STRONG AND WEAK CONVERGENCE IN THE AVERAGING PRINCIPLE FOR SDES WITH HÖLDER COEFFICIENTS

MICHAEL RÖCKNER, XIAOBIN SUN AND LONGJIE XIE

ABSTRACT. Using Zvonkin's transform and the Poisson equation in $\mathbb{R}^d$ with a parameter, we prove the averaging principle for stochastic differential equations with time-dependent Hölder continuous coefficients. Sharp convergence rates with order $(\alpha\wedge 1)/2$ in the strong sense and $(\alpha/2)\wedge 1$ in the weak sense are obtained, considerably extending the existing results in the literature. Moreover, we prove that the convergence of the multi-scale system to the effective equation depends only on the regularity of the coefficients of the equation for the slow variable, and does not depend on the regularity of the coefficients of the equation for the fast component.

AMS 2010 Mathematics Subject Classification: 60H10, 60J60, 35B30
Keywords and Phrases: Averaging principle; Zvonkin's transformation; Poisson equation; multi-scale system.

1. INTRODUCTION

In this paper, we consider the following stochastic slow-fast system in $\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$:

$$\begin{cases} \mathrm{d}X_t^\varepsilon = \varepsilon^{-1}b(X_t^\varepsilon,Y_t^\varepsilon)\mathrm{d}t+\varepsilon^{-1/2}\sigma(X_t^\varepsilon,Y_t^\varepsilon)\mathrm{d}W_t^1, & X_0^\varepsilon = x\in\mathbb{R}^{d_1},\\ \mathrm{d}Y_t^\varepsilon = F(t,X_t^\varepsilon,Y_t^\varepsilon)\mathrm{d}t+G(t,X_t^\varepsilon,Y_t^\varepsilon)\mathrm{d}W_t^2, & Y_0^\varepsilon = y\in\mathbb{R}^{d_2}, \end{cases} \tag{1.1}$$

where $d_1,d_2\geqslant 1$, W_t^1 and W_t^2 are d_1, d_2-dimensional independent standard Brownian motions both defined on some probability space $(\Omega,\mathscr{F},\mathbb{P})$, $b:\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}\to\mathbb{R}^{d_1}$, $F:\mathbb{R}_+\times\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}\to\mathbb{R}^{d_2}$, $\sigma:\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}\to\mathbb{R}^{d_1}\otimes\mathbb{R}^{d_1}$ and $G:\mathbb{R}_+\times\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}\to\mathbb{R}^{d_2}\otimes\mathbb{R}^{d_2}$ are measurable functions, and the parameter $\varepsilon>0$ represents the ratio between the timescales of X_t^ε and Y_t^ε variables. Such multiscale model appears naturally in the theory of nonlinear oscillations, chemical kinetics, biology, climate dynamics and many other areas leading to a mathematical description involving 'slow' and 'fast' phase variables, see e.g. [1, 17, 27, 33] and the references therein. Usually, the underlying system (1.1) is difficult to deal with due to the two widely separated timescales and the cross interactions of slow and fast modes. Hence, the asymptotic study of the behavior of the system as $\varepsilon\to 0$ is of great interest and has attracted much attentions in the past decades.

It is known that under suitable regularity assumptions on the coefficients, the slow part Y_t^ε will converge to the solution of the following reduced equation in $\mathbb{R}^{d_2}$:

$$\mathrm{d}\bar{Y}_t=\bar{F}(t,\bar{Y}_t)\mathrm{d}t+\bar{G}(t,\bar{Y}_t)\mathrm{d}W_t^2,\quad \bar{Y}_0=y, \tag{1.2}$$

This work is supported in part by NSFC (No. 11701233, 11771187, 11931004), NSF of Jiangsu (BK20170226) and the PAPD Project of Jiangsu Higher Education Institutions. Financial support of the DFG through CRC 1283 is gratefully acknowledged.

where the new averaged coefficients are given by

$$\bar{F}(t,y) := \int_{\mathbb{R}^{d_1}} F(t,x,y)\mu^y(\mathrm{d}x) \quad \text{and} \quad \bar{G}(t,y) := \sqrt{\int_{\mathbb{R}^{d_1}} G(t,x,y)G(t,x,y)^* \mu^y(\mathrm{d}x)}. \quad (1.3)$$

Here G^* is the transpose of the matrix G, and $\mu^y(\mathrm{d}x)$ is the unique invariant measure of the transition semigroup of the process X_t^y, which is the solution of the following frozen equation:

$$\mathrm{d}X_t^y = b(X_t^y, y)\mathrm{d}t + \sigma(X_t^y, y)\mathrm{d}W_t^1, \quad X_0^y = x. \quad (1.4)$$

The effective dynamic (1.2) then captures the evolution of the system (1.1) over a long timescale, which does not depend on the fast variable any more and thus is much simpler than SDE (1.1). This theory, known as the averaging principle, was first developed for deterministic ordinary differential equations (ODEs for short) by Bogolyubov and Krylov [25], and extended to the stochastic differential equations (SDEs for short) by Khasminskii [19]. We refer the readers to the book of Freidlin and Wentzell [13] for a comprehensive overview.

As a rule, the averaging method requires certain smoothness on both the original and the averaged coefficients. Various assumptions have been studied in order to guarantee the above convergence. Note that in the stochastic case, the convergence can be analyzed in two different ways: the strong convergence which provides pathwise asymptotic information for the system, and the weak convergence which gives convergence the laws of the processes. To the best of our knowledge, most of the results in the literature, both for the deterministic case and for the stochastic case, require at least local Lipschitz conditions on all the coefficients of system (1.1), see e.g. [9, 16, 18, 22, 23, 29]. There is only one paper by Veretennikov [36] where weak convergence for the time-independent system (1.1) was established under the assumptions that the drift coefficient F in the slow equation is bounded and measurable with respect to the y variable, and all the other coefficients are globally Lipschitz continuous. Therefore, it seems that there are no studies of the averaging principle for SDEs which concentrates on Hölder coefficients.

On the other hand, in the papers mentioned above, no order of convergence in terms of $\varepsilon \to 0$ is provided. But for numerical purposes, it is important to know the rate of convergence of the slow variable to the effective dynamics. The main motivation comes from the well-known Heterogeneous Multi-scale Methods used to approximate the slow component in system (1.1), see e.g. [4, 11]. Moreover, the rate of convergence is also known to be very important for functional limit theorems in probability theory and homogenization, see e.g. [21, 31, 32, 39]. In this direction, the strong convergence with order 1/2 and weak convergence with order 1 are known to be optimal, see [15, 20, 30, 34, 42]. As far as we know, all the known results in the literature concerning the rate of convergence require essentially at least C_b^2-regularity for all the coefficients, and none of them considered the fully coupled cases, i.e., the diffusion coefficient in the slow equation can not depend on the fast term. We also mention that the averaging principle for stochastic partial differential equations and rates of convergence have also been widely studied, we refer to [3, 5, 6, 7, 10, 14] and the references therein.

The main aim of this work is to develop a very general, robust and unified method for establishing the averaging principle, involving both strong and weak convergence, for the multi-scale system (1.1) with **irregular** coefficients, which leads to simplifications and extensions of the existing results. Unlike most previous publications, we mainly focus on the "impact of noises" on the averaging principle for system (1.1). More precisely, we shall prove that under the non-degeneracy of the noises, the averaging principle holds for system (1.1) with only Hölder continuous coefficients, see **Theorem 2.3**. Note that the deterministic system can even be ill-posed under such weak conditions on the coefficients. Moreover, we obtain the strong convergence rate with order $(\alpha\wedge 1)/2$ and the weak convergence rate in the fully coupled case with order $(\alpha/2)\wedge 1$, where $\alpha>0$ is the Hölder index of the coefficients with respect to the slow component (y-variable), see **Theorem 2.1** and **Theorem 2.5** respectively. In particular, the convergence rates do not depend on the regularity of the coefficients with respect to the fast term (x-variable), which appear to be a new observation and which we think provides some new insight for understanding the averaging principle. See **Remark 2.2** and **Remark 2.6** for more detailed comparisons of our results with the previous publications on the subject.

The averaging principle for system (1.1) is also known to be closely related to the behavior of solutions for second-order parabolic and elliptic partial differential equations, see [12, 21, 36] and the references therein. In fact, the infinitesimal operator corresponding to $(X_t^\varepsilon, Y_t^\varepsilon)$ has the form

$$\mathscr{L}^\varepsilon := \varepsilon^{-1}\mathscr{L}_0(x,y)+\mathscr{L}_1(x,y),$$

where

$$\mathscr{L}_0 := \mathscr{L}_0(x,y) := \sum_{i,j} a_{ij}(x,y)\frac{\partial^2}{\partial x_i\partial x_j}+b(x,y)\cdot\nabla_x, \tag{1.5}$$

$$\mathscr{L}_1 := \mathscr{L}_1(x,y) := \sum_{i,j} H_{ij}(t,x,y)\frac{\partial^2}{\partial y_i\partial y_j}+F(t,x,y)\cdot\nabla_y \tag{1.6}$$

with $a(x,y):=\sigma(x,y)\sigma(x,y)^*/2$ and $H(t,x,y):=G(t,x,y)G(t,x,y)^*/2$. Given a $T>0$, consider the following Cauchy problem in $[0,T]\times\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$:

$$\begin{cases}\partial_t u^\varepsilon(t,x,y)+\mathscr{L}^\varepsilon u^\varepsilon(t,x,y)=\psi(y), \quad 0\leqslant t<T,\\ u^\varepsilon(T,x,y)=\varphi(y).\end{cases} \tag{1.7}$$

Using Theorem 2.5, we can study the behavior of the solution u^ε to equation (1.7) as $\varepsilon\to 0$. More precisely, we shall prove that $u^\varepsilon(t,x,y)$ converges to the solution $\bar{u}(t,y)$ of the following reduced Cauchy problem in $[0,T]\times\mathbb{R}^{d_2}$:

$$\begin{cases}\partial_t \bar{u}(t,y)+\bar{\mathscr{L}}\bar{u}(t,y)=\psi(y), \quad 0\leqslant t<T,\\ \bar{u}(T,y)=\varphi(y),\end{cases} \tag{1.8}$$

where ψ is a bounded measurable function, φ is bounded continuous, and $\bar{\mathscr{L}}$ is the infinitesimal generator of the effective SDE (1.2), i.e.,

$$\bar{\mathscr{L}} := \sum_{i,j}\bar{H}_{ij}(t,y)\frac{\partial^2}{\partial y_i\partial y_j}+\bar{F}(t,y)\cdot\nabla_y \tag{1.9}$$

with $\bar{H}(t,y) := \bar{G}(t,y)\bar{G}(t,y)^*/2$, and $\bar{F}, \bar{G}$ are as defined in (1.3). The main result in this direction is given by **Theorem 2.7**.

As mentioned before, the argument that we shall use is rather simple insofar as it does not involve the classical time discretisation procedure, which is commonly used in the literature to prove the averaging principle. Two ingredients are crucial in our proof: Zvonkin's transformation and the Poisson equation in the whole space. First of all, due to the low regularity of the coefficients, we shall use Zvonkin's argument to transform the equation for Y_t^ε and $\bar{Y}_t$ into new ones. Such technique was first developed in [44] and is now widely used to study the strong well-posedness for SDEs with singular coefficients, see e.g. [26, 40, 41, 43]. Then we use the Poisson equation with a parameter to prove both the strong and weak convergence for system (1.1). Here we adopt and improve the idea used in [5], where the convergence rate in the averaging principle for SPDEs with smooth coefficients with the fast equation not depending on the slow component was studied. More precisely, we shall study the following Poisson equation in $\mathbb{R}^{d_1}$:

$$\mathscr{L}_0(x,y)u(x,y) = -f(x,y), \quad x \in \mathbb{R}^{d_1}, \tag{1.10}$$

where $y \in \mathbb{R}^{d_2}$ is a parameter and $\mathscr{L}_0(x,y)$ is defined by (1.5). We note that there is no boundary condition. When the equation is formulated in a compact set, the corresponding theory is well known. However, equation (1.10) in the whole space $\mathbb{R}^{d_1}$ has been studied only very recently, and it turns out to be very useful in the theory of the averaging principle, diffusion approximation and other limit theorems, see the series of papers [2, 31, 32, 38]. We shall derive estimates for the solution of (1.10) in terms of explicit conditions on the coefficients as well as the right hand side, see **Theorem 3.1**, which generalizes the results in [31, 32] and is of independent interest.

The paper is organized as follows. In Section 2, we state our main results. Section 3 is devoted to the study of the Poisson equation in the whole space with a parameter. The proofs of strong convergence and weak convergence are given in Section 4 and Section 5, respectively.

To end this section, we introduce some notations. Let $\mathbb{N} := \{0, 1, \cdots\}$ and $\mathbb{N}^* := \{1, 2\cdots\}$. For $\beta \in (0,1)$, let $C^\beta(\mathbb{R}^d)$ be the usual local Hölder space. For $\beta \in \mathbb{N}^*$, without abuse of notation, we denote by $C^\beta(\mathbb{R}^d)$ the space of all functions f whose $\beta - 1$ order derivative $\partial^{\beta-1}f$ is Lipschitz continuous. While when $\beta \in (0,\infty) \setminus \mathbb{N}^*$, $C^\beta(\mathbb{R}^d)$ consists of all functions satisfying $f \in C^{[\beta]}(\mathbb{R}^d)$ and $\partial^{[\beta]}f \in C^{\beta-[\beta]}(\mathbb{R}^d)$, where $[\beta]$ denotes the largest integer which is smaller than β. For $\beta > 0$, we denote by $C_b^\beta(\mathbb{R}^d)$ the space of all functions $f \in C^\beta(\mathbb{R}^d)$ whose i-order derivative $\partial^i f$ is bounded for any $0 \leqslant i \leqslant ([\beta] - 1) \vee 0$.

