., Approximate method for stochastic chemical kinetics with two-time scales by chemical Langevin equations. *J. Chem. Phys*, 2016, 144(17): 174112.

[17] Xu J., Liu J., Miao Y., Strong averaging principle for two-time-scale SDEs with non-Lipschitz coefficients, *J. Math. Anal. Appl.* 468 (2018), no. 1, 116-140.

[18] Xu Y., Pei B., Guo R., Stochastic averaging for slow-fast dynamical systems with fractional Brownian motion. *Discrete Contin. Dyn. Syst. Ser. B* 20 (2015), no. 7, 2257-2267.

Current address, Liu, W.: School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, 221116, China
Email address: weiliu@jsnu.edu.cn

Current address, Röckner, M.: Fakultät für Mathematik, Universität Bielefeld, D-33501 Bielefeld, Germany
Email address: roeckner@math.uni-bielefeld.de

Current address, Sun, X.: School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, 221116, China
Email address: xbsun@jsnu.edu.cn

Current address, Xie, Y.: School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, 221116, China
Email address: ycxie@jsnu.edu.cn