COMMUN. MATH. SCI.
Vol. 8, No. 4, pp. 999–1020


# STRONG CONVERGENCE OF PRINCIPLE OF AVERAGING FOR MULTISCALE STOCHASTIC DYNAMICAL SYSTEMS*

DI LIU†

**Abstract.** In this paper, we study stochastic differential equations with two well-separated time scales. We prove that the rate of strong convergence to the averaged effective dynamics is of order $O(\varepsilon^{1/2})$ , where $\varepsilon \ll 1$ is the parameter measuring the disparity of the time scales in the system. The convergence rate is shown to be optimal through examples.

**Key words.** Stochastic Differential Equations, Time Scale Separation, Averaging of Perturbations.

**AMS subject classifications.** 60H10, 60H10, 70K65.

## 1. Introduction

Consider the following stochastic dynamical system with a time scale separation measured by $\varepsilon \ll 1$:

$$\begin{aligned} \dot{X}_t^\varepsilon &= a\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) + \sigma\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big)\dot{W}_t, \qquad && X_0^\varepsilon = x, \\ \dot{Y}_t^\varepsilon &= \frac{1}{\varepsilon} B\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big) + \frac{1}{\sqrt{\varepsilon}} C\big(X_t^\varepsilon, Y_t^\varepsilon, \varepsilon\big)\dot{W}_t, \qquad && Y_0^\varepsilon = y, \end{aligned} \tag{1.1}$$

where $X_t^\varepsilon \in \mathbb{R}^n$, and $Y_t^\varepsilon \in \mathbb{R}^m$ are variables in vector spaces and $W_t$ is a standard $d$-dimensional Wiener process. $a(\cdot) \in \mathbb{R}^n$, $B(\cdot) \in \mathbb{R}^m$, $\sigma(\cdot) \in \mathbb{R}^n \times \mathbb{R}^d$ and $C(\cdot) \in \mathbb{R}^m \times \mathbb{R}^d$ are all functions of $O(1)$ magnitude. Systems in the form of (1.1) arise from a wide range of applications including chemical kinetics, material sciences, fluid dynamics, and finance. We have assumed that the phase space can be decomposed into slow degrees of freedom $x$ and fast degrees of freedom $y$. Under appropriate assumptions on $B(\cdot)$ and $C(\cdot)$, the dynamics for $Y_t^\varepsilon$ with $X_t^\varepsilon = x$ fixed is ergodic with a unique invariant measure $\mu_x^\varepsilon(dy)$. In this case, the Principle of Averaging has been proved such that in the limit of $\varepsilon \to 0$, $X_s^\varepsilon$ converges to a stochastic differential equation of the following form:

$$\dot{\bar{X}}_t = \bar{a}\big(\bar{X}_t\big) + \bar{\sigma}\big(\bar{X}_t\big)\dot{W}_t, \qquad \bar{X}_0 = x, \tag{1.2}$$

where

$$\begin{aligned} \bar{a}(x) &= \lim_{\varepsilon \to 0} \int a(x,y,\varepsilon)\mu_x^\varepsilon(dy), \\ \bar{\sigma}(x)\bar{\sigma}^T(x) &= \lim_{\varepsilon \to 0} \int \sigma(x,y,\varepsilon)\sigma^T(x,y,\varepsilon)\mu_x^\varepsilon(dy). \end{aligned} \tag{1.3}$$

From the point of view of numerical analysis, an important question is in which sense, as well as how fast, the system will converge to the effective dynamics. The recent motivation for this problem is the progress on numerical methods for dynamical systems with multiple time scales. In [12], a multiscale integration scheme was

*Received: November 13, 2008; accepted (in revised version): February 17, 2010. Communicated by Jonathan Mattingly.
Supported by NSF-DMS 0609315.
†Department of Mathematics, Michigan State University, USA (richardl@math.msu.edu). URL:www.math.msu.edu/∼richardl

proposed to deal with systems in the form of (1.1) by solving the effective dynamics (1.2). Fitting into the framework of Heterogeneous Multiscale Methods (HMM) [1], the scheme consists of a macro solver to evolve (1.2) and a micro solver for the fast dynamics in (1.1). An estimator is chosen to estimate the coefficients $\bar{a}(\cdot)$ and $\bar{\sigma}(\cdot)$ on-the-fly at each time step of the macro solver using data obtained from the fast simulations with the micro solver. Without having to resolve all the details of the fast process on the $O(\varepsilon)$ time scale, the method is able to overcome the numerical stiffness induced by the time scale separation. To fully justify this strategy, we need an accurate quantitative estimate on the validity of the effective dynamics (1.2). The convergence in probability for (1.1) to (1.2) has been proved in [11, 3] with no explicit convergence rate given. The weak convergence has been proved using the asymptotic expansion of the the backward operator [5, 7], which implies that the convergence rate is $O(\varepsilon)$. What is of further interest is the convergence in the strong sense, which provides pathwise asymptotic information for the dynamical trajectories of the systems.

The problem of the strong convergence rate for (1.1) has been studied in previous literature under the condition that the diffusion in the slow dynamics is independent of the fast variable, i.e.,

$$\sigma(x,y,\varepsilon)=\sigma(x,\varepsilon). \tag{1.4}$$

Assuming (1.4), the strong convergence rate was proved to be $O(\varepsilon^{1/6})$ in [9] and $O(\varepsilon^{1/4})$ in [4]. In this paper, we will show that the strong convergence rate is $O(\varepsilon^{1/2})$ in this case. We will also show through examples that the rate is not only sharper but also optimal. The result here is a generalization of the theorem proved in [2] when $\sigma=0$. We also discuss the fully coupled systems for which assumption (1.4) is not true and the slow diffusion does depend on the fast variables. For this situation, it can be seen through simple examples that although the weak convergence of (1.2) is still valid, the strong convergence does not hold in general. In this paper, we provide a closed form effective dynamics in terms of the slow variables, which takes the form of the slow dynamics with the fast variable being replaced by a quasi-stationary process depending on the slow variable $x$ and parameter $\varepsilon$, i.e.,

$$\dot{\bar{X}}_t^\varepsilon = a\big(\bar{X}_t^\varepsilon,\xi_t\big(\bar{X}_t^\varepsilon,\varepsilon\big),\varepsilon\big)+\sigma\big(\bar{X}_t^\varepsilon,\xi_t\big(\bar{X}_t^\varepsilon,\varepsilon\big),\varepsilon\big)\dot{W}_t, \qquad \bar{X}_0=x, \tag{1.5}$$

where the process $\xi_s(x,\varepsilon)$ has a stationary distribution of $\mu_x^\varepsilon(\cdot)$. The strong convergence rate for this case is proved to be also $O(\varepsilon^{1/2})$.

Throughout the paper, we denote $\mathcal{C}$ to be a generic constant that does not have to have the same value. In chains of inequalities, we will adopt $\mathcal{C}$, $\mathcal{C}'$, $\mathcal{C}''$, … or $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ to avoid confusion.

## 2. A simple example

In this section, we want to illustrate the main result of the paper through a simple example, for which the effective dynamics can be explicitly obtained and the strong convergence rate can be easily calculated.

**Example 1.** Let us consider the following linear equation:

$$\begin{aligned} \dot{X}_t^\varepsilon &= Y_t^\varepsilon+\dot{B}_t, & X_0&=x,\\ \dot{Y}_t^\varepsilon &= -\frac{1}{\varepsilon}Y_t^\varepsilon+\frac{1}{\sqrt{\varepsilon}}\dot{W}_t, & Y_0&=y. \end{aligned} \tag{2.1}$$

The above equation can be solved analytically such that

$$\begin{aligned} X_t^\varepsilon &= x + \int_0^t Y_s^\varepsilon ds + B_t, \\ Y_t^\varepsilon &= e^{-t/\varepsilon} y + \frac{1}{\sqrt{\varepsilon}} \int_0^t e^{-(t-s)/\varepsilon} dW_s. \end{aligned} \tag{2.2}$$

Since the fast process is an Ornstein-Uhlenbeck process which admits a unique invariant measure with mean zero, the effective dynamics prescribed by (1.2) can be simply written as

$$\dot{\bar{X}}_t = \dot{B}_t, \qquad X_0 = x, \tag{2.3}$$

or equivalently,

$$\bar{X}_t = x + B_t. \tag{2.4}$$

The strong rate for $X_t^\varepsilon$ to converge to $\bar{X}_t$ can be obtained through

$$\mathbb{E}\big|X_t^\varepsilon - \bar{X}_t\big| = \mathbb{E}\Big|\int_0^t Y_s^\varepsilon ds\Big|. \tag{2.5}$$

Notice that the process $Y_t^\varepsilon$ is a Gaussian process. Therefore its time integral, as a limit of sums of Gaussian random variables, is also Gaussian with mean

$$\mathbb{E}\int_0^t Y_s^\varepsilon ds = y \int_0^t e^{-s/\varepsilon} ds = O(\varepsilon). \tag{2.6}$$

We can calculate its variance such that

$$\begin{aligned} \mathbb{E}\Big(\int_0^t Y_s^\varepsilon ds\Big)^2 &= 2\mathbb{E}\int_0^t Y_s^\varepsilon ds \int_s^t Y_\tau^\varepsilon d\tau \\ &= 2\int_0^t ds \int_s^t d\tau \Big(\frac{e^{-(\tau-s)/\varepsilon} - e^{-(\tau+s)/\varepsilon}}{2} + e^{-(\tau+s)/\varepsilon} y^2\Big) \\ &= t\varepsilon + O(\varepsilon^2), \end{aligned} \tag{2.7}$$

which, together with (2.5), implies that

$$\mathbb{E}\big|X_t^\varepsilon - \bar{X}_t\big| = O(\sqrt{\varepsilon}). \tag{2.8}$$

The above example shows that the $O(\varepsilon^{1/2})$ convergence rate is optimal in the sense that any sharper rate can be counter-exampled by (2.1). It can also be seen from this example that it is the exponential decay of the correlation function of the fast dynamics that is guaranteeing the $O(\varepsilon^{1/2})$ standard deviation of the time average of the fast process, which leads to the $O(\varepsilon^{1/2})$ strong convergence of the effective dynamics.

## 3. The strong convergence rate

In this section, we want to prove the theorem for the strong convergence of the multiscale dynamics (1.1) to the effective dynamics (1.2). We will first make some assumptions on system (1.1) and provide the theorem for the strong convergence. Then we will elaborate on the proof.