Given a function f and $\gamma_1, \gamma_2, \gamma_3 \in (0,\infty)$, we shall consider the following three cases:
(i) f is defined on $\mathbb{R}^{d_1+d_2}$, i.e., f is a function with variable x and y: we write $f \in C_b^{\gamma_1,\gamma_2}$ if $f \in C_b^{\gamma_1}(\mathbb{R}^{d_1}; C_b^{\gamma_2}(\mathbb{R}^{d_2}))$, and $f \in C_{loc,y}^{\gamma_1,\gamma_2}$ means $f \in C_b^{\gamma_1}(\mathbb{R}^{d_1}; C^{\gamma_2}(\mathbb{R}^{d_2}))$;
(ii) f is defined on $\mathbb{R}_+ \times \mathbb{R}^{d_1+d_2}$, i.e., f is a function of t, x and y: we write $f \in C_b^{\gamma_3,\gamma_1,\gamma_2}$ if for every $t > 0$, $f(t,\cdot,\cdot) \in C_b^{\gamma_1,\gamma_2}$ and for every (x,y), $f(\cdot,x,y) \in C_b^{\gamma_3}(\mathbb{R}_+)$; similarly, $f \in C_{loc,y}^{\gamma_3,\gamma_1,\gamma_2}$ means that $f(t,\cdot,\cdot) \in C_{loc,y}^{\gamma_1,\gamma_2}$ and $f(\cdot,x,y) \in C_b^{\gamma_3}(\mathbb{R}_+)$;
(iii) f is defined on $\mathbb{R}_+ \times \mathbb{R}^{d_2}$, i.e., f is a function of t and y: we write $f \in C_b^{\gamma_3,\gamma_2}$ if for every $t > 0$, $f(t,\cdot) \in C_b^{\gamma_2}$ and for every $y \in \mathbb{R}^{d_2}$, $f(\cdot,y) \in C_b^{\gamma_3}(\mathbb{R}_+)$.

2. Assumptions and main results

Let us first introduce some basic assumptions. We shall assume the following non-degeneracy conditions on the diffusion coefficients:

(Hσ): The coefficient $a = \sigma\sigma^*$ is non-degenerate in x uniformly with respect to y, i.e., there exists a $\lambda > 1$ such that for any $x \in \mathbb{R}^{d_1}$ and $y \in \mathbb{R}^{d_2}$,

$$\lambda^{-1}|\xi|^2 \leqslant a^{ij}(x,y)\xi_i\xi_j \leqslant \lambda|\xi|^2, \quad \forall \xi \in \mathbb{R}^{d_1}.$$

(H$_G$): The coefficient $H = GG^*$ is non-degenerate in y uniformly with respect to (t,x), i.e., there exists a $\lambda > 1$ such that for any $(t,x) \in \mathbb{R}_+ \times \mathbb{R}^{d_1}$ and $y \in \mathbb{R}^{d_2}$,

$$\lambda^{-1}|\xi|^2 \leqslant H^{ij}(t,x,y)\xi_i\xi_j \leqslant \lambda|\xi|^2, \quad \forall \xi \in \mathbb{R}^{d_2}.$$

For the existence of an invariant measure for the frozen SDE (1.4), we assume the following very weak recurrence condition (see [32, 37]):

(Hb): $\qquad \lim_{|x|\to\infty} \sup_y \langle x, b(x,y)\rangle = -\infty.$

Below, we state our main results concerning the strong and weak convergence for the averaging principle for system (1.1) separately.

2.1. Strong convergence.

The following is the first main result of this paper.

Theorem 2.1. *Let* **(Hσ)**-**(H$_G$)**-**(Hb)** *hold, and let*

$$G(t,x,y) \equiv G(t,y). \tag{2.1}$$

Assume that $\sigma \in C_b^{1,1}$, $b \in C_b^{\delta,\alpha}$ *and* $F \in C_b^{\alpha/2,\delta,\alpha}$, $G \in C_b^{\alpha/2,1}$ *with* $0 < \delta, \alpha \leqslant 1$. *Then we have for any* $T > 0$,

$$\sup_{t\in[0,T]} \mathbb{E}|Y_t^\varepsilon - \bar{Y}_t|^2 \leqslant C_T\, \varepsilon^{\alpha\wedge 1}, \tag{2.2}$$

where $C_T > 0$ *is a constant independent of* δ.

We point out that under our assumptions, the strong well-posedness for system (1.1) was obtained by [35] or [43, Theorem 1.3], and the invariant measure $\mu^y(\mathrm{d}x)$ for SDE (1.4) exists and is unique, see [40, Theorem 1.2] or [41, Theorem 2.9]. Meanwhile, we shall show that the averaged drift $\bar{F}$ defined in (1.3) is also Hölder continuous, i.e., $\bar{F} \in C_b^{\alpha/2,\alpha}$ (see Lemma 4.1 below). Thus, there exists a unique strong solution $\bar{Y}_t$ to SDE (1.2).

Let us list some important comments to explain our result.

Remark 2.2. *We first point out that the independence of G with respect to the x-variable in assumption (2.1) is necessary. Otherwise, the strong convergence for SDE (1.1) may not be true, cf.* [30, 36].

(1) [Singular coefficients]. *We do not make any Lipschitz-type assumptions on the drift coefficients b and F. This is due to the regularization effect of the non-degenerate noises. Note that if $\sigma = 0$ or $G = 0$, the system (1.1) may even be ill-posed with only Hölder coefficients.*

(2) [Sharp order]. *Taking $\alpha = 1$ in (2.2), we can obtain the strong convergence with order $1/2$. Thus we get the optimal rate under much weaker regularity conditions both on the diffusion and the drift coefficients than the known results in the literature. Meanwhile, when $0 < \alpha < 1$, we also get that the averaging principle holds with a strong convergence rate $\alpha/2$, which to the best of our knowledge is new. Moreover, we allow the coefficients to be time-dependent, which appears to have not been studied before in estimating the rate of convergence.*

(3) [Dependence of convergence]. *Note that the convergence rate $(\alpha \wedge 1)/2$ does not depend on the index δ. This suggests that the convergence in the averaging principle replies only on the regularity of the coefficients with respect to the y (slow) variable, and does not depend on the regularity with respect to the x (fast) variable, which we think provides some new insight for understanding the averaging principle.*

By a localization technique as in [41, Corollary 2.6], we can drop the boundness condition on the coefficients with respect to the slow variable.

Theorem 2.3. *Let $(\mathbf{H}\sigma)$-$(\mathbf{H}_G)$-$(\mathbf{H}b)$ hold, and let*

$$G(t,x,y) \equiv G(t,y).$$

Assume $\sigma \in C^{1,1}_{loc,y}$, $G \in C^{\alpha/2,1}_{loc,y}$ and $b \in C^{\delta,\alpha}_{loc,y}$, $F \in C^{\alpha/2,\delta,\alpha}_{loc,y}$ with $0 < \delta, \alpha \leqslant 1$, and that the following moment estimate holds:

$(\mathbf{H}^M)$ *For any $T > 0$, there exists a $\beta > 2$ such that*

$$\sup_{\varepsilon\in(0,\varepsilon_0)} \mathbb{E}\Big[\sup_{t\in[0,T]} (|Y^\varepsilon_t|^\beta + |\bar{Y}_t|^\beta)\Big] \leqslant C < \infty,$$

where $\varepsilon_0 > 0$ and $C > 0$ is a constant depending on $T, |x|, |y|$, where x, y are the initial conditions in (1.1).

Then we have for any $T > 0$,

$$\lim_{\varepsilon\to 0} \sup_{t\in[0,T]} \mathbb{E}|Y^\varepsilon_t - \bar{Y}_t|^2 = 0.$$

Remark 2.4. *Local conditions imposed on the coefficients allow functions to have certain growth at infinity. The advantage of Theorem 2.3 lies in that, we only need to show the a priori moment estimate $(\mathbf{H}^M)$ in order to guarantee the strong convergence in the averaging principle for SDE (1.1) with only local Hölder continuous drifts.*

2.2. **Weak convergence.** In the above results, the assumptions $\sigma \in C^{1,1}_b$ and $G \in C^{\alpha/2,1}_b$ are mainly needed to ensure the strong well-posedness for system (1.1). Now, we state our main result concerning the weak convergence of system (1.1) under weaker conditions on the diffusion coefficients.

Theorem 2.5 (Weak convergence)**.** *Let $(\mathbf{H}\sigma)$-$(\mathbf{H}_G)$-$(\mathbf{H}b)$ hold true. Assume that $\sigma, b \in C^{\delta,\alpha}_b$ and $F, G \in C^{\alpha/2,\delta,\alpha}_b$ with $0 < \delta, \alpha \leqslant 2$. Then for any $T > 0$ and every $\varphi \in C^{2+\alpha}_b(\mathbb{R}^{d_2})$, we have*

$$\sup_{t\in[0,T]} \Big|\mathbb{E}[\varphi(Y^\varepsilon_t)] - \mathbb{E}[\varphi(Y_t)]\Big| \leqslant C_T\, \varepsilon^{(\alpha/2)\wedge 1}, \tag{2.3}$$

where $C_T > 0$ is a constant independent of δ.

We now give some comments to explain the above result.

Remark 2.6. *Note that here the diffusion coefficient G in the slow equation can also depend on the fast variable x.*

(1) [Singular coefficients]. *Due to the non-degeneracy of the noises, it is well-known that the system (1.1) is weakly well-posed under our conditions. This is the main reason why we can assume weaker conditions on the diffusion coefficients to prove the above weak convergence.*

(2) [Sharp order]. *Taking $\alpha = 2$ in (2.3), we obtain the optimal weak convergence rate 1. Our result generalizes the known results in the literature by allowing the coefficients to be time-dependent, and more importantly, to be fully coupled, i.e., the diffusion coefficient in the slow equation can depend on the fast variable, which appears to have not been considered before in estimating the rate of convergence. Meanwhile, when $0 < \alpha < 2$, we also get that the weak averaging principle holds with convergence rate $\alpha/2$, which also appears to be new.*

(3) [Dependence of convergence]. *As before, the weak convergence relies only on the regularity of all the coefficients with respect to the y (slow) variable, since the rate $(\alpha/2) \wedge 1$ does not depend on the index δ.*

As a direct consequence of Theorem 2.5, we have the following result concerning the limit behavior of parabolic equations.

Theorem 2.7. *Suppose the assumptions in Theorem 2.5 hold, ψ is bounded measurable and φ is bounded continuous. Let u^ε be the solution to equation (1.7). Then for every $t > 0$, $x \in \mathbb{R}^{d_1}$ and $y \in \mathbb{R}^{d_2}$, the limit*

$$\lim_{\varepsilon \to 0} u^\varepsilon(t, x, y) =: u(t, y)$$

exists, and the function $u(t, y)$ is the unique solution of the Cauchy problem (1.8).

3. Poisson equation in $\mathbb{R}^{d_1}$ with a parameter

This section is devoted to studying the Poisson equation (1.10) in the whole space. We are looking for a solution u for (1.10) which grows at most polynomial in x as $|x| \to \infty$, and the main problem addressed here is the regularity of the solution u with respect to the parameter y. Throughout this section, we shall always assume **(Hσ)** and **(Hb)** to hold. Let us point out that there is no boundary condition. As a result, the solution turns out to be defined up to an additive constant, since $\mathscr{L}(x, y)1 \equiv 0$. To fix this constant, it is necessary to make the following "centering" assumption on the right-hand side:

$$\int_{\mathbb{R}^{d_1}} f(x, y)\mu^y(\mathrm{d}x) = 0, \quad \forall y \in \mathbb{R}^{d_2}, \tag{3.1}$$

which is analogous to the centering in the standard Central Limit Theorem, see [31, 32] for more details.

We shall essentially use the strategy implemented in [32], where the fundamental solution was used to study the equation (1.10). More precisely, note that $\mathscr{L}_0(x, y)$ can be viewed as the infinitesimal generator of the process $X_t^y(x)$, which is the unique strong solution for the frozen SDE (1.4). As a result, the solution u to equation (1.10) should have the following probabilistic representation:

$$u(x,y)=\int_0^\infty \mathbb{E}f\big(X_t^y(x),y\big)\mathrm{d}t. \tag{3.2}$$

As we shall see below, under our assumptions, $X_t^y(x)$ admits a density function $p_t(x,x';y)$, which is also the unique fundamental solution for the operator $\mathscr{L}_0(x,y)$. Let $T_tf(x,y)$denotes the semigroup corresponding to $X_t^y(x)$, i.e.,

$$T_tf(x,y):=\mathbb{E}\big(f(X_t^y(x),y\big)=\int_{\mathbb{R}^{d_1}}p_t(x,x';y)f(x',y)\mathrm{d}x'.$$

Then we can write

$$u(x,y)=\int_0^1 T_tf(x,y)\mathrm{d}t+\int_1^\infty T_tf(x,y)\mathrm{d}t. \tag{3.3}$$

Thus, we need to study the behavior of T_tf as well as its first and second order derivatives with respect to the y-variable both near $t=0$ and as $t\to\infty$. The following is the main result of this section.

Theorem 3.1. *Let* $(\mathbf{H}\sigma)$ *and* $(\mathbf{H}b)$ *hold. Assume that* $a,b\in C_b^{\delta,\ell}$ *with* $0<\delta\leqslant 1$ *and* $\ell=0,1,2$. *Then for every function* $f\in C_b^{\delta,\ell}$ *satisfying (3.1), there exists a unique solution* u *to (1.10) such that for any* $y\in\mathbb{R}^{d_2}$, $u(\cdot,y)\in C^2$ *and for any* $x\in\mathbb{R}^{d_1}$, $u(x,\cdot)\in C_b^\ell$. *Moreover, there exists a constant* $m>0$ *such that for any* $y\in\mathbb{R}^{d_2}$,

$$|u(x,y)|+|\nabla_xu(x,y)|+|\nabla_x^2u(x,y)|\leqslant C_0\|f\|_{C_b^{\delta,0}}(1+|x|^m), \tag{3.4}$$

and when $\ell=1$,

$$\begin{aligned}|\nabla_yu(x,y)|\leqslant &C_0\Big[\big(\|a\|_{C_b^{\delta,0}}+\|b\|_{C_b^{\delta,0}}\big)\|f\|_{C_b^{\delta,1}}\\&+\big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\big)\|f\|_{C_b^{\delta,0}}\Big](1+|x|^m),\end{aligned} \tag{3.5}$$

and when $\ell=2$,

$$\begin{aligned}|\nabla_y^2u(x,y)|\leqslant C_0\Big[&\big(\|a\|_{C_b^{\delta,0}}+\|b\|_{C_b^{\delta,0}}\big)\|f\|_{C_b^{\delta,2}}+\big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\big)\|f\|_{C_b^{\delta,1}}\\&+\Big(\big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\big)^2+\big(\|a\|_{C_b^{\delta,2}}+\|b\|_{C_b^{\delta,2}}\big)\Big)\|f\|_{C_b^{\delta,0}}\Big](1+|x|^m),\end{aligned} \tag{3.6}$$

where C_0 *is a positive constant depending only on* λ,d_1,d_2 *and* $\|a\|_{C_b^{\delta,0}},\|b\|_{C_b^{\delta,0}}$.

Remark 3.2. *Concerning estimates (3.5)-(3.6), usually one does not care about the dependence of constants on the right hand side with respect to the norms of the coefficients. But this will be very important below for us to get the sharp rate of convergence for system (1.1) with only Hölder continuous coefficients. More precisely, since we assume the coefficients belong to the space* $C_b^{\delta,\alpha}$ *in Theorem 2.1 and Theorem 2.5, we need to keep track of the dependence of the constant on the right hand side of (3.5)-(3.6) with respect to the higher order norms of the coefficients* a,b *as well as of the potential term* f.

The proof of the above result relies on the materials addresses in the following two subsections. At the moment, we first give:

Proof of Theorem 3.1. The existence and uniqueness of the solution u to (1.10) are well-known under the above conditions, see e.g. [31, Theorem 1] and [32]. Meanwhile, by regarding $y \in \mathbb{R}^{d_2}$ as a parameter, the estimate (3.4) is true since all coefficients are bounded uniformly in the y variable. Concerning estimate (3.5), we have by (3.3) and Lemma 3.7 below that for any $k \in \mathbb{R}_+$, there exist constants $C_1, m > 0$ such that

$$
\begin{aligned}
|\nabla_y u(x,y)| &\leqslant \int_0^1 |\nabla_y T_t f(x,y)| \mathrm{d}t + \int_1^\infty |\nabla_y T_t f(x,y)| \mathrm{d}t \\
&\leqslant C_1 \Big[\big(\|a\|_{C_b^{\delta,0}} + \|b\|_{C_b^{\delta,0}} \big) \|f\|_{C_b^{\delta,1}} + \big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \big) \|f\|_{C_b^{\delta,0}} \Big] \\
&\quad \times \left(1 + \int_1^\infty \frac{(1+|x|^m)}{(1+t)^k} \mathrm{d}t \right),
\end{aligned}
$$

which in turn yields the desired result. The estimate (3.6) can be proved similarly. The proof is finished. □

Below, we proceed to study the first and second order derivatives of $T_t f(x,y)$ with respect to the y-variable in the following two subsections. We provide the explicit dependence on all higher order norms of the coefficients involved.

3.1. **First order derivative with respect to y.** Let us first recall some classical results concerning the fundamental solution $p_t(x, x'; y)$, see [8, Theorem 2.3], [28, Chapter IV] and [32, Proposition 3].

Lemma 3.3. *Assume* $(\mathbf{H}\sigma)$ *holds and let* $T > 0$. *Let* $a, b \in C_b^{\delta,0}$ *with* $0 < \delta \leqslant 1$. *Then for every* $\ell = 0, 1, 2$ *and any* $0 < t \leqslant T$, *we have*

$$
|\nabla_x^\ell p_t(x, x'; y)| \leqslant C_T t^{-(d+\ell)/2} \exp\big(-c_0 |x - x'|^2 / t \big), \tag{3.7}
$$

and for every $x_1, x_2 \in \mathbb{R}^{d_1}$ *and* $0 < \delta' \leqslant \delta$,

$$
\begin{aligned}
|\nabla_x^2 p_t(x_1, x'; y) - \nabla_x^2 p_t(x_2, x'; y)| &\leqslant C_T |x_1 - x_2|^{\delta'} t^{-(d+2+\delta')/2} \\
&\quad \times \Big(\exp\big(-c_0 |x_1 - x'|^2/t \big) + \exp\big(-c_0 |x_2 - x'|^2/t \big) \Big),
\end{aligned} \tag{3.8}
$$

where $C_T, c_0 > 0$ *are constants independent of* y.

If we further assume $(\mathbf{H}b)$ *holds, then for any* $k, j \in \mathbb{R}_+$, *there exists a constant* $m > 0$ *such that for all* $t \geqslant 1$, $x, x' \in \mathbb{R}^{d_1}$ *and* $y \in \mathbb{R}^{d_2}$,

$$
|p_t(x, x'; y)| \leqslant C_1 \frac{1 + |x|^m}{(1 + |x'|^j)}, \tag{3.9}
$$

and for $\ell = 1, 2$,

$$
|\nabla_x^\ell p_t(x, x'; y)| \leqslant C_2 \frac{1 + |x|^m}{(1+t)^k (1 + |x'|^j)}. \tag{3.10}
$$

Moreover, the limit

$$
p_\infty(x', y) := \lim_{t \to \infty} p_t(x, x'; y)
$$

exists and is independent of x, and for every $k, j \in \mathbb{R}_+$, there exists a constant $m > 0$ such that for any $y \in \mathbb{R}^{d_2}$,

$$|p_\infty(x', y)| \leqslant \frac{C_3}{1 + |x'|^j}, \tag{3.11}$$

and

$$|p_t(x, x'; y) - p_\infty(x', y)| \leqslant C_4 \frac{1 + |x|^m}{(1+t)^k (1 + |x'|^j)}. \tag{3.12}$$

The above positive constants $C_i (i = 1, \cdots, 4)$ depend only on λ, d_1, d_2 and $\|a\|_{C_b^{\delta,0}}, \|b\|_{C_b^{\delta,0}}$.

To study the regularity of $T_t f$ with respect to the y-variable, we first consider the case where $f(x, y) \equiv g(x)$, i.e., the function f does not depend on the parameter y, and

$$\int_{\mathbb{R}^d} g(x) \mu^y(\mathrm{d}x) \equiv 0, \quad \forall y \in \mathbb{R}^{d_2}. \tag{3.13}$$

To shorten the notation, we write for $\ell = 1, 2$,

$$\frac{\partial^\ell \mathscr{L}_0}{\partial y^\ell}(x, y) := \sum_{i,j} \partial_y^\ell a_{ij}(x, y) \frac{\partial^2}{\partial x_i \partial x_j} + \partial_y^\ell b(x, y) \cdot \nabla_x.$$

We have the following result.

Lemma 3.4. *Let* **(Hσ)**-**(Hb)** *and (3.1) hold. Assume that $a, b \in C_b^{\delta,1}$ and $g \in C_b^\delta$ with $0 < \delta \leqslant 1$. Then we have*

$$\nabla_y T_t g(x, y) = \int_0^t \int_{\mathbb{R}^{d_1}} p_{t-s}(x, z; y) \frac{\partial \mathscr{L}_0}{\partial y}(z, y) T_s g(z, y) \mathrm{d}z \mathrm{d}s. \tag{3.14}$$

Moreover, for any $0 < t \leqslant 2$,

$$|\nabla_y T_t g(x, y)| \leqslant C_0 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta}, \tag{3.15}$$

and for any $k \in \mathbb{R}_+$, there exists a constant $m > 0$ such that for all $t > 2$,

$$|\nabla_y T_t g(x, y)| \leqslant C_0 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} \frac{(1 + |x|^m)}{(1+t)^k}, \tag{3.16}$$

where $C_0 > 0$ is a constant depending only on λ, d_1, d_2 and $\|a\|_{C_b^{\delta,0}}, \|b\|_{C_b^{\delta,0}}$.

Proof. The equality (3.14) has been proved in [32, Theorem 10] under sightly stronger assumptions on the coefficients. Let us show that the right hand side is indeed well-defined under our conditions. In fact, since $a, b \in C_b^{\delta,1}$, the operator $\partial \mathscr{L}_0 / \partial y$ is meaningful. On the other hand, since $g \in C_b^\delta$, we can derive by (3.7) that for $\ell = 1, 2$ and any $0 < s \leqslant 2$,

$$\begin{aligned} \nabla_z^\ell T_s g(z, y) &= \int_{\mathbb{R}^{d_1}} \nabla_z^\ell p_s(z, x'; y) \big[g(x') - g(z) \big] \mathrm{d}x' \\ &\leqslant C_1 \|g\|_{C_b^\delta} \int_{\mathbb{R}^{d_1}} s^{-(d+\ell)/2} \exp\big(- c_0 |z - x'|^2 / s \big) \cdot |x' - z|^\delta \mathrm{d}x' \\ &\leqslant C_1 \|g\|_{C_b^\delta} s^{(\delta - \ell)/2}, \end{aligned} \tag{3.17}$$

and for any $s > 1$, we have by (3.10) that

$$\nabla_z^\ell T_s g(z,y) \leqslant C_1 \int_{\mathbb{R}^{d_1}} \frac{1+|z|^{\hat{m}}}{(1+s)^k(1+|x'|^{\hat{j}})} |g(x')|\mathrm{d}x' \leqslant C_1 \|g\|_{C_b^\delta} \frac{1+|z|^{\hat{m}}}{(1+s)^k}, \tag{3.18}$$

where we choose $\hat{j} > d_1$ in the above inequality, and $C_1 > 0$ is a constant independent of s and y. As a result,

$$\frac{\partial \mathscr{L}_0}{\partial y}(z,y) T_s g(z,y)$$

makes sense, and estimate (3.15) follows directly. Below, we proceed to show (3.16). As a consequence of (3.13), we have (see [32, (28)])

$$\begin{aligned}\nabla_y T_t g(x,y) = &\int_0^t \int_{\mathbb{R}^{d_1}} p_{t-s}(x,z;y) \frac{\partial \mathscr{L}_0}{\partial y}(z,y) T_s g(z,y) \mathrm{d}z\mathrm{d}s \\ &- \int_0^\infty \int_{\mathbb{R}^{d_1}} p_\infty(z;y) \frac{\partial \mathscr{L}_0}{\partial y}(z,y) T_s g(z,y) \mathrm{d}z\mathrm{d}s.\end{aligned}$$

We further write

$$\begin{aligned}\nabla_y T_t g(x,y) = &\int_0^{t/2} \int_{\mathbb{R}^{d_1}} \big[p_{t-s}(x,z;y) - p_\infty(z;y)\big] \frac{\partial \mathscr{L}_0}{\partial y}(z,y) T_s g(z,y) \mathrm{d}z\mathrm{d}s \\ &+ \int_{t/2}^t \int_{\mathbb{R}^{d_1}} p_{t-s}(x,z;y) \frac{\partial \mathscr{L}_0}{\partial y}(z,y) T_s g(z,y) \mathrm{d}z\mathrm{d}s \\ &- \int_{t/2}^\infty \int_{\mathbb{R}^{d_1}} p_\infty(z;y) \frac{\partial \mathscr{L}_0}{\partial y}(z,y) T_s g(z,y) \mathrm{d}z\mathrm{d}s =: \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3.\end{aligned}$$

For the first term, we have by (3.12) that for any $k \in \mathbb{R}_+$,

$$\mathcal{I}_1 \leqslant \left(\int_0^1 + \int_1^{t/2} \right) \int_{\mathbb{R}^{d_1}} C_2 \frac{1+|x|^m}{(1+t-s)^k(1+|z|^j)} \left| \frac{\partial \mathscr{L}_0}{\partial y}(z,y) T_s g(z,y) \right| \mathrm{d}z\mathrm{d}s.$$

Using (3.17) and (3.18) , we can derive that

$$\begin{aligned}\mathcal{I}_1 \leqslant\ & C_2 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} \bigg(\int_0^1 \int_{\mathbb{R}^{d_1}} \frac{1+|x|^m}{(1+t-s)^k(1+|z|^j)} s^{\delta/2-1} \mathrm{d}z\mathrm{d}s \\ &+ \int_1^{t/2} \int_{\mathbb{R}^{d_1}} \frac{1+|x|^m}{(1+t-s)^k(1+|z|^j)} \frac{1+|z|^{\hat{m}}}{(1+s)^k} \mathrm{d}z\mathrm{d}s \bigg) \\ \leqslant\ & C_2 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} \frac{(1+|x|^m)}{(1+t)^k},\end{aligned}$$

where we choose $j > d_1 + \hat{m}$ and k large enough in the last inequality. As for the second term, using (3.9) and (3.10), we have

$$\begin{aligned}\mathcal{I}_2 = &\int_0^{t/2} \int_{\mathbb{R}^{d_1}} p_s(x,z;y) \frac{\partial \mathscr{L}_0}{\partial y}(z,y) T_{t-s} g(z,y) \mathrm{d}z\mathrm{d}s \\ \leqslant\ & C_3 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} \bigg(\int_0^1 \int_{\mathbb{R}^{d_1}} s^{-d/2} \exp\big(-c_0|x-z|^2/s \big)\end{aligned}$$

$$\times \frac{1+|z|^m}{(1+(t-s))^k}\mathrm{d}z\mathrm{d}s + \int_1^{t/2}\int_{\mathbb{R}^{d_1}} \frac{1+|x|^m}{(1+|z|^j)}\frac{1+|z|^{\hat{m}}}{(1+(t-s))^k}\mathrm{d}z\mathrm{d}s\Big)$$
$$\leqslant C_3\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}\frac{(1+|x|^m)}{(1+t)^k}.$$

Finally, we have by (3.11) that

$$\mathcal{I}_3 \leqslant C_4\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}\int_{t/2}^\infty\int_{\mathbb{R}^{d_1}}\frac{1}{1+|z|^j}\frac{1+|z|^{\hat{m}}}{(1+s)^k}\mathrm{d}z\mathrm{d}s$$
$$\leqslant C_4\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}\frac{1}{(1+t)^k} \leqslant C_4\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}\frac{(1+|x|^m)}{(1+t)^k}.$$

The proof is finished. □

3.2. **Second order derivative with respect to y.** We shall need the following regularity result for $\nabla_y T_t g(x,y)$ with respect to x to study the second order derivative of $T_t g$ with respect to the parameter y.