**3.1. Assumptions and the convergence theorem.** Define $\mathbb{C}_b^\infty$ to be the space of smooth functions with bounded derivatives of any order. We assume the following conditions for system (1.1):

ASSUMPTION 3.1. *The coefficients $a(\cdot)$, $\sigma(\cdot)$, $B(\cdot)$, and $C(\cdot)$, viewed as functions of $(x,y,\varepsilon)$, are in $\mathbb{C}_b^\infty$. Moreover, $a(\cdot)$ and $\sigma(\cdot)$ are bounded.*

ASSUMPTION 3.2. *There exists a constant $\alpha>0$ such that for any $(x,y,\varepsilon)$,*

$$y^T C(x,y,\varepsilon)C^T(x,y,\varepsilon)y \geq \alpha|y|^2. \tag{3.1}$$

ASSUMPTION 3.3. *There exists a constant $\beta>0$ such that for any $(x,y_1,y_2,\varepsilon)$,*

$$\begin{aligned}&\big\langle y_1-y_2,\ B(x,y_1,\varepsilon)-B(x,y_2,\varepsilon)\big\rangle + \big\|C(x,y_1,\varepsilon)-C(x,y_2,\varepsilon)\big\|^2\\ &\leq -\beta\big|y_1-y_2\big|^2,\end{aligned} \tag{3.2}$$

*where $\|\cdot\|$ denotes the Frobenius norm.*

Suppose $X_t^\varepsilon$ and $\bar{X}_t$ are solutions to (1.1) and (1.2), respectively. We will prove the following theorem for the strong convergence rate:

THEOREM 3.4. *Suppose that Assumptions 3.1–3.3 hold and the following is true:*

$$\sigma=\sigma(x,\varepsilon). \tag{3.3}$$

*Then for any $T_0>0$, there exists a constant $\mathcal{C}>0$ independent of $\varepsilon$ such that*

$$\sup_{0\leq t\leq T_0}\mathbb{E}|X_t^\varepsilon-\bar{X}_t|^2\leq \mathcal{C}\varepsilon. \tag{3.4}$$

Condition (3.3) in Theorem 3.4 implies that the effective dynamics (1.2) takes a simpler form such that

$$\dot{\bar{X}}_t=\bar{a}\big(\bar{X}_t\big)+\sigma\big(\bar{X}_t\big)\dot{W}_t, \qquad \bar{X}_0=x, \tag{3.5}$$

in which the diffusion term is obtained simply by taking the limit of $\varepsilon\to 0$, without averaging with respect to the equilibrium of the fast dynamics:

$$\sigma(x)=\lim_{\varepsilon\to 0}\sigma(x,\varepsilon). \tag{3.6}$$

As we will see later in section 4, the condition (3.3) is necessary for the strong convergence of Principle of Averaging in the form of (3.4). In section 4 we will discuss the case of fully coupled systems when $\sigma$ depends on the fast variable $y$ such that $\sigma=\sigma(x,y,\varepsilon)$. The main purpose of Assumptions 3.2 and 3.3 is to guarantee the exponential convergence of the fast processes to the equilibrium. Based on the recent progress on the theory for stability of Markov processes [10], we believe Assumptions 3.2 and 3.3 can be relaxed. It can be seen from the examples in section 2 that Assumption 3.1 is also not necessary for Theorem 3.4. We will leave finding necessary conditions for Theorem 3.4 to future investigations.

**3.2. Proof of the strong convergence theorem.** Under Assumptions 3.2 and 3.3, it has been shown [6] that for each fixed $(x,\varepsilon)$, the following dynamics

$$\dot{Z}_t = \frac{1}{\varepsilon}B\big(x,Z_t,\varepsilon\big) + \frac{1}{\sqrt{\varepsilon}}C\big(x,Z_t,\varepsilon\big)\dot{W}_t, \qquad Z_0 = z, \tag{3.7}$$

is exponentially mixing with a unique invariant probability measure $\mu_x^\varepsilon(\cdot)$. To facilitate our proof, we define

$$\bar{a}(x,\varepsilon) = \int_{\mathbb{R}^m} a(x,y,\varepsilon)\mu_x^\varepsilon(dy), \tag{3.8}$$

and

$$\bar{\sigma}(x,\varepsilon)\bar{\sigma}^T(x,\varepsilon) = \int_{\mathbb{R}^m} \sigma(x,y,\varepsilon)\sigma^T(x,y,\varepsilon)\mu_x^\varepsilon(dy). \tag{3.9}$$

Notice that the following relations hold between the above functions and those defined by (1.3):

$$\bar{a}(x) = \lim_{\varepsilon\to 0}\bar{a}(x,\varepsilon), \tag{3.10}$$

and

$$\bar{\sigma}(x) = \lim_{\varepsilon\to 0}\bar{\sigma}(x,\varepsilon). \tag{3.11}$$

Let $Z_{x,z,t}^\varepsilon$ denote the solution of (3.7) with initial condition $z$ and parameter $(x,\varepsilon)$. We define the following function:

$$\hat{a}(x,z,t,\varepsilon) = \mathbb{E}a\big(x,Z_{x,z,t}^\varepsilon,\varepsilon\big). \tag{3.12}$$

We also define the following axillary process which is a modification of the process defined in [3]. Partitioning $[0,T_0]$ into subintervals of the same length $\Delta$, we construct for $t\in[k\Delta,(k+1)\Delta)$, $k\geq 0$, the process $(\tilde{X}_t^\varepsilon,\tilde{Y}_t^\varepsilon)$ such that

$$\begin{aligned}
\dot{\tilde{X}}_t^\varepsilon &= a\big(X_{k\Delta}^\varepsilon,\tilde{Y}_t^\varepsilon,\varepsilon\big) + \sigma\big(X_t^\varepsilon,\varepsilon\big)\dot{W}_t,\\
\dot{\tilde{Y}}_t^\varepsilon &= \frac{1}{\varepsilon}B\big(X_{k\Delta}^\varepsilon,\tilde{Y}_t^\varepsilon,\varepsilon\big) + \frac{1}{\sqrt{\varepsilon}}C\big(X_{k\Delta}^\varepsilon,\tilde{Y}_t^\varepsilon,\varepsilon\big)\dot{W}_t,
\end{aligned} \tag{3.13}$$

with the continuity condition at the left end of each subinterval

$$\tilde{X}_{(k+1)\Delta}^\varepsilon = \lim_{t\to(k+1)\Delta^-}\tilde{X}_t^\varepsilon, \qquad \tilde{Y}_{(k+1)\Delta}^\varepsilon = \lim_{t\to(k+1)\Delta^-}\tilde{Y}_t^\varepsilon, \tag{3.14}$$

and also the initial condition

$$\tilde{X}_0^\varepsilon = x, \qquad \tilde{Y}_0^\varepsilon = y. \tag{3.15}$$

Denote $\lfloor x\rfloor$ to be the largest integer less than or equal to $x$; we can also write (3.13)–(3.15) in the integral form

$$\begin{aligned}
\tilde{X}_t^\varepsilon &= x + \int_0^t a\Big(X_{\lfloor s/\Delta\rfloor\Delta}^\varepsilon,\tilde{Y}_s^\varepsilon,\varepsilon\Big)ds + \int_0^t \sigma\Big(X_s^\varepsilon,\varepsilon\Big)dW_s,\\
\tilde{Y}_t^\varepsilon &= y + \frac{1}{\varepsilon}\int_0^t B\Big(X_{\lfloor s/\Delta\rfloor\Delta}^\varepsilon,\tilde{Y}_s^\varepsilon,\varepsilon\Big)ds + \frac{1}{\sqrt{\varepsilon}}\int_0^t C\Big(X_{\lfloor s/\Delta\rfloor\Delta}^\varepsilon,\tilde{Y}_s^\varepsilon,\varepsilon\Big)dW_s.
\end{aligned} \tag{3.16}$$

By Assumption 3.1 on the smoothness of the coefficients $a$ and $\sigma$, $\mathbb{E}\left|X_t^\varepsilon\right|^2$ is bounded over finite time intervals. Here we want to show the stability of the fast processes implied by Assumption 3.3 on the dissipative structure of the fast processes.

LEMMA 3.5. *For any $T_0>0$, there exists a constant $\mathcal{C}$ independent of $(\varepsilon,\Delta)$ such that*

$$\mathbb{E}\left|Y_t^\varepsilon\right|^2,\ \mathbb{E}\left|\tilde{Y}_t^\varepsilon\right|^2\le\mathcal{C}. \tag{3.17}$$

*Proof.* Fixing $y_1=y$ and $y_2=0$ in Assumption 3.3 will give us

$$\left\langle y,B(x,y,\varepsilon)-B(x,0,\varepsilon)\right\rangle+\left\|C(x,y,\varepsilon)-C(x,0,\varepsilon)\right\|^2\le-\beta\left|y\right|^2. \tag{3.18}$$

By Assumption 3.1, we have for any $\gamma>0$,

$$\begin{aligned}\left\|C(x,y,\varepsilon)\right\|^2&\le(1+\gamma)\left\|C(x,y,\varepsilon)-C(x,0,\varepsilon)\right\|^2+(1+1/\gamma)\|C(x,0,\varepsilon)\|^2\\&\le\left\|C(x,y,\varepsilon)-C(x,0,\varepsilon)\right\|^2+\mathcal{C}_1\gamma|y|^2+\mathcal{C}_2(1+1/\gamma)\left(|x|^2+\varepsilon^2+1\right),\end{aligned} \tag{3.19}$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are the Lipschitz and linear growth constants of function $C(x,y,\varepsilon)$. If we choose an appropriate value for $\gamma$ such that $\mathcal{C}_1\gamma\le\beta/4$, by (3.18), (3.19), and Assumption 3.1, we can obtain

$$\begin{aligned}\left\langle y,B(x,y,\varepsilon)\right\rangle+\left\|C(x,y,\varepsilon)\right\|^2&\le\left\langle y,B(x,y,\varepsilon)-B(x,0,\varepsilon)\right\rangle+\left\langle y,B(x,0,\varepsilon)\right\rangle\\&\quad+\left\|C(x,y,\varepsilon)-C(x,0,\varepsilon)\right\|^2+\frac{\beta}{4}|y|^2+\mathcal{C}'\left(|x^2|+\varepsilon^2+1\right)\\&\le-\frac{\beta}{2}|y|^2+\mathcal{C}''\left(|x|^2+\varepsilon^2+1\right).\end{aligned} \tag{3.20}$$

The Ito formula then suggests that

$$\begin{aligned}d\mathbb{E}\left|Y_t^\varepsilon\right|^2&=\frac{2}{\varepsilon}\mathbb{E}\Big\langle Y_t^\varepsilon,\ B\left(X_t^\varepsilon,Y_t^\varepsilon\right)\Big\rangle dt+\frac{1}{\varepsilon}\mathbb{E}\Big\|C\left(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon\right)\Big\|^2dt\\&\le-\frac{\beta}{\varepsilon}\mathbb{E}\left|Y_t^\varepsilon\right|^2+\frac{\mathcal{C}}{\varepsilon}\mathbb{E}\Big(|X_t^\varepsilon|^2+\varepsilon^2+1\Big),\end{aligned} \tag{3.21}$$

which, by the Gronwall inequality, implies the boundedness of $\mathbb{E}\left|Y_t^\varepsilon\right|^2$. Repeating the same argument, we can also obtain the boundedness of $\mathbb{E}\left|\tilde{Y}_t^\varepsilon\right|^2$. ∎

The following Lemma describes how $\tilde{Y}_t^\varepsilon$ deviates from $Y_t^\varepsilon$.

LEMMA 3.6. *For any $T_0>0$, there exists a constant $\mathcal{C}>0$ independent of $(\varepsilon,\Delta)$ such that*

$$\sup_{0\le t\le T_0}\mathbb{E}\left|Y_t^\varepsilon-\tilde{Y}_t^\varepsilon\right|^2\le\mathcal{C}\Delta. \tag{3.22}$$

*Proof.* For each $k\ge0$ and $t\in[k\Delta,(k+1)\Delta)$, direct computation with the Ito formula gives that

$$\begin{aligned}d\mathbb{E}\left|Y_t^\varepsilon-\tilde{Y}_t^\varepsilon\right|^2=&\frac{2}{\varepsilon}\mathbb{E}\left(Y_t^\varepsilon-\tilde{Y}_t^\varepsilon\right)\cdot\Big(B\left(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon\right)-B\left(X_{k\Delta}^\varepsilon,\tilde{Y}_t^\varepsilon,\varepsilon\right)\Big)dt\\&+\frac{1}{\varepsilon}\mathbb{E}\Big\|C\left(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon\right)-C\left(X_{k\Delta}^\varepsilon,\tilde{Y}_t^\varepsilon,\varepsilon\right)\Big\|^2dt.\end{aligned} \tag{3.23}$$