Lemma 3.5. *Let* **(Hσ)**, **(Hb)** *and (3.1) hold. Assume that $a,b\in C_b^{\delta,1}$ and $g\in C_b^\delta$ with $0<\delta\leqslant 1$. Then for every $y\in\mathbb{R}^{d_2}$, we have $\nabla_y T_t g(\cdot,y)\in C^2$. Moreover, for any $0<t\leqslant 2$ and $\ell=1,2$,*

$$|\nabla_x^\ell\nabla_y T_t g(x,y)| \leqslant C_0 t^{(\delta-\ell)/2}\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}, \tag{3.19}$$

and for any $k\in\mathbb{R}_+$, there exists a constant $m>0$ such that for any $t>2$,

$$|\nabla_x^\ell\nabla_y T_t g(x,y)| \leqslant C_0\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}\frac{(1+|x|^m)}{(1+t)^k}, \tag{3.20}$$

where $C_0>0$ is a constant depending only on λ, d_1, d_2 and $\|a\|_{C_b^{\delta,0}}, \|b\|_{C_b^{\delta,0}}$.

Proof. We only prove the estimates (3.19) and (3.20) when $\ell=2$, the case $\ell=1$ can be proved similarly and is easier since it involves less singularities. Recall that

$$\nabla_y T_t g(x,y) = \int_0^t\int_{\mathbb{R}^d} p_{t-s}(x,z;y)\frac{\partial\mathscr{L}_0}{\partial y}(z,y)T_s g(z,y)\mathrm{d}z\mathrm{d}s.$$

By the Hölder assumption on the coefficients and (3.8), it is easy to check that the function

$$z\to\frac{\partial\mathscr{L}_0}{\partial y}(z,y)T_s g(z,y)$$

is δ'-Hölder continuous for any $\delta'<\delta$, i.e., for any $x,z\in\mathbb{R}^{d_1}$ and $s\leqslant 2$, there exists a constant $C_1>0$ such that

$$\Big|\frac{\partial\mathscr{L}_0}{\partial y}(x,y)T_s g(x,y)-\frac{\partial\mathscr{L}_0}{\partial y}(z,y)T_s g(z,y)\Big|$$
$$\leqslant C_1\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}|x-z|^{\delta'}s^{\frac{\delta-\delta'}{2}-1}.$$

Consequently, we can derive as in (3.17) that for any $t \leqslant 2$,

$$\begin{aligned}
|\nabla_x^2 \nabla_y T_t g(x,y)| &\leqslant C_2 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} \int_0^t \int_{\mathbb{R}^d} (t-s)^{-(d+2)/2} \\
&\qquad \times \exp\big(-c_0|z-x'|^2/(t-s)\big) \cdot |x'-z|^{\delta'} s^{\frac{\delta-\delta'}{2}-1} \mathrm{d}x' \mathrm{d}s \\
&\leqslant C_2 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} \int_0^t (t-s)^{\delta'/2-1} s^{\frac{\delta-\delta'}{2}-1} \mathrm{d}s \\
&\leqslant C_2 t^{\delta/2-1} \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta},
\end{aligned}$$

which yields (3.19). For $t > 2$, we write

$$\begin{aligned}
\nabla_x^2 \nabla_y T_t g(x,y) &= \bigg(\int_{t-1}^t + \int_1^{t-1} + \int_0^1 \bigg) \int_{\mathbb{R}^d} \nabla_x^2 p_s(x,z;y) \frac{\partial \mathscr{L}_0}{\partial y}(z,y) T_{t-s} g(z,y) \mathrm{d}z \mathrm{d}s \\
&=: \mathscr{Q}_1 + \mathscr{Q}_2 + \mathscr{Q}_3.
\end{aligned}$$

Note that on $[t-1,t)$, we have $t-s \in (0,1]$. By (3.10) and (3.17), we can get

$$\begin{aligned}
\mathscr{Q}_1 &\leqslant C_3 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} \int_{t-1}^t \int_{\mathbb{R}^d} \frac{1+|x|^m}{(1+s)^k (1+|z|^j)} (t-s)^{\delta/2-1} \mathrm{d}z \mathrm{d}s \\
&\leqslant C_3 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} \frac{(1+|x|^m)}{(1+t)^k}.
\end{aligned}$$

While for the second term, we can use (3.10) and (3.18) to derive that

$$\begin{aligned}
\mathscr{Q}_2 &\leqslant C_4 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} \int_1^{t-1} \int_{\mathbb{R}^d} \frac{1+|x|^m}{(1+s)^k (1+|z|^j)} \frac{1+|z|^{\hat{m}}}{(1+(t-s))^k} \mathrm{d}z \mathrm{d}s \\
&\leqslant C_4 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} \frac{(1+|x|^m)}{(1+t)^k}.
\end{aligned}$$

To control the last term, we first claim that for every $k \in \mathbb{R}_+$, there exist constants $C_5, m > 0$ such that for any $x_1, x_2 \in \mathbb{R}^{d_1}$ and $t \geqslant 1$,

$$\begin{aligned}
\mathscr{J} &:= \frac{\partial \mathscr{L}_0}{\partial y}(x_1,y) T_t g(x_1,y) - \frac{\partial \mathscr{L}_0}{\partial y}(x_2,y) T_t g(x_2,y) \\
&\leqslant C_5 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} |x_1-x_2|^\delta \frac{1+|x_1|^m+|x_2|^m}{(1+t)^k}.
\end{aligned} \tag{3.21}$$

In fact, we can write

$$\begin{aligned}
\mathscr{J} &\leqslant \bigg(\frac{\partial \mathscr{L}_0}{\partial y}(x_1,y) - \frac{\partial \mathscr{L}_0}{\partial y}(x_2,y) \bigg) T_t g(x_1,y) + \frac{\partial \mathscr{L}_0}{\partial y}(x_2,y) \Big(T_t g(x_1,y) - T_t g(x_2,y) \Big) \\
&=: \mathscr{J}_1 + \mathscr{J}_2.
\end{aligned}$$

By the Hölder assumption on the coefficients and (3.18), it is easy to see that

$$\mathscr{J}_1 \leqslant C_6 \Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}} \Big) \|g\|_{C_b^\delta} |x_1-x_2|^\delta \frac{1+|x_1|^m}{(1+t)^k}.$$

On the other hand, we have by (3.8) that

$$
\begin{aligned}
\mathscr{J}_2 &\leqslant C_7\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}\sum_{\ell=1,2}\left|\int_{\mathbb{R}^d}\big[\nabla_x^\ell p_t(x_1,x';y)-\nabla_x^\ell p_t(x_2,x';y)\big]g(x')\mathrm{d}x'\right|\\
&=C_7\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}\sum_{\ell=1,2}\bigg|\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\big[\nabla_x^\ell p_1(x_1,z;y)-\nabla_x^\ell p_1(x_2,z;y)\big]\\
&\qquad\qquad\times\big[p_{t-1}(z,x';y)-p_\infty(x';y)\big]g(x')\mathrm{d}z\mathrm{d}x'\bigg|\\
&\leqslant C_7\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}|x_1-x_2|^\delta\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\Big(\exp(-c_0|x_1-z|^2)\\
&\qquad\qquad+\exp(-c_0|x_2-z|^2)\Big)\times\frac{1+|z|^m}{(1+t)^k(1+|x'|^j)}|g(x')|\mathrm{d}z\mathrm{d}x'\\
&\leqslant C_7\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}|x_1-x_2|^\delta\frac{1+|x_1|^m+|x_2|^m}{(1+t)^k},
\end{aligned}
$$

where in the third inequality we also used (3.12). Thus (3.21) is true. It then follows by the same argument as in (3.17) that

$$
\begin{aligned}
\mathscr{Q}_3 &\leqslant C_8\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}\int_0^1\int_{\mathbb{R}^d}s^{-(d+2)/2}\exp\big(-c_0|x-z|^2/2s\big)\\
&\qquad\qquad\times|x-z|^\delta\frac{1+|z|^m+|x|^m}{(1+(t-s))^k}\mathrm{d}z\mathrm{d}s\\
&\leqslant C_8\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}\int_0^1\frac{1+|x|^m}{(1+(t-s))^k}s^{\delta/2-1}\mathrm{d}s\\
&\leqslant C_8\Big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\Big)\|g\|_{C_b^\delta}\frac{(1+|x|^m)}{(1+t)^k}.
\end{aligned}
$$

The proof is finished. □

We now establish the second order differentiability of T_tg with respect to the y-variable. We have the following result.

Lemma 3.6. *Let* **(H**σ**)**, **(H**b**)** *and (3.1) hold. Assume that* $a,b\in C_b^{\delta,2}$ *and* $g\in C_b^\delta$ *with* $0<\delta\leqslant 1$. *Then we have*

$$
\begin{aligned}
\nabla_y^2T_tg(x,y)&=2\int_0^t\int_{\mathbb{R}^d}p_{t-s}(x,z;y)\frac{\partial\mathscr{L}_0}{\partial y}(z,y)\nabla_yT_sg(z,y)\mathrm{d}z\mathrm{d}s\\
&\quad+\int_0^t\int_{\mathbb{R}^d}p_{t-s}(x,z;y)\frac{\partial^2\mathscr{L}_0}{\partial y^2}(z,y)T_sg(z,y)\mathrm{d}z\mathrm{d}s. \qquad (3.22)
\end{aligned}
$$

Moreover, for any $0<t\leqslant 2$,

$$
|\nabla_y^2T_tg(x,y)|\leqslant C_0\Big[\big(\|a\|_{C_b^{\delta,1}}+\|b\|_{C_b^{\delta,1}}\big)^2+\big(\|a\|_{C_b^{\delta,2}}+\|b\|_{C_b^{\delta,2}}\big)\Big]\|g\|_{C_b^\delta}, \qquad (3.23)
$$

and for any $k \in \mathbb{R}_+$, there exists a constant $m > 0$ such that for every $t > 2$,

$$|\nabla_y^2 T_t g(x,y)| \leqslant C_0\Big[\big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\big)^2 + \big(\|a\|_{C_b^{\delta,2}} + \|b\|_{C_b^{\delta,2}}\big)\Big]\|g\|_{C_b^\delta}\frac{(1+|x|^m)}{(1+t)^k}, \quad (3.24)$$

where $C_0 > 0$ is a constant depending only on λ, d_1, d_2 and $\|a\|_{C_b^{\delta,0}}, \|b\|_{C_b^{\delta,0}}$.

Proof. The formula (3.22) has been proven in [32, formula (34)]. Let us focus on estimates (3.23) and (3.24). In fact, for $t \leqslant 2$, we can use (3.19) and the same argument as in (3.15) to get that

$$|\nabla_y^2 T_t g(x,y)| \leqslant C_0\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)^2\|g\|_{C_b^\delta} + \Big(\|a\|_{C_b^{\delta,2}} + \|b\|_{C_b^{\delta,2}}\Big)\|g\|_{C_b^\delta},$$

which implies (3.23). Now we prove the estimate (3.24). To this end, we write

$$\begin{aligned}
\nabla_y^2 T_t g(x,y) = 2\bigg(&\int_0^t\int_{\mathbb{R}^d} p_{t-s}(x,z;y)\frac{\partial \mathscr{L}_0}{\partial y}(z,y)\nabla_y T_s g(z,y)\mathrm{d}z\mathrm{d}s \\
&- \int_0^\infty\int_{\mathbb{R}^d} p_\infty(z;y)\frac{\partial \mathscr{L}_0}{\partial y}(z,y)\nabla_y T_s g(z,y)\mathrm{d}z\mathrm{d}s\bigg) \\
&+ \bigg(\int_0^t\int_{\mathbb{R}^d} p_{t-s}(x,z;y)\frac{\partial^2 \mathscr{L}_0}{\partial y^2}(z,y)T_s g(z,y)\mathrm{d}z\mathrm{d}s \\
&- \int_0^\infty\int_{\mathbb{R}^d} p_\infty(z;y)\frac{\partial^2 \mathscr{L}_0}{\partial y^2}(z,y)T_s g(z,y)\mathrm{d}z\mathrm{d}s\bigg) =: \mathscr{J}_1 + \mathscr{J}_2.
\end{aligned}$$

Note by our assumption that $a, b \in C_b^{\delta,2}$, the second part $\mathscr{J}_2$ can be controlled in exactly the same way as in the estimate of $\nabla_y T_t g(x,y)$, i.e., we can get

$$\mathscr{J}_2 \leqslant C_1\Big(\|a\|_{C_b^{\delta,2}} + \|b\|_{C_b^{\delta,2}}\Big)\|g\|_{C_b^\delta}\frac{(1+|x|^m)}{(1+t)^k},$$

where $C_1 > 0$ depends only on λ, d_1, d_2 and $\|a\|_{C_b^{\delta,0}}, \|b\|_{C_b^{\delta,0}}$. Below we shall focus on the estimate of $\mathscr{J}_1$. As before, we write

$$\begin{aligned}
\tfrac{1}{2}\mathscr{J}_1 = &\int_0^{t/2}\int_{\mathbb{R}^d}\big[p_{t-s}(x,z;y) - p_\infty(z;y)\big]\frac{\partial \mathscr{L}_0}{\partial y}(z,y)\nabla_y T_s g(z,y)\mathrm{d}z\mathrm{d}s \\
&+ \int_{t/2}^t\int_{\mathbb{R}^d} p_{t-s}(x,z;y)\frac{\partial \mathscr{L}_0}{\partial y}(z,y)\nabla_y T_s g(z,y)\mathrm{d}z\mathrm{d}s \\
&+ \int_{t/2}^\infty\int_{\mathbb{R}^d} p_\infty(z;y)\frac{\partial \mathscr{L}_0}{\partial y}(z,y)\nabla_y T_s g(z,y)\mathrm{d}z\mathrm{d}s =: \mathscr{J}_{11} + \mathscr{J}_{12} + \mathscr{J}_{13}.
\end{aligned}$$

For the first term, we have by (3.12), (3.19)and (3.20) that

$$\begin{aligned}
\mathscr{J}_{11} \leqslant\; &C_1\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)^2\|g\|_{C_b^\delta}\int_0^1\int_{\mathbb{R}^d}\frac{1+|x|^m}{(1+t-s)^k(1+|z|^j)}\mathrm{d}z\mathrm{d}s \\
&+ C_1\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)^2\|g\|_{C_b^\delta}\int_1^{t/2}\int_{\mathbb{R}^d}\frac{1+|x|^m}{(1+t-s)^k(1+|z|^j)}\frac{1+|z|^{\hat{m}}}{(1+s)^k}\mathrm{d}z\mathrm{d}s \\
\leqslant\; &C_1\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)^2\|g\|_{C_b^\delta}\frac{(1+|x|^m)}{(1+t)^k}.
\end{aligned}$$

Using (3.9) and (3.20) again, we can control the second term by

$$\mathcal{J}_{12} \leqslant C\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)^2 \|g\|_{C_b^\delta} \int_0^1 \int_{\mathbb{R}^d} s^{-d/2} \exp\big(-c_0|x-z|^2/2s\big) \frac{1+|z|^m}{(1+(t-s))^k} \mathrm{d}z\mathrm{d}s$$
$$+ C\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)^2 \|g\|_{C_b^\delta} \int_1^{t/2} \int_{\mathbb{R}^d} \frac{1+|x|^m}{(1+|z|^k)} \frac{1+|z|^{\hat{m}}}{(1+(t-s))^k} \mathrm{d}z\mathrm{d}s$$
$$\leqslant C\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)^2 \|g\|_{C_b^\delta} \frac{(1+|x|^m)}{(1+t)^k}.$$

Finally, we have by (3.11) and (3.20) that

$$\mathcal{I}_3 \leqslant C\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)^2 \|g\|_{C_b^\delta} \int_{t/2}^\infty \int_{\mathbb{R}^d} \frac{1}{1+|z|^j} \cdot \frac{1+|z|^{\hat{m}}}{(1+s)^k} \mathrm{d}z\mathrm{d}s$$
$$\leqslant C\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)^2 \|g\|_{C_b^\delta} \frac{(1+|x|^m)}{(1+t)^k}.$$

The proof is finished. □

With the above preparations, we can establish the following regularity of $T_t f$ with respect to the parameter y.

Lemma 3.7. *Let* **(H**σ**)**, **(H**b**)** *and (3.1) hold. Assume* $a, b \in C_b^{\delta,\ell}$ *and* $f \in C_b^{\delta,\ell}$ *with* $0 < \delta \leqslant 1$ *and* $\ell = 1, 2$. *Then we have:*
(i) (Case $\ell = 1$ *and* $0 < t \leqslant 2$*):*

$$|\nabla_y T_t f(x,y)| \leqslant C_0\Big[\big(\|a\|_{C_b^{\delta,0}} + \|b\|_{C_b^{\delta,0}}\big)\|f\|_{C_b^{\delta,1}} + \big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\big)\|f\|_{C_b^{\delta,0}}\Big];$$

(ii) (Case $\ell = 1$ *and* $t > 2$*): for any* $k \in \mathbb{R}_+$, *there exists a constant* $m > 0$ *such that for every* $t > 2$,

$$|\nabla_y T_t f(x,y)| \leqslant C_0\Big[\big(\|a\|_{C_b^{\delta,0}} + \|b\|_{C_b^{\delta,0}}\big)\|f\|_{C_b^{\delta,1}}$$
$$+ \big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\big)\|f\|_{C_b^{\delta,0}}\Big] \frac{(1+|x|^m)}{(1+t)^k};$$

(iii) (Case $\ell = 2$ *and* $0 < t \leqslant 2$*):*

$$|\nabla_y^2 T_t f(x,y)| \leqslant C_0\Big[\big(\|a\|_{C_b^{\delta,0}} + \|b\|_{C_b^{\delta,0}}\big)\|f\|_{C_b^{\delta,2}} + \big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\big)\|f\|_{C_b^{\delta,1}}$$
$$+ \Big(\big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\big)^2 + \big(\|a\|_{C_b^{\delta,2}} + \|b\|_{C_b^{\delta,2}}\big)\Big)\|f\|_{C_b^{\delta,0}}\Big]; \tag{3.25}$$

(iv) (Case $\ell = 2$ *and* $t > 2$*): for any* $k \in \mathbb{R}_+$, *there exists a constant* $m > 0$ *such that for every* $t > 2$,

$$|\nabla_y^2 T_t f(x,y)| \leqslant C_0\Big[\big(\|a\|_{C_b^{\delta,0}} + \|b\|_{C_b^{\delta,0}}\big)\|f\|_{C_b^{\delta,2}} + \big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\big)\|f\|_{C_b^{\delta,1}}$$
$$+ \Big(\big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\big)^2 + \big(\|a\|_{C_b^{\delta,2}} + \|b\|_{C_b^{\delta,2}}\big)\Big)\|f\|_{C_b^{\delta,0}}\Big] \frac{(1+|x|^m)}{(1+t)^k}, \tag{3.26}$$

where $C_0 > 0$ *is a constant depending only on* λ, d_1, d_2 *and* $\|a\|_{C_b^{\delta,0}}, \|b\|_{C_b^{\delta,0}}$.

Proof. We only prove the above estimates when $\ell = 2$. The corresponding estimates for $\ell = 1$ follows by the same arguments. In fact, we have

$$\begin{aligned}\nabla_y^2 T_t f(x,y) &= \sum_{\ell=0}^{2} C_2^\ell \nabla_y^\ell T_t g(x,y)\Big|_{g=\nabla_y^{2-\ell} f}\\ &= T_t g(x,y)\Big|_{g=\nabla_y^2 f} + 2\nabla_y T_t \nabla_y g(x,y)\Big|_{g=\nabla_y f} + \nabla_y^2 T_t g(x,y)\Big|_{g=f}\\ &=: \mathcal{K}_1 + \mathcal{K}_2 + \mathcal{K}_3.\end{aligned}$$

When $t \leqslant 2$, it is obvious that

$$\mathcal{K}_1 \leqslant C_1\Big(\|a\|_{C_b^{\delta,0}} + \|b\|_{C_b^{\delta,0}}\Big)\|\nabla_y^2 f\|_{C_b^{\delta,0}} \leqslant C_1\Big(\|a\|_{C_b^{\delta,0}} + \|b\|_{C_b^{\delta,0}}\Big)\|f\|_{C_b^{\delta,2}}.$$

For the second term, we have by (3.15) that

$$\mathcal{K}_2 \leqslant C_2\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)\|\nabla_y f\|_{C_b^{\delta,0}} \leqslant C_2\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)\|f\|_{C_b^{\delta,1}}.$$

Finally, using (3.23) we can control the third term by

$$\mathcal{K}_3 \leqslant C_3\Big[\big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\big)^2 + \big(\|a\|_{C_b^{\delta,2}} + \|b\|_{C_b^{\delta,2}}\big)\Big]\|f\|_{C_b^{\delta,0}},$$

which in turn yields (3.25). The estimate (3.26) can be proved similarly by replacing (3.15) and (3.23) with (3.16) and (3.24), respectively. The proof is finished. ☐

4. Strong convergence with order $(\alpha \wedge 1)/2$

In this section, we study the strong convergence of the multi-scale system (1.1) to the effective equation (1.2). To this end, we assume that

$$G(t,x,y) \equiv G(t,y),$$

i.e., the diffusion coefficient G in the slow equation does not dependent on the x-variable. Note that in this case, we have

$$\bar{G}(t,y) = G(t,y).$$

We shall always assume $(\mathbf{H}\sigma)$, $(\mathbf{H}_G)$, $(\mathbf{H}b)$ hold, and that the coefficients a and b are Hölder continuous with respect to x uniformly in y, and that the coefficient G is Hölder continuous with respect to y uniformly in t.

4.1. **Zvonkin transform.** Due to the low regularity assumptions on the coefficients of the system (1.1), it is not possible to prove the strong convergence of Y_t^ε to $\bar{Y}_t$ directly. For this reason, we shall use Zvonkin's argument to transform the equations for Y_t^ε and $\bar{Y}_t$ into new ones. Let us first prove the following regularity result for the averaged drift coefficient.

Lemma 4.1. *Assume that $a, b \in C_b^{\delta,\alpha}$ and $F \in C_b^{\alpha/2,\delta,\alpha}$ with $0 < \delta, \alpha \leqslant 1$. Let $\bar{F}$ be defined as in (1.3). Then we have $\bar{F} \in C_b^{\alpha/2,\alpha}$.*

Proof. The $\alpha/2$-Hölder continuity with respect to the t variable follows directly by the definition of $\bar{F}$. Let us prove the Hölder continuous with respect to y. We write for $y_1, y_2 \in \mathbb{R}^{d_2}$

$$\begin{aligned}\bar{F}(t,y_1) - \bar{F}(t,y_2) &= \int_{\mathbb{R}^{d_1}} [F(t,x,y_1) - F(t,x,y_2)]\mu^{y_1}(\mathrm{d}x) \\ &\quad + \int_{\mathbb{R}^{d_1}} F(t,x,y_2)\big[\mu^{y_1}(\mathrm{d}x) - \mu^{y_2}(\mathrm{d}x)\big] =: \mathscr{K}_1 + \mathscr{K}_2.\end{aligned}$$

It is easy to see that there exists a constant $C_1 > 0$ such that

$$\mathscr{K}_1 \leqslant C_1\big(|y_1 - y_2|^\alpha \wedge 1\big).$$

For the second term, by the same argument as in (3.14), we get

$$\begin{aligned}\mathscr{K}_2 &= \int_{\mathbb{R}^{d_1}} F(t,x',y_2)\big[p_\infty(x';y_1)\mathrm{d}x' - p_\infty(x';y_2)\mathrm{d}x'\big] \\ &= \lim_{t\to\infty}\int_0^t\int_{\mathbb{R}^{d_1}} p_{t-s}(x,z;y_2)[\mathscr{L}_0(z,y_1) - \mathscr{L}_0(z,y_2)]\left(\int_{\mathbb{R}^{d_1}} p_s(z,x';y_2)F(t,x',y_2)\mathrm{d}x'\right)\mathrm{d}z\mathrm{d}s \\ &= \int_0^\infty\int_{\mathbb{R}^{d_1}} p_\infty(z;y_2)[\mathscr{L}_0(z,y_1) - \mathscr{L}_0(z,y_2)]\left(\int_{\mathbb{R}^{d_1}} p_s(z,x';y_2)F(t,x',y_2)\mathrm{d}x'\right)\mathrm{d}z\mathrm{d}s.\end{aligned}$$

Thus, we have by (3.11), (3.17) and (3.18) that

$$\begin{aligned}\mathscr{K}_2 &\leqslant C_2\big(|y_1 - y_2|^\alpha \wedge 1\big)\bigg(\int_0^2\int_{\mathbb{R}^{d_1}} \frac{1}{(1+|z|^j)} s^{\delta/2-1}\mathrm{d}z\mathrm{d}s \\ &\quad + \int_2^\infty\int_{\mathbb{R}^{d_1}} \frac{1}{(1+|z|^j)}\frac{(1+|z|)^m}{(1+s)^k}\mathrm{d}z\mathrm{d}s\bigg) \leqslant C_2\big(|y_1 - y_2|^\alpha \wedge 1\big),\end{aligned}$$

where $C_2 > 0$ is a constant. The proof is finished. □

Below, we shall fix a $T > 0$ to be sufficiently small. Recall that $\bar{\mathscr{L}}$ is defined by (1.9). Consider the following backward PDE in $\mathbb{R}^{d_2}$:

$$\begin{cases} \partial_t v(t,y) + \bar{\mathscr{L}} v(t,y) + \bar{F}(t,y) = 0, \quad t \in [0,T), \\ v(T,y) = 0. \end{cases} \tag{4.1}$$

Under our assumptions on the coefficients and by Lemma 4.1, it is well known that there exits a unique solution $v \in L^\infty\big([0,T]; C_b^{2+\alpha}(\mathbb{R}^{d_2})\big) \cap C_b^{1+\alpha/2}\big([0,T]; L^\infty(\mathbb{R}^{d_2})\big)$ for equation (4.1). Moreover, we can choose T small enough so that for any $0 < t < T$,

$$|\nabla_y v(t,y)| \leqslant 1/2, \quad \forall y \in \mathbb{R}^{d_2}.$$

Define the transformed function by

$$\Phi(t,y) := y + v(t,y).$$

Then, the map $y \to \Phi(t,y)$ forms a C^1-diffeomorphism and

$$1/2 \leqslant \|\nabla_y \Phi\|_\infty \leqslant 3. \tag{4.2}$$

Now, let us define the new processes by

$$\bar{V}_t := \Phi(t,\bar{Y}_t) \quad \text{and} \quad V_t^\varepsilon := \Phi(t,Y_t^\varepsilon). \tag{4.3}$$

We have the following result.