By Assumptions 3.1 and 3.3, we have

$$\begin{aligned}
&\big(Y_t^\varepsilon-\tilde Y_t^\varepsilon\big)\cdot\Big(B\big(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon\big)-B\big(X_{k\Delta}^\varepsilon,\tilde Y_t^\varepsilon,\varepsilon\big)\Big)+\frac{1}{2}\Big\|C\big(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon\big)-C\big(X_{k\Delta}^\varepsilon,\tilde Y_t^\varepsilon,\varepsilon\big)\Big\|^2\\
&\le\big(Y_t^\varepsilon-\tilde Y_t^\varepsilon\big)\cdot\Big(B\big(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon\big)-B\big(X_t^\varepsilon,\tilde Y_t^\varepsilon,\varepsilon\big)\Big)\\
&\qquad+\big(Y_t^\varepsilon-\tilde Y_t^\varepsilon\big)\cdot\Big(B\big(X_t^\varepsilon,\tilde Y_t^\varepsilon,\varepsilon\big)-B\big(X_{k\Delta}^\varepsilon,\tilde Y_t^\varepsilon,\varepsilon\big)\Big)\\
&\qquad+\Big\|C\big(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon\big)-C\big(X_t^\varepsilon,\tilde Y_t^\varepsilon,\varepsilon\big)\Big\|^2+\Big\|C\big(X_t^\varepsilon,\tilde Y_t^\varepsilon,\varepsilon\big)-C\big(X_{k\Delta}^\varepsilon,\tilde Y_t^\varepsilon,\varepsilon\big)\Big\|^2\\
&\le-\beta\big|Y_t^\varepsilon-\tilde Y_t^\varepsilon\big|^2+\mathcal{C}\Big(\big|Y_t^\varepsilon-\tilde Y_t^\varepsilon\big|\big|X_t^\varepsilon-X_{k\Delta}^\varepsilon\big|+\big|X_t^\varepsilon-X_{k\Delta}^\varepsilon\big|^2\Big).
\end{aligned}\tag{3.24}$$

Note that since $\beta>0$, we have

$$\mathcal{C}\big|Y_t^\varepsilon-\tilde Y_t^\varepsilon\big|\big|X_t^\varepsilon-X_{k\Delta}^\varepsilon\big|\le\frac{1}{2}\beta\big|Y_t^\varepsilon-\tilde Y_t^\varepsilon\big|^2+\frac{\mathcal{C}^2}{2\beta}\big|X_t^\varepsilon-X_{k\Delta}^\varepsilon\big|^2,\tag{3.25}$$

which easily leads to

$$\begin{aligned}
&\big(Y_t^\varepsilon-\tilde Y_t^\varepsilon\big)\cdot\Big(B\big(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon\big)-B\big(X_{k\Delta}^\varepsilon,\tilde Y_t^\varepsilon,\varepsilon\big)\Big)\\
&\qquad+\frac{1}{2}\Big\|C\big(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon\big)-C\big(X_{k\Delta}^\varepsilon,\tilde Y_t^\varepsilon,\varepsilon\big)\Big\|^2\\
&\le-\frac{1}{2}\beta\big|Y_t^\varepsilon-\tilde Y_t^\varepsilon\big|^2+\mathcal{C}\big|X_t^\varepsilon-X_{k\Delta}^\varepsilon\big|^2.
\end{aligned}\tag{3.26}$$

By the boundedness of $a(\cdot)$ and $\sigma(\cdot)$ and the quadratic variation of the Brownian motion we have for $t\in[k\Delta,(k+1)\Delta)$,

$$\mathbb{E}\big|X_t^\varepsilon-X_{k\Delta}^\varepsilon\big|^2\le\mathcal{C}\Delta.\tag{3.27}$$

Combining (3.23), (3.26), and (3.27), it follows that

$$d\mathbb{E}\big|Y_t^\varepsilon-\tilde Y_t^\varepsilon\big|^2\le-\frac{\beta}{\varepsilon}\mathbb{E}\big|Y_t^\varepsilon-\tilde Y_t^\varepsilon\big|^2dt+\mathcal{C}\frac{\Delta}{\varepsilon}dt.\tag{3.28}$$

The Gronwall inequality implies that

$$\mathbb{E}\big|Y_t^\varepsilon-\tilde Y_t^\varepsilon\big|^2\le e^{-\beta\frac{(t-k\Delta)}{\varepsilon}}\mathbb{E}\big|Y_{k\Delta}^\varepsilon-\tilde Y_{k\Delta}^\varepsilon\big|^2+\mathcal{C}\big(1-e^{-\beta\frac{(t-k\Delta)}{\varepsilon}}\big)\Delta.\tag{3.29}$$

By the continuity condition (3.14) we can take $t=(k+1)\Delta$ in the above inequality, which gives

$$\mathbb{E}\big|Y_{(k+1)\Delta}^\varepsilon-\tilde Y_{(k+1)\Delta}^\varepsilon\big|^2\le e^{-\beta\frac{\Delta}{\varepsilon}}\mathbb{E}\big|Y_{k\Delta}^\varepsilon-\tilde Y_{k\Delta}^\varepsilon\big|^2+\mathcal{C}\big(1-e^{-\frac{\beta}{\varepsilon}\Delta}\big)\Delta.\tag{3.30}$$

Applying the above inequality recursively for reducing value of $k$ until $k=0$, we can obtain by the initial condition (3.15) that

$$\mathbb{E}\big|Y_{(k+1)\Delta}^\varepsilon-\tilde Y_{(k+1)\Delta}^\varepsilon\big|^2\le\mathcal{C}\big(1-e^{-\frac{\beta}{\varepsilon}\Delta}\big)\Delta\sum_{0\le\ell\le k}e^{-\ell\beta\frac{\Delta}{\varepsilon}}\le\mathcal{C}\Delta,\tag{3.31}$$

which, together with (3.29), gives (3.22). ∎

Lemma 3.6 easily gives the following asymptotic behavior of $\tilde X_t^\varepsilon$.

PROPOSITION 3.7. *For any $T_0>0$, there exists a constant $\mathcal{C}>0$ independent of $(\varepsilon,\Delta)$ such that*

$$\sup_{0\le t\le T_0}\mathbb{E}\big|X_t^\varepsilon-\tilde{X}_t^\varepsilon\big|^2\le\mathcal{C}\Delta. \tag{3.32}$$

*Proof.* By the smoothness of the coefficients and Lemma 3.6, we can write

$$\begin{aligned}\mathbb{E}\big|X_t^\varepsilon-\tilde{X}_t^\varepsilon\big|^2&=\mathbb{E}\Big|\int_0^t\Big(a\Big(X_s^\varepsilon,Y_s^\varepsilon,\varepsilon\Big)-a\Big(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta},\tilde{Y}_s^\varepsilon,\varepsilon\Big)ds\Big|^2\\&\le\mathcal{C}\mathbb{E}\int_0^t\Big(\big|X_s^\varepsilon-X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta}\big|^2+\big|Y_s^\varepsilon-\tilde{Y}_s^\varepsilon\big|^2\Big)ds\\&\le\mathcal{C}\Delta.\end{aligned} \tag{3.33}$$

∎

Now we want to give an estimate for the expectation $\mathbb{E}\big|\tilde{X}_t^\varepsilon-\bar{X}_t\big|^2$.

PROPOSITION 3.8. *For any $T_0>0$, there exists a constant $\mathcal{C}>0$ independent of $(\varepsilon,\Delta)$ such that*

$$\sup_{0\le t\le T_0}\mathbb{E}|\tilde{X}_t^\varepsilon-\bar{X}_t|^2\le\mathcal{C}\big(\Delta+\varepsilon\big). \tag{3.34}$$

*Proof.* First of all, we notice that

$$\begin{aligned}\mathbb{E}|\tilde{X}_t^\varepsilon-\bar{X}_t|^2\le 2\mathbb{E}&\left(\int_0^t\Big(a\Big(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta},\tilde{Y}_s^\varepsilon,\varepsilon\Big)-\bar{a}\big(\bar{X}_s\big)\Big)ds\right)^2\\&+2\mathbb{E}\left(\int_0^t\Big(\sigma\big(X_s^\varepsilon,\varepsilon\big)-\sigma\big(\bar{X}_s^\varepsilon\big)\Big)dW_s\right)^2.\end{aligned} \tag{3.35}$$

For the second term on the right hand side of the above inequality, we have by Ito Isometry

$$\mathbb{E}\Big(\int_0^t\Big(\sigma\big(X_s^\varepsilon,\varepsilon\big)-\sigma\big(\bar{X}_s^\varepsilon\big)\Big)dW_s\Big)^2=\int_0^t\mathbb{E}\Big(\sigma\big(X_s^\varepsilon,\varepsilon\big)-\sigma\big(\bar{X}_s^\varepsilon\big)\Big)^2ds. \tag{3.36}$$

Proposition 3.7 and the smoothness of the coefficients gives the following estimate:

$$\begin{aligned}\mathbb{E}\Big(\sigma\big(X_s^\varepsilon,\varepsilon\big)-\sigma\big(\bar{X}_s^\varepsilon\big)\Big)^2&\le\mathcal{C}\Big(\mathbb{E}\big|X_s^\varepsilon-\bar{X}_s\big|^2+\varepsilon^2\Big)\\&\le\mathcal{C}'\left(\mathbb{E}\big|\tilde{X}_s^\varepsilon-\bar{X}_s\big|^2+\Delta+\varepsilon^2\right).\end{aligned} \tag{3.37}$$

Therefore we have

$$\mathbb{E}\Big(\int_0^t\Big(\sigma\big(X_s^\varepsilon,\varepsilon\big)-\sigma\big(\bar{X}_s^\varepsilon\big)\Big)dW_s\Big)^2\le\mathcal{C}\left(\int_0^t\mathbb{E}\big|\tilde{X}_s^\varepsilon-\bar{X}_s\big|^2ds+\Delta+\varepsilon^2\right). \tag{3.38}$$

Now we want to give an estimate for the first term on the right hand side of (3.35). Notice that

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^t \left(a\left(X^{\varepsilon}_{\lfloor s/\Delta\rfloor \Delta},\tilde{Y}^{\varepsilon}_s,\varepsilon\right)-\bar{a}\left(\bar{X}_s\right)\right)ds\right)^2\\
&\leq 3\mathbb{E}\left(\int_0^t \left(a\left(X^{\varepsilon}_{\lfloor s/\Delta\rfloor \Delta},\tilde{Y}^{\varepsilon}_s,\varepsilon\right)-\bar{a}\left(X^{\varepsilon}_{\lfloor s/\Delta\rfloor \Delta}\right)\right)ds\right)^2\\
&\qquad +3\mathbb{E}\left(\int_0^t \left(\bar{a}\left(X^{\varepsilon}_{\lfloor s/\Delta\rfloor \Delta}\right)-\bar{a}\left(X^{\varepsilon}_s\right)\right)ds\right)^2+3\mathbb{E}\left(\int_0^t \left(\bar{a}\left(X^{\varepsilon}_s\right)-\bar{a}\left(\bar{X}_s\right)\right)ds\right)^2.
\end{aligned}
\tag{3.39}
$$

By the smoothness of $\bar{a}(x,\varepsilon)$ proved in the Appendix, we have

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^t \left(a\left(X^{\varepsilon}_{\lfloor s/\Delta\rfloor \Delta},\tilde{Y}^{\varepsilon}_s,\varepsilon\right)-\bar{a}\left(\bar{X}_s\right)\right)ds\right)^2\\
&\leq 3\mathbb{E}\left(\int_0^t \left(a\left(X^{\varepsilon}_{\lfloor s/\Delta\rfloor \Delta},\tilde{Y}^{\varepsilon}_s,\varepsilon\right)-\bar{a}\left(X^{\varepsilon}_{\lfloor s/\Delta\rfloor \Delta}\right)\right)ds\right)^2+\mathcal{C}\left(\Delta+\mathbb{E}\int_0^t \left|X^{\varepsilon}_s-\bar{X}_s\right|^2 ds\right).
\end{aligned}
\tag{3.40}
$$