Lemma 4.2. *Let $\bar{V}_t$ and V_t^ε be defined by (4.3). Then we have*

$$\mathrm{d}\bar{V}_t = G(t,\bar{Y}_t)\nabla_y\Phi(t,\bar{Y}_t)\mathrm{d}W_t^2, \quad V_0 = \Phi(0,y) \tag{4.4}$$

and

$$\begin{aligned}\mathrm{d}V_t^\varepsilon = &\big[F(t,X_t^\varepsilon,Y_t^\varepsilon) - \bar{F}(t,Y_t^\varepsilon)\big]\nabla_y\Phi(t,Y_t^\varepsilon)\mathrm{d}t\\ &+ G(t,Y_t^\varepsilon)\nabla_y\Phi(t,Y_t^\varepsilon)\mathrm{d}W_t^2, \quad V_0^\varepsilon = \Phi(0,y).\end{aligned} \tag{4.5}$$

Proof. Using Itô's formula, we have

$$\begin{aligned}v(t,Y_t^\varepsilon) &= v(0,y) + \int_0^t \big(\partial_s + \mathscr{L}_1\big)v(s,Y_s^\varepsilon)\mathrm{d}s + \int_0^t G(s,Y_s^\varepsilon)\nabla_y v(s,Y_s^\varepsilon)\mathrm{d}W_s^2\\
&= v(0,y) + \int_0^t \big(\partial_s + \bar{\mathscr{L}}\big)v(s,Y_s^\varepsilon)\mathrm{d}s + \int_0^t G(s,Y_s^\varepsilon)\nabla_y v(s,Y_s^\varepsilon)\mathrm{d}W_s^2\\
&\quad + \int_0^t \big[F(s,X_s^\varepsilon,Y_s^\varepsilon) - \bar{F}(s,Y_s^\varepsilon)\big]\nabla_y v(s,Y_s^\varepsilon)\mathrm{d}s\\
&= v(0,y) - \int_0^t \bar{F}(s,Y_s^\varepsilon)\mathrm{d}s + \int_0^t G(s,Y_s^\varepsilon)\nabla_y v(s,Y_s^\varepsilon)\mathrm{d}W_s^2\\
&\quad + \int_0^t \big[F(s,X_s^\varepsilon,Y_s^\varepsilon) - \bar{F}(s,Y_s^\varepsilon)\big]\nabla_y v(s,Y_s^\varepsilon)\mathrm{d}s,\end{aligned}$$

where in the last equality we used (4.1). This together with the equation for Y_t^ε yields (4.5). The proof of (4.4) is easier and follows by the same argument. ☐

4.2. **Proof of Theorem 2.1.** We first prepare the following mollifying approximation result. For simplification, let us set

$$\hat{F}(t,x,y) := \big[F(t,x,y) - \bar{F}(t,y)\big]\nabla_y\Phi(t,y). \tag{4.6}$$

Let $\rho_1 : \mathbb{R} \to [0,1]$ and $\rho_2 : \mathbb{R}^{d_2} \to [0,1]$ be two smooth radial convolution kernel functions such that $\int_{\mathbb{R}}\rho_1(r)\mathrm{d}r = \int_{\mathbb{R}^{d_2}}\rho_2(y)\mathrm{d}y = 1$, and for any $k \geqslant 1$, $|\nabla^k\rho_1| \leqslant C_k\rho_1(x)$ and $|\nabla^k\rho_2| \leqslant C_k\rho_2(x)$, where $C_k > 0$ are constants. For every $n \in \mathbb{N}^*$, set

$$\rho_1^n(r) := n^2\rho_1(n^2 r) \quad \text{and} \quad \rho_2^n(y) := n^{d_2}\rho_2(ny).$$

We define the mollifying approximations of $\hat{F}$ by

$$F_n(t,x,y) := \int_{\mathbb{R}^{d_2+1}} \hat{F}(t-s,x,y-z)\rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s. \tag{4.7}$$

Similarly, we define the mollifying approximations of a, b by

$$a_n(x,y) := \int_{\mathbb{R}^{d_2}} a(x,y-z)\rho_2^n(z)\mathrm{d}z, \quad b_n(x,y) := \int_{\mathbb{R}^{d_2}} b(x,y-z)\rho_2^n(z)\mathrm{d}z. \tag{4.8}$$

Let $\bar{F}_n$ be the average of F_n with respect to $\mu^y(\mathrm{d}x)$, i.e.,

$$\bar{F}_n := \int_{\mathbb{R}^{d_1}} F_n(t,x,y)\mu^y(\mathrm{d}x).$$

We have the following easy result, which will play important role below.

Lemma 4.3. *Assume that $a, b \in C_b^{\delta,\alpha}$ and $F \in C_b^{\alpha/2,\delta,\alpha}$ with $0 < \delta, \alpha \leqslant 1$. Then we have*

$$\|\hat{F} - F_n\|_\infty + \|a - a_n\|_\infty + \|b - b_n\|_\infty + \|\bar{F}_n\|_\infty \leqslant C_0 n^{-\alpha}, \tag{4.9}$$

and

$$\|F_n\|_{C_b^{1,\delta,\alpha}} + \|F_n\|_{C_b^{\alpha,\delta,2}} + \|\bar{F}_n\|_{C_b^{1,\alpha}} + \|\bar{F}_n\|_{C_b^{\alpha,2}} + \|a_n\|_{C_b^{\delta,2}} + \|b_n\|_{C_b^{\delta,2}} \leqslant C_0 n^{2-\alpha}, \tag{4.10}$$

where $C_0 > 0$ is a constant independent of n.

Proof. According to Lemma 4.1, it is easy to check that $\hat{F} \in C_b^{\alpha/2,\delta,\alpha}$. By the definition of $\hat{F}_n$, we have

$$\begin{aligned}|\hat{F}(t,x,y) - F_n(t,x,y)| &\leqslant \int_{\mathbb{R}^{d_2+1}} \big|\hat{F}(t-s,x,y-z) - \hat{F}(t,x,y)\big| \cdot \rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s \\ &\leqslant C_1 \int_{\mathbb{R}^{d_2+1}} \big(s^{\alpha/2} + |z|^\alpha\big) \cdot \rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s \leqslant C_1 n^{-\alpha}.\end{aligned}$$

Furthermore, we have

$$\begin{aligned}|\partial_t F_n(t,x,y)| &\leqslant \int_{\mathbb{R}^{d_2+1}} \big|\hat{F}(t-s,x,y-z) - \hat{F}(t,x,y-z)\big| \cdot |\rho_2^n(z)|\partial_s\rho_1^n(s)\mathrm{d}z\mathrm{d}s \\ &\leqslant C_2 n^2 \int_{\mathbb{R}^{d_2+1}} s^{\alpha/2}\rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s \leqslant C_2 n^{2-\alpha},\end{aligned}$$

and

$$\begin{aligned}|\nabla_y^2 F_n(t,x,y)| &\leqslant \int_{\mathbb{R}^{d_2+1}} \big|\hat{F}(t-s,x,y-z) - \hat{F}(t-s,x,y)\big| \cdot |\nabla_z^2\rho_2^n(z)|\rho_1^n(s)\mathrm{d}z\mathrm{d}s \\ &\leqslant C_2 n^2 \int_{\mathbb{R}^{d_2+1}} |z|^\alpha \rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s \leqslant C_2 n^{2-\alpha}.\end{aligned}$$

The other estimates can be proved similarly. □

Now, we are in the position to give:

Proof of Theorem 2.1. Let us first assume that $T > 0$ is sufficiently small so that (4.2) holds. As a result, we have for any $t \in [0,T]$,

$$\mathbb{E}\big|Y_t^\varepsilon - \bar{Y}_t\big|^2 \leqslant 2\mathbb{E}\big|V_t^\varepsilon - \bar{V}_t\big|^2. \tag{4.11}$$

Hence, we shall focus on the convergence of V_t^ε to $\bar{V}_t$. Recall the definition of $\hat{F}$ by (4.6), and let F_n be given by (4.7), $\hat{F}_n := F_n - \bar{F}_n$. According to (4.4) and (4.5), we write

$$\begin{aligned}V_t^\varepsilon - \bar{V}_t &= \int_0^t \big[G(s,Y_s^\varepsilon)\nabla_y\Phi(s,Y_s^\varepsilon) - G(s,\bar{Y}_s)\nabla_y\Phi(s,\bar{Y}_s)\big]\mathrm{d}W_s^2 \\ &\quad + \int_0^t \big[\hat{F}(s,X_s^\varepsilon,Y_s^\varepsilon) - \hat{F}_n(s,X_s^\varepsilon,Y_s^\varepsilon)\big]\mathrm{d}s + \int_0^t \hat{F}_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s.\end{aligned}$$

Thus, taking expectation and using Burkholder-Davis-Gundy's inequality we can get that there exists a $C_0 > 0$ such that

$$\mathbb{E}|V_t^\varepsilon - V_t|^2 \leqslant C_0\mathbb{E}\left(\int_0^t \big|G(s,Y_s^\varepsilon)\nabla_y\Phi(s,Y_s^\varepsilon) - G(s,\bar{Y}_s)\nabla_y\Phi(s,\bar{Y}_s)\big|^2\mathrm{d}s\right)$$

$$
\begin{aligned}
&+ C_0\mathbb{E}\left|\int_0^t \big[\hat{F}(s,X_s^\varepsilon,Y_s^\varepsilon) - \hat{F}_n(s,X_s^\varepsilon,Y_s^\varepsilon)\big]\mathrm{d}s\right|^2 \\
&+ C_0\mathbb{E}\left|\int_0^t \hat{F}_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right|^2 =: \mathscr{Q}_1(t,\varepsilon) + \mathscr{Q}_2(t,\varepsilon) + \mathscr{Q}_3(t,\varepsilon).
\end{aligned}
$$

Below, we divide the proof into three steps to control each term on the right hand side separately.

Step 1 (Control of $\mathscr{Q}_1(t,\varepsilon)$). Note that the function

$$
y \to G(t,\cdot)\nabla_y\Phi(t,\cdot) \in C_b^1(\mathbb{R}^{d_2}).
$$

As a result, we easily have that

$$
\mathscr{Q}_1(t,\varepsilon) \leqslant C_1\mathbb{E}\left(\int_0^t |Y_s^\varepsilon - \bar{Y}_s|^2\mathrm{d}s\right), \tag{4.12}
$$

where $C_1 > 0$ is a constant independent of ε.

Step 2 (Control of $\mathscr{Q}_2(t,\varepsilon)$). The estimate of this term follows by an easy consequence of (4.9), which in turn yields that

$$
\mathscr{Q}_2(t,\varepsilon) \leqslant C_2\|\hat{F} - \hat{F}_n\|_\infty^2 \leqslant C_2 n^{-2\alpha}, \tag{4.13}
$$

where C_2 is a positive constant independent of n and ε.

Step 3 (Control of $\mathscr{Q}_3(t,\varepsilon)$). We use the technique of the Poisson equation to control the third part. Let a_n, b_n be defined by (4.8), and denote by $\mathscr{L}_0^n(x,y)$ the operator $\mathscr{L}_0(x,y)$ with coefficients a, b replaced by a_n, b_n, i.e.,

$$
\mathscr{L}_0^n(x,y) := \sum_{i,j} a_n^{ij}(x,y)\frac{\partial^2}{\partial x_i \partial x_j} + b_n(x,y)\cdot\nabla_x. \tag{4.14}
$$

Let Ψ_n be the solution to the following Poisson equation in $\mathbb{R}^{d_1}$:

$$
\mathscr{L}_0^n(x,y)\Psi_n(t,x,y) = \hat{F}_n(t,x,y),
$$

where $(t,y) \in \mathbb{R}_+ \times \mathbb{R}^{d_2}$ are viewed as parameters. Note that $\hat{F}_n$ satisfies the centering condition (3.1). Thus, according to Theorem 3.1, we can use Itô's formula to get that for any $t > 0$,

$$
\begin{aligned}
\Psi_n(t,X_t^\varepsilon,Y_t^\varepsilon) = \Psi_n(0,x,y) &+ \int_0^t \big(\partial_s + \mathscr{L}_1\big)\Psi_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s \\
&+ \int_0^t \frac{1}{\varepsilon}\mathscr{L}_0\Psi_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s + \frac{1}{\sqrt{\varepsilon}}M_t^1 + M_t^2,
\end{aligned}
$$

where $\mathscr{L}_1$ is given by (1.6), and for $i = 1, 2$, M_t^i are martingales defined by

$$
M_t^1 := \int_0^t \nabla_x\Psi_n(s,X_s^\varepsilon,Y_s^\varepsilon)\sigma(s,X_s^\varepsilon)\mathrm{d}W_s^1,
$$

and

$$
M_t^2 := \int_0^t \nabla_y\Psi_n(s,X_s^\varepsilon,Y_s^\varepsilon)G(s,Y_s^\varepsilon)\mathrm{d}W_s^2.
$$

This in turn yields that

$$\int_0^t \hat{F}_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s = \varepsilon\Psi_n(t,X_t^\varepsilon,Y_t^\varepsilon) - \varepsilon\Psi_n(0,x,y) - \sqrt{\varepsilon}M_t^1 - \varepsilon M_t^2$$
$$+ \int_0^t \big[b_n(X_s^\varepsilon,Y_s^\varepsilon) - b(X_s^\varepsilon,Y_s^\varepsilon)\big]\nabla_x\Psi_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s$$
$$+ \int_0^t \big[a_n(X_s^\varepsilon,Y_s^\varepsilon) - a(X_s^\varepsilon,Y_s^\varepsilon)\big]\nabla_x^2\Psi_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s$$
$$- \varepsilon\int_0^t \big(\partial_s + \mathscr{L}_1\big)\Psi_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s.$$

Taking this back into the definition of $\mathscr{Q}_3(t,\varepsilon)$ and by (3.4), we have that there exists a constant $m>0$ such that

$$\mathscr{Q}_3(t,\varepsilon) \leqslant C_3\Bigg[\varepsilon^2\mathbb{E}(1+|X_t^\varepsilon|^{2m}) + \varepsilon\mathbb{E}|M_t^1|^2 + \varepsilon^2\mathbb{E}|M_t^2|^2$$
$$+\mathbb{E}\left(\int_0^t\Big(\big|b_n(X_s^\varepsilon,Y_s^\varepsilon) - b(X_s^\varepsilon,Y_s^\varepsilon)\big|^2 + \big|a_n(X_s^\varepsilon,Y_s^\varepsilon) - a(X_s^\varepsilon,Y_s^\varepsilon)\big|^2\Big)(1+|X_s^\varepsilon|^{2m})\mathrm{d}s\right)$$
$$+\varepsilon^2\mathbb{E}\left|\int_0^t \big(\partial_s + \mathscr{L}_y\big)\Psi_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right|^2\Bigg] =: \mathscr{Q}_{31}(t,\varepsilon) + \mathscr{Q}_{32}(t,\varepsilon) + \mathscr{Q}_{33}(t,\varepsilon).$$

Note that the assumptions **(Hσ)** and **(Hb)** hold uniformly in y. Hence, it follows by [37, Lemma 1] (see also [32, Lemma 2]) that for any $k>0$,

$$\mathbb{E}|X_t^\varepsilon|^k \leqslant C(1+|x|^k), \tag{4.15}$$

where C is a positive constant independent of ε. As a result, we can control the first term by (3.5) and (4.10) that

$$\mathscr{Q}_{31}(t,\varepsilon) \leqslant C_4\Big(\varepsilon + \varepsilon^2\mathbb{E}\int_0^t \|\nabla_y\Psi(\cdot,X_s^\varepsilon,\cdot)\|_\infty^2\mathrm{d}s\Big)$$
$$\leqslant C_4\Big(\varepsilon + \varepsilon^2\Big(\|a_n\|_{C_b^{\delta,1}} + \|b_n\|_{C_b^{\delta,1}} + \|\hat{F}_n\|_{C_b^{0,\delta,1}}\Big)\mathbb{E}\left(\int_0^t(1+|X_s^\varepsilon|^{2m})\mathrm{d}s\right)$$
$$\leqslant C_4(1+|x|^{2m})\Big(\varepsilon + \varepsilon^2 n^{2(1-\alpha)}\Big).$$

For the second term, by (4.9) and (4.15), it is easy to see that

$$\mathscr{Q}_{32}(t,\varepsilon) \leqslant C_5\Big(\|b_n - b\|_\infty^2 + \|a_n - a\|_\infty^2\Big)\mathbb{E}\left(\int_0^t(1+|X_s^\varepsilon|^{2m})\mathrm{d}s\right) \leqslant C_5(1+|x|^{2m})n^{-2\alpha}.$$

To estimate the last part, we first note that by (3.2) and viewing t as a parameter,we have that for any $s>0$, $x\in\mathbb{R}^{d_1}$ and $y\in\mathbb{R}^{d_2}$,

$$|\partial_s\Psi_n(s,x,y)| \leqslant C_6\|\partial_s\hat{F}_n\|_{C_b^{0,\delta,0}}(1+|x|^m) \leqslant C_6 n^{2-\alpha}(1+|x|^m),$$

where the last inequality follows by (4.10). On the other hand, by reviewing y as a parameter, we have by (3.6) and (4.10) that

$$\|\mathscr{L}_y\Psi_n(\cdot,x,\cdot)\|_\infty \leqslant C_6\Big[\Big(\|a_n\|_{C_b^{\delta,0}} + \|b_n\|_{C_b^{\delta,0}}\Big)\|\hat{F}_n\|_{C_b^{0,\delta,2}} + \Big(\|a_n\|_{C_b^{\delta,1}} + \|b_n\|_{C_b^{\delta,1}}\Big)\|\hat{F}_n\|_{C_b^{0,\delta,1}}$$

$$+\Big(\big(\|a_n\|_{C_b^{\delta,1}}+\|b_n\|_{C_b^{\delta,1}}\big)^2+\big(\|a_n\|_{C_b^{\delta,2}}+\|b_n\|_{C_b^{\delta,2}}\big)\Big)\|\hat{F}_n\|_{C_b^{0,\delta,0}}\Big](1+|x|^m)$$
$$\leqslant C_6\Big(n^{2-\alpha}+n^{2-2\alpha}\Big)(1+|x|^m)\leqslant C_6 n^{2-\alpha}(1+|x|^m).$$

As a result, we have

$$\mathscr{Q}_{33}(t,\varepsilon)\leqslant C_7(1+|x|^{2m})\varepsilon^2 n^{2(2-\alpha)}.$$

Combing the above estimates, we get

$$\mathscr{Q}_3(t,\varepsilon)\leqslant C_8(1+|x|^{2m})\big(\varepsilon+n^{-2\alpha}+\varepsilon^2 n^{2(2-\alpha)}\big).$$

Now, in view of (4.11), (4.12) and (4.13), we arrive at

$$\mathbb{E}\big|Y_t^\varepsilon-\bar{Y}_t\big|^2\leqslant C_9\mathbb{E}\left(\int_0^t|Y_s^\varepsilon-\bar{Y}_s|^2\mathrm{d}s\right)+C_9(1+|x|^{2m})\Big(n^{-2\alpha}+\varepsilon+\varepsilon^2 n^{2(2-\alpha)}\Big).$$

Taking $n=\varepsilon^{-1/2}$ (which is optimal for minimizing the right hand side of the above inequality), we get

$$\mathbb{E}\big|Y_t^\varepsilon-\bar{Y}_t\big|^2\leqslant C_9\mathbb{E}\left(\int_0^t|Y_s^\varepsilon-\bar{Y}_s|^2\mathrm{d}s\right)+C_9(1+|x|^{2m})\varepsilon^{\alpha\wedge 1},$$

which in turn yields by Gronwall's inequality that

$$\mathbb{E}\big|Y_t^\varepsilon-\bar{Y}_t\big|^2\leqslant C_T(1+|x|^{2m})\varepsilon^{\alpha\wedge 1}.$$

For general $T>0$ and in view of (4.15), the result can be proved by induction and analogous arguments. So, the whole proof is finished. □

4.3. **Proof of Theorem 2.3.** We use Theorem 2.1 and the radial truncation technique to prove Theorem 2.3.

Proof of Theorem 2.3. For each $n\in\mathbb{N}$, define the new coefficients by

$$b_n(x,y):=\begin{cases}b(x,y), & |y|\leqslant n,\\ b(x,ny/|y|) & |y|>n,\end{cases},\quad \sigma_n(x,y):=\begin{cases}\sigma(x,y), & |y|\leqslant n,\\ \sigma(x,ny/|y|) & |y|>n,\end{cases}$$

and

$$F_n(t,x,y):=\begin{cases}F(t,x,y), & |y|\leqslant n,\\ F(t,x,ny/|y|) & |y|>n,\end{cases},\quad G_n(t,y):=\begin{cases}G(t,y), & |y|\leqslant n,\\ G(t,ny/|y|) & |y|>n.\end{cases}$$

It is easy to check that b_n,σ_n,F_n,G_n satisfy the conditions in Theorem 2.1. Let $(X_t^{n,\varepsilon},Y_t^{n,\varepsilon})$ be the solution to SDE (1.1) with coefficients b,σ,F,G replaced by b_n,σ_n,F_n,G_n. Then for any $T>0$, we have by Theorem 2.1 that

$$\sup_{t\in[0,T]}\mathbb{E}\big|Y_t^{n,\varepsilon}-\bar{Y}_t^n\big|^2\to 0\quad\text{as}\quad\varepsilon\to 0,$$

where $\bar{Y}_t^n$ is the solution of the following new averaged equation:

$$\mathrm{d}\bar{Y}_t^n=\bar{F}_n(t,\bar{Y}_t^n)\mathrm{d}t+G_n(t,\bar{Y}_t^n)\mathrm{d}W_t^2,\quad \bar{Y}_0^n=y.$$

Here, $\bar{F}_n(t,y):=\int_{\mathbb{R}^{d_1}}F_n(t,x,y)\mu_n^y(dx)$, and $\mu_n^y(\mathrm{d}x)$ is the unique invariant measure of the transition semigroup of the following frozen equation:

$$\mathrm{d}X_t^{n,y}=b_n(X_t^{n,y},y)\mathrm{d}t+\sigma_n(X_t^{n,y},y)\mathrm{d}W_t^1,\quad X_0^{n,y}=x.$$

For every $\varepsilon > 0$, define the stopping time by

$$\tau_n^\varepsilon := \inf\{t \geqslant 0 : |Y_t^\varepsilon| + |\bar{Y}_t| \geqslant n\}.$$

Then, by the construction of the new coefficients and the uniqueness of the strong solution to SDE (1.1), it holds

$$Y_t^\varepsilon = Y_t^{n,\varepsilon}, \quad \forall t \in [0, \tau_n^\varepsilon].$$

On the other hand, note that for every $|y| \leqslant n$, we also have $\mu_n^y(dx) = \mu^y(dx)$. This implies that for $|y| \leqslant n$,

$$\bar{F}_n(t,y) = \int_{\mathbb{R}^{d_1}} F_n(t,x,y)\mu_n^y(dx) = \int_{\mathbb{R}^{d_1}} F(t,x,y)\mu^y(dx) = \bar{F}(t,y),$$

which together with the uniqueness of the strong solution to SDE (1.2) means

$$\bar{Y}_t^n = \bar{Y}_t, \quad \forall t \in [0, \tau_n^\varepsilon].$$

As a result, we can deduce that for some $\beta > 2$ and $C > 0$,

$$\begin{aligned}
\sup_{t\in[0,T]} \mathbb{E}\big|Y_t^\varepsilon - \bar{Y}_t\big|^2 &\leqslant \sup_{t\in[0,T]} \mathbb{E}\big(|Y_t^\varepsilon - \bar{Y}_t|^2 \cdot 1_{\{t\leqslant\tau_n^\varepsilon\}}\big) + \sup_{t\in[0,T]} \mathbb{E}\big(|Y_t^\varepsilon - \bar{Y}_t|^2 \cdot 1_{\{t>\tau_n^\varepsilon\}}\big) \\
&\leqslant \sup_{t\in[0,T]} \mathbb{E}|Y_t^{n,\varepsilon} - \bar{Y}_t^n|^2 + C\left[\sup_{t\in[0,T]} \mathbb{E}\big(|Y_t^\varepsilon|^\beta + |\bar{Y}_t|^\beta\big)\right]^{2/\beta} [\mathbb{P}(T > \tau_n^\varepsilon)]^{(\beta-2)/\beta} \\
&\leqslant \sup_{t\in[0,T]} \mathbb{E}|Y_t^{n,\varepsilon} - \bar{Y}_t^n|^2 + C/n^{\beta-2},
\end{aligned}$$

where the last inequality follows by Chebyshev's inequality and condition $(\mathbf{H}^M)$. Letting $\varepsilon \to 0$ first and then $n \to \infty$, we can get the desired result. $\square$

5. Weak convergence with order $(\alpha/2) \wedge 1$

Now we study the weak convergence of the multi-scale system (1.1) to the effective system (1.2) in the fully coupled case, i.e., the diffusion coefficient $G(t,x,y)$ in the slow part also depends on the fast term. We first prove the following regularity result for the averaged coefficients.

Lemma 5.1. *Assume that $a, b \in C_b^{\delta,\alpha}$ and $F, G \in C_b^{\alpha/2,\delta,\alpha}$ with $0 < \delta, \alpha \leqslant 2$. Let $\bar{F}$ and $\bar{G}$ be defined by (1.3). Then we have $\bar{F}, \bar{H} \in C_b^{\alpha/2,\alpha}$.*

Proof. We only sketch the proof of the regularity for $\bar{F}$. Note that when $0 < \alpha \leqslant 1$, the conclusion has been proven in Lemma 4.1. Let us focus on the case $1 < \alpha \leqslant 2$. We write for $y_1, y_2 \in \mathbb{R}^{d_2}$

$$\begin{aligned}
\nabla_y \bar{F}(t,y_1) - \nabla_y \bar{F}(t,y_2) &= \int_{\mathbb{R}^{d_1}} [\nabla_y F(t,x,y_1) - \nabla_y F(t,x,y_2)]\mu^{y_1}(\mathrm{d}x) \\
&\quad + \int_{\mathbb{R}^{d_1}} F(t,x,y_2)\nabla_y\big[\mu^{y_1}(\mathrm{d}x) - \mu^{y_2}(\mathrm{d}x)\big] =: \tilde{\mathscr{K}_1} + \tilde{\mathscr{K}_2}.
\end{aligned}$$

It is easy to see that there exists a constant $C_1 > 0$ such that

$$\tilde{\mathscr{K}_1} \leqslant C_1\big(|y_1 - y_2|^{\alpha-1} \wedge 1\big).$$

For the second term, by (3.14) we write

$$
\begin{aligned}
\tilde{\mathscr{K}}_2 &= \int_{\mathbb{R}^{d_1}} F(t,x',y_2)\big[\nabla_y p_\infty(x';y_1) - \nabla_y p_\infty(x';y_2)\big]\mathrm{d}x' \\
&= \lim_{t\to\infty}\int_0^t\int_{\mathbb{R}^{d_1}} p_{t-s}(x,z;y_2)\left[\frac{\partial\mathscr{L}_0(z,y_1)}{\partial y} - \frac{\partial\mathscr{L}_0(z,y_2)}{\partial y}\right] \\
&\qquad\qquad\times\left(\int_{\mathbb{R}^{d_1}} p_s(z,x';y_2)F(t,x',y_2)\mathrm{d}x'\right)\mathrm{d}z\mathrm{d}s \\
&= \int_0^\infty\int_{\mathbb{R}^{d_1}} p_\infty(z;y_2)\left[\frac{\partial\mathscr{L}_0(z,y_1)}{\partial y} - \frac{\partial\mathscr{L}_0(z,y_2)}{\partial y}\right]\left(\int_{\mathbb{R}^{d_1}} p_s(z,x';y_2)F(t,x',y_2)\mathrm{d}x'\right)\mathrm{d}z\mathrm{d}s.
\end{aligned}
$$

Then, the desired estimates follow by exactly the same arguments as in the proof of Lemma 4.1. We omit the details. □

Recall that $\bar{\mathscr{L}}$ is defined by (1.9). Given a function $\varphi \in C_b^{2+\alpha}$ and $T>0$, we consider the following Cauchy problem:

$$
\left\{\begin{array}{l}
\partial_t\hat{u}(t,y) - \bar{\mathscr{L}}\hat{u}(t,y) = 0, \quad t\in[0,T), \\
\hat{u}(0,y) = \varphi(y).
\end{array}\right. \tag{5.1}
$$