We can evaluate the above double integral in the above inequality as

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^t \left(a\left(X^{\varepsilon}_{\lfloor s/\Delta\rfloor \Delta},\tilde{Y}^{\varepsilon}_s,\varepsilon\right)-\bar{a}\left(\bar{X}_{\lfloor s/\Delta\rfloor \Delta}\right)\right)ds\right)^2\\
&\leq \mathbb{E}\sum_{0\leq k\leq \lfloor t/\Delta\rfloor}\left(\int_{k\Delta}^{((k+1)\Delta)\wedge t} \left(a\left(X^{\varepsilon}_{k\Delta},\tilde{Y}^{\varepsilon}_s,\varepsilon\right)-\bar{a}\left(X^{\varepsilon}_{k\Delta}\right)\right)ds\right)^2\\
&\quad +2\mathbb{E}\sum_{0\leq i<j\leq \lfloor t/\Delta\rfloor}\left(\int_{i\Delta}^{((i+1)\Delta)\wedge t} \left(a\left(X^{\varepsilon}_{i\Delta},\tilde{Y}^{\varepsilon}_s,\varepsilon\right)-\bar{a}\left(X^{\varepsilon}_{i\Delta}\right)\right)ds\right)\\
&\qquad\qquad \cdot\left(\int_{j\Delta}^{((j+1)\Delta)\wedge t} \left(a\left(X^{\varepsilon}_{j\Delta},\tilde{Y}^{\varepsilon}_\tau,\varepsilon\right)-\bar{a}\left(X^{\varepsilon}_{j\Delta}\right)\right)d\tau\right)\\
&\stackrel{def}{=} A_1+2A_2.
\end{aligned}
\tag{3.41}
$$

For $A_1$, we have the following estimate:

$$
\begin{aligned}
&\mathbb{E}\left(\int_{k\Delta}^{((k+1)\Delta)\wedge t} \left(a\left(X^{\varepsilon}_{k\Delta},\tilde{Y}^{\varepsilon}_s,\varepsilon\right)-\bar{a}\left(X^{\varepsilon}_{k\Delta}\right)\right)ds\right)^2\\
&\leq 2\int_{k\Delta}^{((k+1)\Delta)\wedge t}\mathbb{E}\left(a\left(X^{\varepsilon}_{k\Delta},\tilde{Y}^{\varepsilon}_s,\varepsilon\right)-\bar{a}\left(X^{\varepsilon}_{k\Delta}\right)\right)ds\\
&\qquad\qquad \cdot\int_s^{(k+1)\Delta}\mathbb{E}_s\left(a\left(X^{\varepsilon}_{k\Delta},\tilde{Y}^{\varepsilon}_\tau,\varepsilon\right)-\bar{a}\left(X^{\varepsilon}_{k\Delta}\right)\right)d\tau,
\end{aligned}
\tag{3.42}
$$

where $\mathbb{E}_s$ denotes the conditional probability for information up to time $s$. By the smoothness of $\bar{a}(x,\varepsilon)$, we have

$$\begin{aligned}
&\left|\mathbb{E}_s\Big(a\big(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_s,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{k\Delta}\big)\Big)\right|\\
\le&\left|\mathbb{E}_s\Big(a\big(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_s,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{k\Delta},\varepsilon\big)\Big)\right|+\left|\mathbb{E}_s\Big(\bar{a}\big(X^\varepsilon_{k\Delta},\varepsilon\big)-\bar{a}\big(X^\varepsilon_{k\Delta}\big)\Big)\right|\\
\le&\left|\mathbb{E}_s\Big(a\big(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_s,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{k\Delta},\varepsilon\big)\Big)\right|+\mathcal{C}\varepsilon.
\end{aligned}\tag{3.43}$$

By the exponential mixing (A.17) given in the Appendix and Lemma 3.5 for the boundedness of $\mathbb{E}\big|Y^\varepsilon_t\big|$, we can write

$$\begin{aligned}
\mathbb{E}\left|\mathbb{E}_s\Big(a\big(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_\tau,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{k\Delta},\varepsilon\big)\Big)\right|&\le\mathcal{C}\mathbb{E}\Big(\big|X^\varepsilon_s\big|+\big|Y^\varepsilon_s\big|+\varepsilon+1\Big)e^{-\beta\frac{(\tau-s)}{\varepsilon}}\\
&\le\mathcal{C}'e^{-\beta\frac{(\tau-s)}{\varepsilon}}.
\end{aligned}\tag{3.44}$$

Therefore, by the Assumption that $a$ is bounded, we have

$$\begin{aligned}
A_1\le&\,\mathcal{C}\sum_{0\le k\le\lfloor t/\Delta\rfloor}\int_{k\Delta}^{(k+1)\Delta\wedge t}ds\int_s^{(k+1)\Delta\wedge t}d\tau\Big(e^{-\beta\frac{(\tau-s)}{\varepsilon}}+\varepsilon\Big)\\
\le&\,\mathcal{C}'\varepsilon.
\end{aligned}\tag{3.45}$$

To estimate $A_2$, we define the auxiliary process $Z^\varepsilon_{i,\tau}$ for $i\Delta\le\tau$ such that it satisfies (3.7) with parameter $x=X^\varepsilon_{i\Delta}$ and initial condition $\tilde{Y}^\varepsilon_{i\Delta}$, i.e.,

$$\begin{aligned}
&\dot{Z}^\varepsilon_{i,\tau}=\frac{1}{\varepsilon}B\big(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon\big)+\frac{1}{\sqrt{\varepsilon}}C\big(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon\big)\dot{W}_\tau,\\
&Z_{i,i\Delta}=\tilde{Y}^\varepsilon_{i\Delta}.
\end{aligned}\tag{3.46}$$

Notice that by the above definition, we have

$$Z^\varepsilon_{k,\tau}=\tilde{Y}^\varepsilon_\tau,\quad\text{when}\quad\tau\in[k\Delta,(k+1)\Delta),\tag{3.47}$$

and continuity implies that

$$Z^\varepsilon_{(k+1),(k+1)\Delta}=Z^\varepsilon_{k,(k+1)\Delta}=\tilde{Y}^\varepsilon_{(k+1)\Delta}.\tag{3.48}$$

Using the boundedness of $\bar{a}(x,\varepsilon)$, we obtain

$$\begin{aligned}
&\left|\mathbb{E}\Big(a\big(X^\varepsilon_{i\Delta},\tilde{Y}^\varepsilon_s,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{i\Delta}\big)\Big)\cdot\Big(a\big(X^\varepsilon_{j\Delta},\tilde{Y}^\varepsilon_\tau,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{j\Delta}\big)\Big)\right|\\
=&\left|\mathbb{E}\Big(a\big(X^\varepsilon_{i\Delta},\tilde{Y}^\varepsilon_s,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{i\Delta}\big)\Big)\cdot\mathbb{E}_{i\Delta}\Big(a\big(X^\varepsilon_{j\Delta},\tilde{Y}^\varepsilon_\tau,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{j\Delta}\big)\Big)\right|\\
\le&\,\mathcal{C}\mathbb{E}\left|\mathbb{E}_{i\Delta}\Big(a\big(X^\varepsilon_{j\Delta},\tilde{Y}^\varepsilon_\tau,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{j\Delta}\big)\Big)\right|\\
\le&\,\mathcal{C}\mathbb{E}\left|\mathbb{E}_{i\Delta}\Big\{\Big(a\big(X^\varepsilon_{j\Delta},\tilde{Y}^\varepsilon_\tau,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{j\Delta}\big)\Big)-\Big(a\big(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon\big)-\bar{a}\big(X^\varepsilon_{i\Delta}\big)\Big)\Big\}\right|\\
&+\mathcal{C}\mathbb{E}\left|\mathbb{E}_{i\Delta}\Big(a\big(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon\big)-\bar{a}\big(X^\varepsilon_{i\Delta}\big)\Big)\right|.
\end{aligned}\tag{3.49}$$

By the exponential mixing (A.17) and the smoothness of $\bar{a}(x,\varepsilon)$, we have

$$\mathbb{E}\left|\mathbb{E}_{i\Delta}\left(a\left(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon\right)-\bar{a}\left(X^\varepsilon_{i\Delta}\right)\right)\right|\le\mathcal{C}\left(e^{-\beta\frac{(\tau-i\Delta)}{\varepsilon}}+\varepsilon\right). \tag{3.50}$$

The smoothness of $\bar{a}(x,\varepsilon)$ suggests that

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{E}_{i\Delta}\left\{\left(a\left(X^\varepsilon_{j\Delta},\tilde{Y}^\varepsilon_{\tau},\varepsilon\right)-\bar{a}\left(X^\varepsilon_{j\Delta}\right)\right)-\left(a\left(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon\right)-\bar{a}\left(X^\varepsilon_{i\Delta}\right)\right)\right\}\right|\\
\le&\mathbb{E}\left|\mathbb{E}_{i\Delta}\left\{\left(a\left(X^\varepsilon_{j\Delta},\tilde{Y}^\varepsilon_{\tau},\varepsilon\right)-\bar{a}\left(X^\varepsilon_{j\Delta},\varepsilon\right)\right)-\left(a\left(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon\right)-\bar{a}\left(X^\varepsilon_{i\Delta},\varepsilon\right)\right)\right\}\right|+\mathcal{C}\varepsilon\\
\le&\sum_{k=i}^{j-1}\mathbb{E}\Big|\mathbb{E}_{k\Delta}\Big\{\left(a\left(X^\varepsilon_{(k+1)\Delta},Z^\varepsilon_{k+1,\tau},\varepsilon\right)-\bar{a}\left(X^\varepsilon_{(k+1)\Delta},\varepsilon\right)\right)\\
&\qquad-\left(a\left(X^\varepsilon_{k\Delta},Z^\varepsilon_{k,\tau},\varepsilon\right)-\bar{a}\left(X^\varepsilon_{k\Delta},\varepsilon\right)\right)\Big\}\Big|+\mathcal{C}\varepsilon.
\end{aligned}\tag{3.51}$$

From the definition of $\hat{a}(x,z,t,\varepsilon)$ by (3.12), we can see that for $i\le k$ and $(k+1)\Delta\le t$,

$$\begin{aligned}
\mathbb{E}_{i\Delta}a\left(X^\varepsilon_{k\Delta},Z^\varepsilon_{k,t},\varepsilon\right)&=\mathbb{E}_{i\Delta}\hat{a}\left(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{k\Delta},t-k\Delta,\varepsilon\right)\\
&=\mathbb{E}_{i\Delta}\hat{a}\left(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{(k+1)\Delta},t-(k+1)\Delta,\varepsilon\right).
\end{aligned}\tag{3.52}$$

Let $\tilde{a}(x,z,t,\varepsilon)=\hat{a}(x,z,t,\varepsilon)-\bar{a}(x,\varepsilon)$. By the smoothness of $\hat{a}(x,z,t,\varepsilon)$ given in the Appendix, we can perform the following Taylor expansion:

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{E}_{k\Delta}\left\{\left(a\left(X^\varepsilon_{(k+1)\Delta},Z^\varepsilon_{k+1,\tau},\varepsilon\right)-\bar{a}\left(X^\varepsilon_{(k+1)\Delta},\varepsilon\right)\right)-\left(a\left(X^\varepsilon_{k\Delta},Z^\varepsilon_{k,\tau},\varepsilon\right)-\bar{a}\left(X^\varepsilon_{k\Delta},\varepsilon\right)\right)\right\}\right|\\
=&\mathbb{E}\left|\mathbb{E}_{k\Delta}\left\{\tilde{a}\left(X^\varepsilon_{(k+1)\Delta},\tilde{Y}^\varepsilon_{(k+1)\Delta},\tau-(k+1)\Delta,\varepsilon\right)-\tilde{a}\left(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{(k+1)\Delta},\tau-(k+1)\Delta,\varepsilon\right)\right\}\right|\\
\le&\mathbb{E}\left|\mathbb{E}_{k\Delta}\sum_{|I|=1}^{3}\nabla^I_x\tilde{a}\left(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{(k+1)\Delta},\tau-(k+1)\Delta,\varepsilon\right)\cdot\left(X^\varepsilon_{(k+1)\Delta}-X^\varepsilon_{k\Delta}\right)^I\right|+\mathcal{C}\Delta^2.
\end{aligned}\tag{3.53}$$

We can further have for any multi-index $I$,

$$\begin{aligned}
&\mathbb{E}\left|\mathbb{E}_{k\Delta}\nabla^I_x\tilde{a}\left(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{(k+1)\Delta},\tau-(k+1)\Delta,\varepsilon\right)\left(X^\varepsilon_{(k+1)\Delta}-X^\varepsilon_{k\Delta}\right)^I\right|\\
\le&\mathbb{E}\left|\mathbb{E}_{k\Delta}\nabla^I_x\tilde{a}\left(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{k\Delta},\tau-(k+1)\Delta,\varepsilon\right)\cdot\left(X^\varepsilon_{(k+1)\Delta}-X^\varepsilon_{k\Delta}\right)^I\right|\\
&+\mathbb{E}\Big|\mathbb{E}_{k\Delta}\nabla^I_x\Big\{\hat{a}\left(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{(k+1)\Delta},\tau-(k+1)\Delta,\varepsilon\right)\\
&\qquad-\hat{a}\left(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{k\Delta},\tau-(k+1)\Delta,\varepsilon\right)\Big\}\cdot\left(X^\varepsilon_{(k+1)\Delta}-X^\varepsilon_{k\Delta}\right)^I\Big|.
\end{aligned}\tag{3.54}$$

Using independent increments and exponential mixing for the derivatives of $\bar{a}(x,\varepsilon)$ given by (A.19) in the Appendix and the fact that for any diffusion process $x_t$,

$$\mathbb{E}(dx_t)^I = O\Big(dt^{\lfloor \frac{|I|+1}{2} \rfloor}\Big), \tag{3.55}$$

we have

$$\begin{aligned}
&\mathbb{E}\bigg|\mathbb{E}_{k\Delta}\nabla_x^I \tilde{a}\Big(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{k\Delta},\tau-(k+1)\Delta,\varepsilon\Big)\cdot\Big(X^\varepsilon_{(k+1)\Delta}-X^\varepsilon_{k\Delta}\Big)^I\bigg| \\
&\le \Delta^{\lfloor \frac{|I|+1}{2} \rfloor}\mathbb{E}\bigg|\mathbb{E}_{k\Delta}\nabla_x^I\Big\{\hat{a}\Big(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{k\Delta},\tau-(k+1)\Delta,\varepsilon\Big)-\bar{a}\Big(X^\varepsilon_{k\Delta},\varepsilon\Big)\Big\}\bigg| \\
&\le \mathcal{C}\Delta e^{-\beta\frac{(\tau-(k+1)\Delta)}{\varepsilon}},
\end{aligned} \tag{3.56}$$

and

$$\begin{aligned}
&\mathbb{E}\bigg|\mathbb{E}_{k\Delta}\nabla_x^I\Big\{\hat{a}\Big(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{(k+1)\Delta},\tau-(k+1)\Delta,\varepsilon\Big) \\
&\qquad -\hat{a}\Big(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{k\Delta},\tau-(k+1)\Delta,\varepsilon\Big)\Big\}\cdot\Big(X^\varepsilon_{(k+1)\Delta}-X^\varepsilon_{k\Delta}\Big)^I\bigg| \\
&\le \Delta^{\lfloor \frac{|I|+2}{2} \rfloor}\mathbb{E}\bigg|\mathbb{E}_{k\Delta}\nabla_x^I\sum_{|J|=1}\nabla_z^J\hat{a}\Big(X^\varepsilon_{k\Delta},\tilde{Y}^\varepsilon_{k\Delta},\tau-(k+1)\Delta,\varepsilon\Big)\bigg|+\mathcal{C}\Delta^2 \\
&\le \mathcal{C}'\Big((\Delta e^{-\beta\frac{(\tau-(k+1)\Delta)}{\varepsilon}}+\Delta^2\Big).
\end{aligned} \tag{3.57}$$

Combining (3.53), (3.54), (3.56), and (3.57), we have

$$\begin{aligned}
&\mathbb{E}\bigg|\mathbb{E}_{k\Delta}\Big\{\Big(a\big(X^\varepsilon_{(k+1)\Delta},Z^\varepsilon_{k+1,\tau},\varepsilon\big)-\bar{a}\big(X^\varepsilon_{(k+1)\Delta},\varepsilon\big)\Big)-\Big(a\big(X^\varepsilon_{k\Delta},Z^\varepsilon_{k,\tau},\varepsilon\big)-\bar{a}\big(X^\varepsilon_{k\Delta},\varepsilon\big)\Big)\Big\}\bigg| \\
&\le \mathcal{C}\Big(\Delta^2+\Delta e^{-\beta\frac{(\tau-(k+1)\Delta)}{\varepsilon}}\Big).
\end{aligned} \tag{3.58}$$

Substituting the above inequality into (3.51) gives

$$\begin{aligned}
&\mathbb{E}\bigg|\mathbb{E}_{i\Delta}\Big\{\Big(a\big(X^\varepsilon_{j\Delta},\tilde{Y}^\varepsilon_\tau,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{j\Delta}\big)\Big)-\Big(a\big(X^\varepsilon_{i\Delta},Z^\varepsilon_{i,\tau},\varepsilon\big)-\bar{a}\big(X^\varepsilon_{i\Delta}\big)\Big)\Big\}\bigg| \\
&\le \mathcal{C}\left(\sum_{k=i}^{j-1}\Big(\Delta^2+\Delta e^{-\beta\frac{(\tau-(k+1)\Delta)}{\varepsilon}}\Big)+\varepsilon\right) \\
&\le \mathcal{C}\left((j-i)\Delta^2+\Delta\frac{e^{-\beta\frac{(\tau-j\Delta)}{\varepsilon}}}{1-e^{-\beta\frac{\Delta}{\varepsilon}}}+\varepsilon\right).
\end{aligned} \tag{3.59}$$

Using (3.49), (3.50), and (3.59), we have

$$\begin{aligned}
&\bigg|\mathbb{E}\Big(a\big(X^\varepsilon_{i\Delta},\tilde{Y}^\varepsilon_s,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{i\Delta}\big)\Big)\cdot\Big(a\big(X^\varepsilon_{j\Delta},\tilde{Y}^\varepsilon_\tau,\varepsilon\big)-\bar{a}\big(X^\varepsilon_{j\Delta}\big)\Big)\bigg| \\
&\le \mathcal{C}\left((j-i)\Delta^2+\Delta\frac{e^{-\beta\frac{(\tau-j\Delta)}{\varepsilon}}}{1-e^{-\beta\frac{\Delta}{\varepsilon}}}+e^{-\frac{\beta}{\varepsilon}(\tau-i\Delta)}+\varepsilon\right).
\end{aligned} \tag{3.60}$$

Therefore, by the boundedness of $a$, we have

$$
\begin{aligned}
A_2 &\le \mathcal{C} \sum_{0\le i<j\le \lfloor t/\Delta \rfloor} \int_{i\Delta}^{(i+1)\Delta} ds \int_{j\Delta}^{(j+1)\Delta} d\tau \left( (j-i)\Delta^2 + \Delta \frac{e^{-\beta\frac{(\tau - j\Delta)}{\varepsilon}}}{1-e^{-\beta\frac{\Delta}{\varepsilon}}} + e^{-\frac{\beta}{\varepsilon}(\tau - i\Delta)} + \varepsilon \right) \\
&\le \mathcal{C} \sum_{0\le i<j\le \lfloor t/\Delta \rfloor} \left( (j-i)\Delta^4 + \varepsilon\Delta^2 + e^{-\beta\frac{(j-i)\Delta}{\varepsilon}} \left(1-e^{-\beta\frac{\Delta}{\varepsilon}}\right)\varepsilon\Delta \right) \\
&\le \mathcal{C}\Big(\Delta + \varepsilon\Big).
\end{aligned} \tag{3.61}
$$

Combining (3.38), (3.40), (3.45), and (3.61), we have

$$
\mathbb{E}\left|\tilde{X}_t^\varepsilon - \bar{X}_t\right|^2 \le \mathcal{C}\left( \int_0^t \mathbb{E}\left|\tilde{X}_s^\varepsilon - \bar{X}_s\right|^2 ds + \Delta + \varepsilon \right). \tag{3.62}
$$

The Gronwall inequality then implies (3.34). ∎

Now we can finish the proof for the $O(\varepsilon^{1/2})$ strong convergence rate.

*Proof.* [Proof of Theorem 3.4] In Proposition 3.7 and 3.8, taking $\Delta = \varepsilon$, we have

$$
\mathbb{E}\left|X_t^\varepsilon - \bar{X}_t\right|^2 \le 2\mathbb{E}\left|X_t^\varepsilon - \tilde{X}_t^\varepsilon\right|^2 + 2\mathbb{E}\left|\tilde{X}_t^\varepsilon - \bar{X}_t\right|^2 \le \mathcal{C}\varepsilon. \tag{3.63}
$$

∎

## 4. Effective dynamics for fully coupled systems

The previous sections concern only the situation when $\sigma = \sigma(x,\varepsilon)$. In this section, we want to discuss the situation where $\sigma = \sigma(x,y,\varepsilon)$, i.e., the diffusion term in the slow dynamics does depend on the fast variable. It is easy to show that the weak convergence of (1.2) is still true for this case [7]. But the strong convergence does not hold. In the following, we will first show this by an example. Then we will provide the effective dynamics for the fully coupled system and prove the the strong convergence rate is also $O(\varepsilon^{1/2})$.

### 4.1. An illustrative example.

Consider the following example:

$$
\begin{aligned}
\dot{X}_t^\varepsilon &= Y_t^\varepsilon \dot{W}_t, & X_0 &= x, \\
\dot{Y}_t^\varepsilon &= -\frac{1}{\varepsilon} Y_t^\varepsilon + \frac{1}{\sqrt{\varepsilon}} \dot{W}_t, & Y_0 &= y.
\end{aligned} \tag{4.1}
$$

According to (1.3), we have $\bar{a} = 0$ and $\bar{\sigma}^2 = \frac{1}{2}$. Then (1.2) takes the form

$$
\dot{\bar{X}}_t = \bar{\sigma} \dot{W}_t. \tag{4.2}
$$

Although the above effective dynamics is true in the weak sense, we are going to show that strong convergence does not hold. Using Ito Isometry, we can easily calculate that

$$
\mathbb{E}\left|X_t^\varepsilon - \bar{X}_t\right|^2 = \mathbb{E}\left|\int_0^t \left(Y_t^\varepsilon - \bar{\sigma}\right) dW_t\right|^2 = \int_0^t \mathbb{E}\left|Y_t^\varepsilon - \bar{\sigma}\right|^2 dt = O(t), \tag{4.3}
$$

which implies that the strong convergence of effective dynamics according to (1.2) is not valid.

To give the effective dynamics for (4.1) in the strong sense, we define the following stationary process $\xi_t^\varepsilon$ with an Gaussian invariant measure of mean zero and variance $\frac{1}{2}$ satisfying the following SDE over the whole time domain $t\in(-\infty,\infty)$:

$$\dot{\xi}_t^\varepsilon = -\frac{1}{\varepsilon}\xi_t^\varepsilon + \frac{1}{\sqrt{\varepsilon}}\dot{W}_t. \tag{4.4}$$

The existence of the above process is provided in the Appendix. It is also shown in the Appendix that

$$\mathbb{E}\big|Y_t^\varepsilon - \xi_t^\varepsilon\big|^2 \le (y^2+1)e^{-2\frac{t}{\varepsilon}}. \tag{4.5}$$

The above estimate suggests the effective slow dynamics could be given in the following form:

$$\dot{\bar{X}} = \xi_t^\varepsilon \dot{W}_t, \qquad \bar{X}_0 = x, \tag{4.6}$$

for which the strong convergence holds with the rate

$$\mathbb{E}\big|X_t^\varepsilon - \bar{X}_t\big|^2 = \mathbb{E}\Big|\int_0^t \big(Y_t^\varepsilon - \xi_t^\varepsilon\big)dW_t\Big|^2 = \int_0^t \mathbb{E}\big|Y_t^\varepsilon - \xi_t^\varepsilon\big|^2 dt = O(\varepsilon). \tag{4.7}$$

Equation (4.6) gives a closed form dynamics in terms of the slow variable $x$, but its coefficient still depends on $\varepsilon$. A natural question is whether there exists a closed form dynamics in terms of only $x$ such that (4.6) converges strongly when $\varepsilon\to 0$. We are going to show that the answer is no. From the simple form of (4.6), we can see that this problem can be reduced to the existence of a random process independent of $\varepsilon$ to which the process $\int_0^t \xi_s^\varepsilon dW_s$ converges strongly. We want to show that this is not possible. Direct calculation shows that

$$\begin{aligned}\mathbb{E}\xi_t^\varepsilon \xi_t^{\varepsilon'} &= \mathbb{E}\Big(\frac{1}{\sqrt{\varepsilon}}\int_{-\infty}^t e^{-(t-s)/\varepsilon}dW_s \frac{1}{\sqrt{\varepsilon'}}\int_{-\infty}^t e^{-(t-s)/\varepsilon'}dW_s\Big)\\ &= \mathbb{E}\frac{1}{\sqrt{\varepsilon\varepsilon'}}\Big(\int_{-\infty}^t e^{-(t-s)\left(\frac{1}{\varepsilon}+\frac{1}{\varepsilon'}\right)}ds\Big)\\ &= \frac{\sqrt{\varepsilon\varepsilon'}}{\varepsilon+\varepsilon'}.\end{aligned} \tag{4.8}$$

Letting $\varepsilon' = \varepsilon^2$, we have

$$\mathbb{E}\xi_t^\varepsilon \xi_t^{\varepsilon^2} = O(\varepsilon^{1/2}) \to 0, \quad \varepsilon\to 0. \tag{4.9}$$

The above inequality means that as Gaussian random variables, $\xi_t^\varepsilon$ and $\xi_t^{\varepsilon^2}$ are asymptotically independent to each other when $\varepsilon\to 0$. Ito Isometry then implies that

$$\mathbb{E}\Big(\int_0^t \xi_s^\varepsilon dW_s \int_0^t \xi_s^{\varepsilon^2} dW_s\Big) = \int_0^t \mathbb{E}\xi_s^\varepsilon \xi_s^{\varepsilon^2} ds = O(\varepsilon^{1/2}). \tag{4.10}$$

Suppose that there is a stochastic process $\eta_t$ such that

$$\mathbb{E}\big|\int_0^t \xi_s^\varepsilon dW_s - \eta_t\big|^2 \to 0, \quad \varepsilon\to 0. \tag{4.11}$$

Then we have

$$\begin{aligned}&\mathbb{E}\Big|\int_0^t \xi_s^\varepsilon dW_s - \int_0^t \xi_s^{\varepsilon^2} dW_s\Big|^2\\ &\le 2\mathbb{E}\Big|\int_0^t \xi_s^\varepsilon dW_s - \eta_t\Big|^2 + 2\mathbb{E}\Big|\int_0^t \xi_s^{\varepsilon^2} dW_s - \eta_t\Big|^2 \to 0, \quad \varepsilon\to 0. \end{aligned}\tag{4.12}$$

This contradicts the fact (4.10) since by the stationarity of $\xi_t^\varepsilon$, we have

$$\begin{aligned}&\mathbb{E}\Big(\int_0^t \xi_s^\varepsilon dW_s - \int_0^t \xi_t^{\varepsilon^2} dW_s\Big)^2\\ &= \mathbb{E}\Big(\int_0^t \xi_s^\varepsilon dW_s\Big)^2 + \mathbb{E}\Big(\int_0^t \xi_s^{\varepsilon^2} dW_s\Big)^2 + 2\mathbb{E}\Big(\int_0^t \xi_s^\varepsilon dW_s \int_0^t \xi_s^{\varepsilon^2} dW_s\Big)\\ &= \int_0^t \mathbb{E}\big(\xi_s^\varepsilon\big)^2 ds + \int_0^t \mathbb{E}\big(\xi_s^{\varepsilon^2}\big)^2 ds + 2\int_0^t \mathbb{E}\xi_s^\varepsilon \xi_s^{\varepsilon^2} ds\\ &= t + O(\varepsilon^{1/2}). \end{aligned}\tag{4.13}$$

**4.2. Effective dynamics with strong convergence.** Now we want to establish the validity of effective dynamics of type (4.10) by proving its strong convergence. For each $(x,\varepsilon)$, the existence of a stationary process with invariant distribution $\mu_x^\varepsilon$ is guaranteed by Lemma A.2 provided in the Appendix such that for $t\in(-\infty,\infty)$,

$$\dot{\xi}_t(x,\varepsilon) = \frac{1}{\varepsilon}B\Big(x,\xi_t(x,\varepsilon),\varepsilon\Big) + \frac{1}{\sqrt{\varepsilon}}C\Big(x,\xi_t(x,\varepsilon),\varepsilon\Big)\dot{W}_t. \tag{4.14}$$

Using the integration by parts and the Ito formula, we obtain for any process $X_t$ with quadratic variation

$$\begin{aligned} d\xi_t(X_t,\varepsilon) = {}& \frac{1}{\varepsilon}B\Big(X_t,\xi_t(X_t,\varepsilon),\varepsilon\Big)dt + \frac{1}{\sqrt{\varepsilon}}C\Big(X_t,\xi_t(X_t,\varepsilon),\varepsilon\Big)dW_t\\ &+\eta_t(X_t,\varepsilon)dX_t + \frac{1}{2}\theta_t(X_t,\varepsilon) < dX_t, dX_t >\\ &+ < d\eta_t(X_t,\varepsilon), dX_t >, \end{aligned}\tag{4.15}$$

where

$$\eta_t(x,\varepsilon) = \nabla_x \xi_t(x,\varepsilon), \quad \theta_t(x,\varepsilon) = \nabla_{xx}\xi_t(x,\varepsilon). \tag{4.16}$$

Let $X_t^\varepsilon$ and $\bar{X}_t^\varepsilon$ be solutions of (1.1) and (1.5), respectively. The following theorem says that when $\varepsilon \ll 1$, the difference between $X_t^\varepsilon$ and $\bar{X}_t^\varepsilon$ in the strong sense is $O(\varepsilon^{1/2})$.

THEOREM 4.1. *Suppose Assumption 3.1– 3.3 holds. Then for any $T_0>0$, there exists a constant $\mathcal{C}>0$ independent of $\varepsilon$ such that*

$$\sup_{0\le t\le T_0} \mathbb{E}|X_t^\varepsilon - \bar{X}_t^\varepsilon|^2 \le \mathcal{C}\varepsilon. \tag{4.17}$$

To prove the above theorem, we define the following process that is similar to (3.16):

$$\begin{aligned} \hat{X}_t^\varepsilon &= x + \int_0^t a\Big(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta}, \hat{Y}_s^\varepsilon,\varepsilon\Big)ds + \int_0^t \sigma\Big(X_s^\varepsilon, Y_s^\varepsilon,\varepsilon\Big)dW_s,\\ \hat{Y}_t^\varepsilon &= y + \frac{1}{\varepsilon}\int_0^t B\Big(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta}, \hat{Y}_s^\varepsilon,\varepsilon\Big)ds + \frac{1}{\sqrt{\varepsilon}}\int_0^t C\Big(X^\varepsilon_{\lfloor s/\Delta\rfloor\Delta}, \hat{Y}_s^\varepsilon,\varepsilon\Big)dW_s. \end{aligned}\tag{4.18}$$

By the same proof for Lemma 3.5 and 3.6, we can easily have the following Lemmas.

Lemma 4.2. *For any $T_0>0$, there exists a constant $\mathcal{C}$ independent of $(\varepsilon,\Delta)$ such that*

$$\mathbb{E}\left|\hat{Y}_t^{\varepsilon}\right|^2\le\mathcal{C}. \tag{4.19}$$

Lemma 4.3. *For any $T_0>0$, there exists a constant $\mathcal{C}>0$ independent of $(\varepsilon,\Delta)$ such that*

$$\sup_{0\le t\le T_0}\mathbb{E}\left|Y_t^{\varepsilon}-\hat{Y}_t^{\varepsilon}\right|^2\le\mathcal{C}\Delta. \tag{4.20}$$

We also easily have the following Proposition by the same argument as for Propositon 3.7.

Proposition 4.4. *For any $T_0>0$, there exists a constant $\mathcal{C}>0$ independent of $(\varepsilon,\Delta)$ such that*

$$\sup_{0\le t\le T_0}\mathbb{E}\left|X_t^{\varepsilon}-\hat{X}_t^{\varepsilon}\right|^2\le\mathcal{C}\Delta. \tag{4.21}$$

Now we want to provide the counterpart of Proposition 3.8 for fully coupled systems.

Proposition 4.5. *For any $T_0>0$, there exists a constant $\mathcal{C}>0$ independent of $(\varepsilon,\Delta)$ such that*

$$\sup_{0\le t\le T_0}\mathbb{E}|\hat{X}_t^{\varepsilon}-\bar{X}_t^{\varepsilon}|^2\le\mathcal{C}\Big(\Delta+\varepsilon\Big). \tag{4.22}$$

*Proof.* By Assumption 3.1 and Ito Isometry, we have

$$\begin{aligned}\mathbb{E}|\hat{X}_t^{\varepsilon}-\bar{X}_t^{\varepsilon}|^2\le&\,2\mathbb{E}\left(\int_0^t\Big\{a\Big(X^{\varepsilon}_{\lfloor s/\Delta\rfloor\Delta},\hat{Y}_s^{\varepsilon},\varepsilon\Big)-a\Big(\bar{X}_s^{\varepsilon},\xi_s(\bar{X}_s^{\varepsilon},\varepsilon),\varepsilon\Big)\Big\}ds\right)^2\\&+2\mathbb{E}\left(\int_0^t\Big\{\sigma\Big(X_s^{\varepsilon},Y_s^{\varepsilon},\varepsilon\Big)-\sigma\Big(\bar{X}_s^{\varepsilon},\xi_s(\bar{X}_s^{\varepsilon},\varepsilon),\varepsilon\Big)\Big\}dW_s\right)^2\\\le&\,\mathcal{C}\mathbb{E}\int_0^t ds\Big\{\Big|X^{\varepsilon}_{\lfloor s/\Delta\rfloor\Delta}-X_s^{\varepsilon}\Big|^2+\Big|X_s^{\varepsilon}-\hat{X}_s^{\varepsilon}\Big|^2+\Big|\hat{X}_s^{\varepsilon}-\bar{X}_s^{\varepsilon}\Big|^2\\&+\Big|Y_s^{\varepsilon}-\hat{Y}_s^{\varepsilon}\Big|^2+\Big|Y_s^{\varepsilon}-\xi_s\Big(\bar{X}_s^{\varepsilon},\varepsilon\Big)\Big|^2\Big\}.\end{aligned} \tag{4.23}$$

By the quadratic variation of Brownian motions, Lemma 4.3 and Proposition 4.4, we have by the same argument for (3.26) that

$$\mathbb{E}\Big|X^{\varepsilon}_{\lfloor s/\Delta\rfloor\Delta}-X_s^{\varepsilon}\Big|^2,\ \mathbb{E}\Big|X_s^{\varepsilon}-\hat{X}_s^{\varepsilon}\Big|^2,\ \mathbb{E}\Big|Y_s^{\varepsilon}-\hat{Y}_s^{\varepsilon}\Big|^2\le\mathcal{C}\Delta. \tag{4.24}$$

By the Ito formula and Assumptions 3.1 and 3.3 , it is easy to show that

$$\mathbb{E}\big|\eta_t(x,\varepsilon)\big|^2,\ \mathbb{E}\big|\theta_t(x,\varepsilon)\big|^2\le\mathcal{C}. \tag{4.25}$$

Using the above estimate and (4.15), we can have

$$d\mathbb{E}\Big|Y_s^{\varepsilon}-\xi_s\Big(\bar{X}_s^{\varepsilon},\varepsilon\Big)\Big|^2\le\frac{2}{\varepsilon}\mathbb{E}\Big\langle Y_s^{\varepsilon}-\xi_s\Big(\bar{X}_s^{\varepsilon},\varepsilon\Big),\ B\Big(X_t^{\varepsilon},Y_t^{\varepsilon},\varepsilon\Big)-B\Big(\bar{X}_t,\xi_t(\bar{X}_t,\varepsilon),\varepsilon\Big)\Big\rangle dt$$

$$
\begin{aligned}
&+\frac{5}{4\varepsilon}\mathbb{E}\Big\|C\Big(X_t^\varepsilon,Y_t^\varepsilon,\varepsilon\Big)-C\Big(\bar{X}_t,\xi_t(\bar{X}_t,\varepsilon),\varepsilon\Big)\Big\|^2dt+\mathcal{C}dt\\
&\le-\frac{\beta}{\varepsilon}\Big|Y_s^\varepsilon-\xi_s\Big(\bar{X}_s^\varepsilon,\varepsilon\Big)\Big|^2dt+\frac{\mathcal{C}'}{\varepsilon}\Big|X_s^\varepsilon-\bar{X}_s^\varepsilon\Big|^2dt+\mathcal{C}''dt,
\end{aligned}
\tag{4.26}
$$

which, by the Gronwall inequality, implies that

$$
\mathbb{E}\Big|Y_s^\varepsilon-\xi_s\Big(\bar{X}_s^\varepsilon,\varepsilon\Big)\Big|^2\le\mathcal{C}\Big(\sup_{0\le\tau\le s}\Big|X_\tau^\varepsilon-\bar{X}_\tau^\varepsilon\Big|^2+\varepsilon\Big).
\tag{4.27}
$$

Taking the supremum on both sides of (4.23), we obtain

$$
\sup_{0\le s\le t}\mathbb{E}|\hat{X}_s^\varepsilon-\bar{X}_s^\varepsilon|^2\le\mathcal{C}\mathbb{E}\int_0^t ds\Big(\sup_{0\le\tau\le s}\Big|\hat{X}_\tau^\varepsilon-\bar{X}_\tau^\varepsilon\Big|^2+\Delta+\varepsilon\Big),
\tag{4.28}
$$

which gives (4.22). ∎

Now we can finish the proof for the $O(\varepsilon^{1/2})$ strong convergence rate for the fully coupled system.

*Proof.* [Proof of Theorem 4.1.] In Proposition 4.4 and 4.5, taking $\Delta=\varepsilon$ we have

$$
\mathbb{E}\big|X_t^\varepsilon-\bar{X}_t^\varepsilon\big|^2\le 2\mathbb{E}\big|X_t^\varepsilon-\hat{X}_t^\varepsilon\big|^2+2\mathbb{E}\big|\hat{X}_t^\varepsilon-\bar{X}_t^\varepsilon\big|^2\le\mathcal{C}\varepsilon.
\tag{4.29}
$$

∎

**Conclusion.** We proved the strong convergence for the Principle of Averaging for stochastic differential equations with two well separated time scales. The optimal rate of convergence was provided. The effective dynamics for fully coupled system was investigated. The analytical results will shed light on efficient and accurate numerical schemes for systems of this type.

**Acknowledgment.** We want to thank Weinan E and Eric Vanden-Eijnden for stimulating discussions.

**Appendix A. Limiting properties of the fast processes.** Here we want to provide some properties for the fast process $Z_{x,t}^\varepsilon$ defined in (3.7) on the infinite time horizon. The ergodicity and uniqueness of invariant measures of $Z_{x,t}^\varepsilon$ for each fixed $(x,\varepsilon)$ under Assumptions 3.1–3.3 have been established in [6]. We want to provide some sharper estimates for the purpose of proving our theorems. First, we give an energy estimate for $Z_{x,t}^\varepsilon$ and its invariant measure.

LEMMA A.1. *There exists a constant $\mathcal{C}$ such that for all $t\ge 0$, we have*

$$
\mathbb{E}\big|Z_{x,t}^\varepsilon\big|^2\le e^{-\frac{\beta}{\varepsilon}t}|z|^2+\mathcal{C}\Big(|x|^2+\varepsilon^2+1\Big),
\tag{A.1}
$$

*and the invariant measure $\mu_x^\varepsilon$ has a finite second order moment*

$$
\int z^2\mu_x^\varepsilon(dz)\le\mathcal{C}\Big(|x|^2+\varepsilon^2+1\Big),
\tag{A.2}
$$

*where $\mathcal{C}$ is the same constant as in* (A.1).

*Proof.* By (3.20) and Ito formula, we have

$$\begin{aligned} d\mathbb{E}\big|Z^{\varepsilon}_{x,t}\big|^2 &= \frac{2}{\varepsilon}\mathbb{E}\big\langle Z^{\varepsilon}_{x,t}, B\big(x, Z^{\varepsilon}_{x,t}\big)\big\rangle dt + \frac{1}{\varepsilon}\mathbb{E}\big\|C\big(x, Z^{\varepsilon}_{x,t},\varepsilon\big)\big\|^2 dt \\ &\le -\frac{\beta}{\varepsilon}\mathbb{E}\big|Z^{\varepsilon}_{x,t}\big|^2 + \mathcal{C}\big(|x|^2+\varepsilon^2+1\big). \end{aligned} \tag{A.3}$$

The Gronwall inequality then suggests

$$\mathbb{E}_z\big|Z^{\varepsilon}_{x,t}\big|^2 \le e^{-\frac{\beta}{\varepsilon}t}|z|^2 + \mathcal{C}\big(|x|^2+\varepsilon^2+1\big). \tag{A.4}$$

Taking $t\to\infty$, ergodicity gives (A.2). ∎

Based on ergodicity and Lemma A.1, we can prove existence of a stationary solution for Equation (3.7) satisfied by $Z^{\varepsilon}_{x,t}$.

LEMMA A.2. *For each fixed* $(x,\varepsilon)$, *there exists a process* $\xi^{\varepsilon}_{x,t}$ *defined over the whole time domain* $t\in(-\infty,\infty)$ *such that it satisfies* (3.7) *with a stationary probability distribution that agrees with the invariant measure of* (3.7), *i.e.,*

$$\dot{\xi}^{\varepsilon}_{x,t} = \frac{1}{\varepsilon}B\big(x, \xi^{\varepsilon}_{x,t},\varepsilon\big) + \frac{1}{\sqrt{\varepsilon}}C(x,\xi^{\varepsilon}_{x,t},\varepsilon)\dot{W}_t, \qquad \mathcal{L}(\xi^{\varepsilon}_{x,t}) = \mu^{\varepsilon}_x. \tag{A.5}$$

*Moreover, we have*

$$\mathbb{E}\big|Z^{\varepsilon}_{x,t}-\xi^{\varepsilon}_{x,t}\big|^2 \le \mathcal{C}\Big(|z|^2+|x|^2+\varepsilon^2+1\Big)e^{-2\frac{\beta}{\varepsilon}t}. \tag{A.6}$$

*Proof.* Define the process $Z^{\varepsilon}_{x,\tau,t}$ to be the solution of the following equation on the time domain $(\tau,\infty)$:

$$\dot{Z}^{\varepsilon}_{x,\tau,t} = \frac{1}{\varepsilon}B\big(x, Z^{\varepsilon}_{x,\tau,t},\varepsilon\big) + \frac{1}{\sqrt{\varepsilon}}C(x,Z^{\varepsilon}_{x,\tau,t},\varepsilon)\dot{W}_t, \qquad Z^{\varepsilon}_{x,\tau,\tau} = z. \tag{A.7}$$

For $\tau_2<\tau_1\le 0$, by the Ito formula and Assumption 3.3 we have

$$\begin{aligned} d\mathbb{E}\big|Z^{\varepsilon}_{x,\tau_1,t}-Z^{\varepsilon}_{x,\tau_2,t}\big|^2 &= \frac{2}{\varepsilon}\mathbb{E}\Big\langle Z^{\varepsilon}_{x,\tau_1,t}-Z^{\varepsilon}_{x,\tau_2,t},\ B\big(x,Z^{\varepsilon}_{x,\tau_1,t},\varepsilon\big)-B\big(x,Z^{\varepsilon}_{x,\tau_2,t},\varepsilon\big)\Big\rangle dt \\ &\quad + \frac{1}{\sqrt{\varepsilon}}\mathbb{E}\ \Big\|C\big(x,Z^{\varepsilon}_{x,\tau_1,t},\varepsilon\big)-C\big(x,Z^{\varepsilon}_{x,\tau_2,t},\varepsilon\big)\Big\|^2 dt \\ &\le -2\frac{\beta}{\varepsilon}\big|Z^{\varepsilon}_{x,\tau_1,t}-Z^{\varepsilon}_{x,\tau_2,t}\big|^2. \end{aligned} \tag{A.8}$$

Using Lemma A.1 we have

$$\begin{aligned} \mathbb{E}\big|Z^{\varepsilon}_{x,\tau_1,t}-Z^{\varepsilon}_{x,\tau_2,t}\big|^2 &\le \mathbb{E}\big|Z^{\varepsilon}_{x,\tau_2,\tau_1}-z\big|^2 e^{-2\beta\frac{(t-\tau_1)}{\varepsilon}} \\ &\le \mathcal{C}\Big(x^2+z^2+\varepsilon^2+1\Big)e^{-2\beta\frac{(t-\tau_1)}{\varepsilon}}. \end{aligned} \tag{A.9}$$

Therefore for any sequence $\{\tau_n\}$ such that $\tau_{n+1}\le\tau_n-1$, we have

$$\begin{aligned} \mathbb{P}\Big\{\big|Z^{\varepsilon}_{x,\tau_{n+1},t}-Z^{\varepsilon}_{x,\tau_n,t}\big|>\frac{1}{|\tau_n|^2}\Big\} &\le |\tau_n|^2\mathbb{E}\big|Z^{\varepsilon}_{x,\tau_{n+1},t}-Z^{\varepsilon}_{x,\tau_n,t}\big|^2 \\ &\le \mathcal{C}\Big(x^2+z^2+\varepsilon^2+1\Big)|\tau_n|^2 e^{-\beta\frac{(t-\tau_n)}{\varepsilon}}, \end{aligned} \tag{A.10}$$

which implies that

$$\sum_n |\tau_n|^2 e^{\frac{\beta}{\varepsilon}\tau_n} \le \sum_n |\tau_n|^2 |\tau_n - \tau_{n-1}| e^{\frac{\beta}{\varepsilon}\tau_n} \le \mathcal{C}\int_{-\infty}^0 x^2 e^{\frac{\beta}{\varepsilon}t} dt < \infty, \tag{A.11}$$

where we have assumed $\tau_0 = 0$. By the Borel-Cantelli Lemma, we know that with probability one, $Z^\varepsilon_{x,\tau_n,t}$ satisfies

$$\left|Z^\varepsilon_{x,\tau_{n+1},t} - Z^\varepsilon_{x,\tau_n,t}\right| \le \frac{1}{|\tau_n|^2}, \tag{A.12}$$

when $n \ge N(\omega)$. Since $\tau_n \le -n$ by the way $\{\tau_n\}$ is chosen, we know that $\{Z^\varepsilon_{x,\tau_n,t}\}$ is a converging sequence when $n \to \infty$. By the arbitrariness of $\{\tau_n\}$ we know that with probability one $Z^\varepsilon_{x,\tau,t}$ is converging when $\tau \to -\infty$. Otherwise with a nontrivial probability, we have for some $\delta(\omega) > 0$ such that for each $\tau_n$ we can find $\tau_a < \tau_n - 1$, $\tau_b < \tau_n - 1$ and $\left|Z^\varepsilon_{x,\tau_a,t} - Z^\varepsilon_{x,\tau_b,t}\right| > \delta$. Picking up $\tau_{n+1}$ such that $\left|Z^\varepsilon_{x,\tau_n,t} - Z^\varepsilon_{x,\tau_{n+1},t}\right| > \delta/2$, we can construct a sequence $Z^\varepsilon_{x,\tau_n,t}$ that contradicts (A.12). So we can define

$$\xi^\varepsilon_{x,t} = \lim_{\tau \to -\infty} Z^\varepsilon_{x,\tau,t}. \tag{A.13}$$

Notice that for any $s < t$, $Z^\varepsilon_{x,\tau,t}$ satisfies the integral equation

$$Z^\varepsilon_{x,\tau,t} = Z^\varepsilon_{x,\tau,s} + \frac{1}{\varepsilon}\int_s^t B\left(x, Z^\varepsilon_{x,\tau,\omega}, \varepsilon\right) d\omega + \frac{1}{\varepsilon}\int_s^t C\left(x, Z^\varepsilon_{x,\tau,\omega}, \varepsilon\right) dW_\omega. \tag{A.14}$$

Taking $\tau \to -\infty$ in the above equation implies that $\xi^\varepsilon_{x,t}$ really satisfies (3.7). By ergodicty, we have

$$\lim_{t \to \infty} \mathcal{L}\left(\xi^\varepsilon_{x,t}\right) = \mu^\varepsilon_x. \tag{A.15}$$

Meanwhile, the translation invariance of (A.13) suggests that $\xi^\varepsilon_{x,t}$ has a stationary distribution. Therefore we have $\mathcal{L} = \mu^\varepsilon_x$. Finally, taking $\tau_1 = 0$ and $\tau_2 \to -\infty$ in (A.9), by (A.2), we have

$$\mathbb{E}\left|Z^\varepsilon_{x,t} - \xi^\varepsilon_{x,t}\right|^2 \le \mathbb{E}\left|\xi^\varepsilon_{x,0} - z\right|^2 e^{-2\frac{\beta}{\varepsilon}t} \le \mathcal{C}\left(|z|^2 + |x|^2 + \varepsilon^2 + 1\right) e^{-2\frac{\beta}{\varepsilon}t}. \tag{A.16}$$

∎

Now we want to give the strong rate of convergence to the equilibrium for the process $Z^\varepsilon_{x,t}$.

Lemma A.3. *There exists a constant $\mathcal{C}$ such that for any function $f$ with bounded derivatives,*

$$\left|\mathbb{E}f\left(Z^\varepsilon_{x,t}\right) - \int f(z)\mu^\varepsilon_x(dz)\right| \le \mathcal{C}\sup|f'|\Big(|x| + |z| + \varepsilon + 1\Big) e^{-\frac{\beta}{\varepsilon}t}. \tag{A.17}$$

*Proof.* Let $\xi^\varepsilon_{x,t}$ be given by Lemma A.2. Suppose $f$ is a smooth function with bounded derivatives; we have

$$\begin{aligned}\left|\mathbb{E}f\left(Z^\varepsilon_{x,t}\right) - \int f(z)\mu^\varepsilon_x(dz)\right| &= \left|\mathbb{E}f\left(Z^\varepsilon_{x,t}\right) - \mathbb{E}f(\xi^\varepsilon_{x,t})\right| \\ &\le \sup|f'|\ \mathbb{E}\left|Z^\varepsilon_{x,t} - \xi^\varepsilon_{x,t}\right|, \end{aligned} \tag{A.18}$$

which, together with (A.6), gives (A.17). ∎

The following Lemma describes the behavior of the solution $Z^\varepsilon_{x,z,t}$ of (3.7) at equilibrium under the perturbation in $x$.

LEMMA A.4. *Functions $\hat{a}(x,z,t,\varepsilon)$ and $\bar{a}(x,\varepsilon)$ defined by (3.8) and (3.12) are smooth functions with bounded derivatives. In addition, for any multi-index $I=(I_1,\cdots,I_n)$, we have*

$$\left|\nabla_x^I\Big(\hat{a}\big(x,z,t,\varepsilon\big)-\bar{a}(x,\varepsilon)\Big)\right|\le \mathcal{C}_I e^{-\frac{\beta}{\varepsilon}t}, \tag{A.19}$$

*and for any multi-indices $J=(J_1,\cdots,J_m)$, we have*

$$\left|\nabla_z^J\nabla_x^I\Big(\hat{a}\big(x,z,t,\varepsilon\big)\Big)\right|\le \mathcal{C}_{I,J} e^{-\frac{\beta}{\varepsilon}t}. \tag{A.20}$$

*Proof.* Let us first focus on the first order derivatives $\nabla_x\hat{a}(\cdot)$, $\nabla_z\hat{a}(\cdot)$ and $\nabla_x\bar{a}(\cdot)$. The differentiability of the solution of (3.7) with respect to the initial condition and parameters under Assumption 3.1 is established in [8]. Therefore we have the smoothness of $\hat{a}(\cdot)$. Here we want to prove the boundedness of the derivatives. Letting $y_1=y+\theta z$ and $y_2=y$ in Assumption 3.3, we have

$$\begin{aligned}&\Big\langle z,\frac{1}{\theta}\Big(B(x,y+\theta z,\varepsilon)-B(x,y,\varepsilon)\Big)\Big\rangle+\frac{1}{\theta^2}\big\|C(x,y+\theta z,\varepsilon)-C(x,y,\varepsilon)\big\|^2\\ &\le -\beta|z|^2.\end{aligned} \tag{A.21}$$

Taking $\theta\to 0$ in the above inequality, we obtain

$$\big\langle z,\ \nabla_y B(x,y,\varepsilon)z\big\rangle+\big\|\nabla_y C(x,y,\varepsilon)z\big\|^2\le -\beta|z|^2. \tag{A.22}$$

Note that $Z^\varepsilon_{x,z,t}$ denotes the solution of Equation (3.7) with initial condition $z$ and parameter $(x,\varepsilon)$. Define

$$U_t=\nabla_x Z^\varepsilon_{x,z,t}, \tag{A.23}$$

then we have

$$\begin{aligned}dU_t=&\nabla_y B\big(x,Z^\varepsilon_{x,z,t},\varepsilon\big)U_t dt+\nabla_x B\big(x,Z^\varepsilon_{x,z,t},\varepsilon\big)dt\\ &+\nabla_y C\big(x,Z^\varepsilon_{x,z,t},\varepsilon\big)U_t dW_t+\nabla_x C\big(x,Z^\varepsilon_{x,z,t},\varepsilon\big)dW_t.\end{aligned} \tag{A.24}$$

Applying the Ito formula and Assumption 3.1, we have

$$\begin{aligned}d\mathbb{E}\big|U_t\big|^2=&2\mathbb{E}\Big\langle U_t,\nabla_y B\big(x,Z^\varepsilon_{x,z,t},\varepsilon\big)U_t\Big\rangle dt+2\mathbb{E}\Big\langle U_t,\nabla_x B\big(x,Z^\varepsilon_{x,z,t},\varepsilon\big)\Big\rangle dt\\ &+\mathbb{E}\Big\|\nabla_y C\big(x,Z^\varepsilon_{x,z,t},\varepsilon\big)U_t+\nabla_x C\big(x,Z^\varepsilon_{x,z,t},\varepsilon\big)\Big\|^2 dt\\ \le& -2\beta\big|U_t\big|^2+2\mathbb{E}\Big\langle U_t,\nabla_x B\big(x,Z^\varepsilon_{x,z,t},\varepsilon\big)\Big\rangle dt+2\mathbb{E}\Big\|\nabla_x C\big(x,Z^\varepsilon_{x,z,t},\varepsilon\big)U_t\Big\|^2 dt\\ \le& -\beta\big|U_t\big|^2+\mathcal{C},\end{aligned} \tag{A.25}$$

where $\mathcal{C}$ is a constant that only depends on the Lipschitz coefficients of $B$ and $C$. The Gronwall inequality and the initial condition $U_0=0$ imply that

$$\mathbb{E}\big|U_t\big|^2\le\mathcal{C}. \tag{A.26}$$

The boundedness of $\nabla_x \hat{a}(\cdot)$ follows from (A.26) and Assumption 3.1 since

$$\nabla_x \mathbb{E} a(x,z,t,\varepsilon) = \mathbb{E}\nabla_x a\big(x, Z^\varepsilon_{x,z,t},\varepsilon\big) + \mathbb{E}\nabla_y a\big(x, Z^\varepsilon_{x,z,t},\varepsilon\big) U_t. \tag{A.27}$$

Letting

$$V_t = \nabla_z Z^\varepsilon_{x,z,t}, \tag{A.28}$$

the boundedness of $\nabla_z \hat{a}(\cdot)$ can be obtained by repeating the above argument for $V_t$.

To prove (A.19), we notice that by the exponential mixing (A.17), the convergence to the invariant measure is uniform for $x$ in any compact set of $\mathbb{R}^n$. Therefore we can interchange the following limits:

$$\begin{aligned} \nabla_x \bar{a}(x,\varepsilon) &= \nabla_x \lim_{t\to\infty} \mathbb{E} a\big(x, Z^\varepsilon_{x,z,t},\varepsilon\big) \\ &= \lim_{t\to\infty} \nabla_x a\big(x, Z^\varepsilon_{x,z,t},\varepsilon\big) + \nabla_y a\big(x, Z^\varepsilon_{x,z,t},\varepsilon\big) U_t. \end{aligned} \tag{A.29}$$

Let $\xi^\varepsilon_{x,t}$ be the process defined in (4.14) and define

$$\eta^\varepsilon_{x,t} = \nabla_x \xi^\varepsilon_{x,t}. \tag{A.30}$$

Using (A.22) and (A.6), we can easily show that

$$\mathbb{E}\big|\eta^\varepsilon_{x,t}\big|^2 \le \mathcal{C} \tag{A.31}$$

and

$$\mathbb{E}\big|U_t - \eta^\varepsilon_{x,t}\big|^2 \le \mathcal{C} e^{-\frac{\beta}{\varepsilon} t}. \tag{A.32}$$

At the same time, we have

$$\begin{aligned} \nabla_x \bar{a}(x,\varepsilon) &= \lim_{t\to\infty} \nabla_x \mathbb{E}\Big(a\big(x, \xi^\varepsilon_{x,t},\varepsilon\big)\Big) \\ &= \lim_{t\to\infty} \mathbb{E}\Big(\nabla_x a\big(x, \xi^\varepsilon_{x,t},\varepsilon\big) + \nabla_y a\big(x, \xi^\varepsilon_{x,t},\varepsilon\big)\eta^\varepsilon_{x,t}\Big). \end{aligned} \tag{A.33}$$

Using (A.6) and (A.32), we have (A.19) for $\nabla_x \bar{a}(\cdot)$. (A.20) for $\nabla_z \hat{a}(\cdot)$ can be proved similarly as above by replacing $U_t$ with $V_t$ defined in (A.28).

Reiterating the same argument as above, we have have boundedness for higher order derivatives of $\hat{a}(x,z,t,\varepsilon)$ and $\bar{a}(x,\varepsilon)$, and prove (A.19) and (A.20) for arbitrary $I$ and $J$. ∎

## REFERENCES


[1] W. E and B. Engquist, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1(1), 87–133, 2003.

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

[3] M.I. Freidlin and A.D. Wentzell, *Random Perturbations of Dynamical Systems*, 2nd edition, Springer-Verlag, New York, 1998.

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

[5] R.Z. Khasminskii, *Principle of averaging for parabolic and elliptic differential equations and for Markov processes with small diffusion*, Theory Probab. Appl., 8, 1–21, 1963.

[6] R.Z. Khasminskii, *Stochastic stability of differential equations*, Sijthoff & Noordhoff, 1980.

[7] R.Z. Khasminskii and G. Yin, *On averaging principles: an asymptotic expansion approach*, SIAM J. Math. Anal., 35(6), 1534–1560, 2004.

[8] H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, Cambridge University Press, Cambridge, UK, 1990.

[9] Y. Kifer, *Stochastic versions of Anosov and Neistadt theorems on averaging*, Stoch. Dyn., 1(1), 1–21, 2001.

[10] S.P. Meyn and R.L. Tweedie, *Stability of Markovian processes, I, II* & *III*, Adv. in Appl. Probab., 24(3), 542-574, 1992; 25(3), 487-517, 1993; 25(3), 518–548, 1993.

[11] A. Yu. Veretennikov, *On an averaging principle for systems of stochastic differential equations*, Math. Sbornik, 181(2), 256–268, 1990.

[12] E. Vanden-Eijnden, *Numerical techniques for multi-scale dynamical systems with stochastic effects*, Commun. Math. Sci., 1, 377–384, 2003.