It is known that there exists a unique solution $\hat{u}$ to (5.1) which is given by

$$
\hat{u}(t,y) = \mathbb{E}\varphi(\bar{Y}_t(y)).
$$

Moreover, we have $\nabla_y^2\hat{u} \in C_b^{\alpha/2,\alpha}$, see e.g. [28, Chapter IV, Section 5]. Set

$$
\tilde{u}(t,y) := \hat{u}(T-t,y), \quad t\in[0,T].
$$

By Itô's formula, we deduce that

$$
\tilde{u}(T,Y_T^\varepsilon) = \tilde{u}(0,y) + \int_0^T \partial_s\tilde{u}(s,Y_s^\varepsilon) + \mathscr{L}_1(X_s^\varepsilon,Y_s^\varepsilon)\tilde{u}(s,Y_s^\varepsilon)\mathrm{d}s + \tilde{M}_t,
$$

where $\tilde{M}_t$ is a martingale given by

$$
\tilde{M}_t := \int_0^t G(s,X_s^\varepsilon,Y_s^\varepsilon)\nabla_y\tilde{u}(s,Y_s^\varepsilon)\mathrm{d}W_s^2.
$$

Note that

$$
\tilde{u}(T,Y_T^\varepsilon) = \hat{u}(0,Y_T^\varepsilon) = \varphi(Y_T^\varepsilon), \quad \text{and} \quad \tilde{u}(0,y) = \hat{u}(T,y) = \mathbb{E}[\varphi(Y_T)],
$$

and

$$
\begin{aligned}
&\partial_s\tilde{u}(s,Y_s^\varepsilon) + \mathscr{L}_1\big(X_s^\varepsilon,Y_s^\varepsilon)\tilde{u}(s,Y_s^\varepsilon) = \mathscr{L}_1(X_s^\varepsilon,Y_s^\varepsilon)\tilde{u}(s,Y_s^\varepsilon) - \bar{\mathscr{L}}_y\tilde{u}(s,Y_s^\varepsilon) \\
&\ = [H(s,X_s^\varepsilon,Y_s^\varepsilon) - \bar{H}(s,Y_s^\varepsilon)]\nabla_y^2\tilde{u}(s,Y_s^\varepsilon) + [F(s,X_s^\varepsilon,Y_s^\varepsilon) - \bar{F}(s,Y_s^\varepsilon)]\nabla_y\tilde{u}(s,Y_s^\varepsilon).
\end{aligned}
$$

We thus get

$$
\begin{aligned}
\mathbb{E}[\varphi(Y_T^\varepsilon)] - \mathbb{E}[\varphi(Y_T)] &= \mathbb{E}\bigg(\int_0^T [H(s,X_s^\varepsilon,Y_s^\varepsilon) - \bar{H}(s,Y_s^\varepsilon)]\nabla_y^2\tilde{u}(s,Y_s^\varepsilon)\mathrm{d}s\bigg) \\
&\quad + \mathbb{E}\bigg(\int_0^T [F(s,X_s^\varepsilon,Y_s^\varepsilon) - \bar{F}(s,Y_s^\varepsilon)]\nabla_y\tilde{u}(s,Y_s^\varepsilon)\mathrm{d}s\bigg) := \mathscr{U}_1 + \mathscr{U}_2.
\end{aligned} \tag{5.2}
$$

Define

$$\tilde{H}(t,x,y) := [H(t,x,y) - \bar{H}(t,y)]\nabla_y^2 \tilde{u}(t,y)$$

and

$$\tilde{F}(t,x,y) := [F(t,x,y) - \bar{F}(t,y)]\nabla_y \tilde{u}(t,y).$$

Let $\tilde{H}_n$, $\tilde{F}_n$ be the mollifying approximations of $\tilde{H}$ and $\tilde{F}$ defined similarly as in (4.7), respectively. We prepare the following approximation result, which is similar to Lemma 4.3.

Lemma 5.2. *Assume that $a, b \in C_b^{\delta,\alpha}$ and $F \in C_b^{\alpha/2,\delta,\alpha}$ with $0 < \delta \leqslant 1$, $0 < \alpha \leqslant 2$. Then we have*

$$\|\tilde{F} - \tilde{F}_n\|_\infty + \|\tilde{H} - \tilde{H}_n\|_\infty + \|a - a_n\|_\infty + \|b - b_n\|_\infty \leqslant C_0 n^{-\alpha}, \tag{5.3}$$

and

$$\|\tilde{F}_n\|_{C_b^{1,\delta,\alpha}} + \|\tilde{F}_n\|_{C_b^{\alpha,\delta,2}} + \|\tilde{H}_n\|_{C_b^{1,\delta,\alpha}} + \|\tilde{H}_n\|_{C_b^{\alpha,\delta,2}} + \|a_n\|_{C_b^{\delta,2}} + \|b_n\|_{C_b^{\delta,2}} \leqslant C_0 n^{2-\alpha}, \tag{5.4}$$

where $C_0 > 0$ is a constant independent of n.

Proof. Note that when $0 < \alpha \leqslant 1$, the conclusion has been proved in Lemma 4.1. Below, we shall focus on the case $1 < \alpha \leqslant 2$, and prove the corresponding estimates for $\tilde{H}$. The other estimates can be proved similarly. According to Lemma 5.1, it is easy to check that $\tilde{H} \in C_b^{\alpha/2,\delta,\alpha}$. As a result, we have

$$\begin{aligned}
|\tilde{H}(t,x,y) - \tilde{H}_n(t,x,y)| &\leqslant \int_{\mathbb{R}^{d_2+1}} \big|\tilde{H}(t-s,x,y-z) + \tilde{H}(t-s,x,y+z) \\
&\qquad - 2\tilde{H}(t,x,y)\big| \cdot \rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s \\
&\leqslant C_1 \int_{\mathbb{R}^{d_2+1}} \left(s^{\alpha/2} + |z|^\alpha\right) \cdot \rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s \leqslant C_1 n^{-\alpha},
\end{aligned}$$

and

$$\begin{aligned}
|\nabla_y^2 \tilde{H}_n(t,x,y)| &\leqslant \int_{\mathbb{R}^{d_2+1}} \big|\nabla_y \tilde{H}(t-s,x,y-z) - \nabla_y \tilde{H}(t-s,x,y)\big| \cdot |\nabla_z \rho_2^n(z)|\rho_1^n(s)\mathrm{d}z\mathrm{d}s \\
&\leqslant C_2 n \int_{\mathbb{R}^{d_2+1}} |z|^{\alpha-1} \cdot \rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s \leqslant C_2 n^{2-\alpha}.
\end{aligned}$$

So, the proof is finished. □

We are now in the position to give:

Proof of Theorem 2.5. We begin from (5.2) and proceed to control the first term. Let $\bar{H}_n$ be the average of $\tilde{H}_n$ with respect to $\mu^y(\mathrm{d}x)$, and set $\hat{H}_n := \tilde{H}_n - \bar{H}_n$. We write

$$\begin{aligned}
\mathscr{U}_1 &\leqslant \mathbb{E}\left|\int_0^T \left[\tilde{H}(s, X_s^\varepsilon, Y_s^\varepsilon) - \hat{H}_n(s, X_s^\varepsilon, Y_s^\varepsilon)\right]\mathrm{d}s\right| \\
&\quad + \mathbb{E}\left(\int_0^T \hat{H}_n(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\right) =: \mathscr{U}_{11} + \mathscr{U}_{12}.
\end{aligned}$$

Using (5.3), we can control the first term easily by

$$\mathscr{U}_{11} \leqslant C_1 n^{-\alpha}.$$

To control the second term, let $\tilde{\Psi}_n$ be the solution to the following Poisson equation in $\mathbb{R}^{d_1}$:

$$\mathscr{L}_0^n(x,y)\tilde{\Psi}_n(t,x,y) = \hat{H}_n(t,x,y),$$

where $\mathscr{L}_0^n$ is defined by (4.14) and $(t,y) \in \mathbb{R}_+ \times \mathbb{R}^{d_2}$ are viewed as parameters. Note that $\tilde{H}_n$ satisfies the centering condition (3.1). Thus, according to Theorem 3.1, we can use the Itô's formula to get that

$$\begin{aligned}
\mathbb{E}\left(\int_0^T \hat{H}_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right) &= \varepsilon\tilde{\Psi}_n(T,X_T^\varepsilon,Y_T^\varepsilon) - \varepsilon\tilde{\Psi}_n(0,x,y) \\
&\quad + \left(\int_0^T \big[b_n(X_s^\varepsilon,Y_s^\varepsilon) - b(X_s^\varepsilon,Y_s^\varepsilon)\big]\nabla_x\tilde{\Psi}_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right. \\
&\quad \left. + \int_0^T \big[a_n(X_s^\varepsilon,Y_s^\varepsilon) - a(X_s^\varepsilon,Y_s^\varepsilon)\big]\nabla_x^2\tilde{\Psi}_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right) \\
&\quad - \varepsilon\int_0^T \big(\partial_s + \mathscr{L}_1\big)\tilde{\Psi}_n(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s \\
&=: \tilde{\mathscr{Q}}_1(T,\varepsilon) + \tilde{\mathscr{Q}}_2(T,\varepsilon) + \tilde{\mathscr{Q}}_3(T,\varepsilon).
\end{aligned}$$

Using (5.3), (5.4) and exactly the same arguments as before, we get

$$\tilde{\mathscr{Q}}_1(T,\varepsilon) + \tilde{\mathscr{Q}}_2(T,\varepsilon) \leqslant C(\varepsilon + n^{-\alpha}).$$

and

$$\tilde{\mathscr{Q}}_3(T,\varepsilon) \leqslant C\varepsilon n^{2-\alpha}.$$

As a result, we have

$$\mathscr{U}_1 \leqslant C(\varepsilon + n^{-\alpha} + \varepsilon n^{2-\alpha}).$$

Using exactly the same arguments as above, we can also get

$$\mathscr{U}_2 \leqslant C(\varepsilon + n^{-\alpha} + \varepsilon n^{2-\alpha}).$$

Hence, taking $n = \varepsilon^{-1/2}$, we arrive at

$$\big|\mathbb{E}[\varphi(Y_T^\varepsilon)] - \mathbb{E}[\varphi(\bar{Y}_T)]\big| \leqslant C_T\big(\varepsilon^{\alpha/2} + \varepsilon\big) \leqslant C_T\varepsilon^{(\alpha/2)\wedge 1}.$$

The proof is finished. □

Finally, we give:

Proof of Theorem 2.7. It is well-known that the solution u^ε to equation (1.7) has the following probabilistic representation (see [24]):

$$u^\varepsilon(t,x,y) = \mathbb{E}\left(\int_0^{T-t} \psi(Y_s^\varepsilon)\mathrm{d}s + \varphi(Y_{T-t}^\varepsilon)\right).$$

Since φ is continuous, we can always find a sequence of functions $\varphi_n \in C_b^3$ such that $\|\varphi_n - \varphi\|_\infty \to 0$ as $n \to \infty$. As a result, we deduce by Theorem 2.5 that

$$\mathbb{E}\varphi(Y_{T-t}^\varepsilon) - \mathbb{E}\varphi(\bar{Y}_{T-t}) \leqslant \big[\mathbb{E}\varphi_n(Y_{T-t}^\varepsilon) - \mathbb{E}\varphi_n(\bar{Y}_{T-t})\big] + 2\|\varphi_n - \varphi\|_\infty.$$

Taking $\varepsilon \to 0$ first and then $n \to \infty$, we get

$$\lim_{\varepsilon\to 0}\big|\mathbb{E}\varphi(Y_{T-t}^\varepsilon) - \mathbb{E}\varphi(\bar{Y}_{T-t})\big| = 0.$$

On the other hand, since ψ is bounded, we can always find a sequence of functions $\psi_n \in C_b^3$ such that for every $p \geqslant 1$, $\|\psi_n - \psi\|_{L^p_{loc}} \to 0$ as $n \to \infty$. Then, for every $R > 0$, we write

$$\begin{aligned}
&\mathbb{E}\left(\int_0^{T-t} \left[\psi(Y_s^\varepsilon) - \psi(\bar{Y}_s)\right] \mathrm{d}s\right) = \mathbb{E}\left(\int_0^{T-t} \left[\psi_n(Y_s^\varepsilon) - \psi_n(\bar{Y}_s)\right] \mathrm{d}s\right) \\
&+ \mathbb{E}\left(\int_0^{T-t} \left[\psi(Y_s^\varepsilon) - \psi_n(Y_s^\varepsilon)\right] 1_{\{|Y_s^\varepsilon| \leqslant R\}} \mathrm{d}s\right) + \mathbb{E}\left(\int_0^{T-t} \left[\psi(\bar{Y}_s) - \psi_n(\bar{Y}_s)\right] 1_{\{|\bar{Y}_s| \leqslant R\}} \mathrm{d}s\right) \\
&+ \mathbb{E}\left(\int_0^{T-t} \left[\psi(\bar{Y}_s) - \psi_n(\bar{Y}_s)\right] 1_{\{|\bar{Y}_s| > R\}} \mathrm{d}s\right) + \mathbb{E}\left(\int_0^{T-t} \left[\psi(\bar{Y}_s) - \psi_n(\bar{Y}_s)\right] 1_{\{|\bar{Y}_s| > R\}} \mathrm{d}s\right).
\end{aligned}$$

Due to Theorem 2.5, the first term goes to 0 as $\varepsilon \to 0$. By Krylov's estimate (see [24]) we have that for some $p > d_1 + d_2$,

$$\mathbb{E}\left(\int_0^{T-t} \left[\psi(Y_s^\varepsilon) - \psi_n(Y_s^\varepsilon)\right] 1_{\{|Y_s^\varepsilon| \leqslant R\}} \mathrm{d}s\right) \leqslant C\|\psi - \psi_n\|_{L^p_{loc}},$$

which goes to 0 as $n \to \infty$. Finally, the last part goes to 0 as $R \to \infty$ by Chebyshev's inequality. This finishes the proof. $\square$

References

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

[2] Bogachev V. I., Shaposhnikov S. V. and Veretennikov A. Yu.: Differentiability of solutions of stationary Fokker-Pllanck-Kolmogorov equations with respect to a parameter. *Discrete and Continuous Dynamical Systems - Series B*, **36** (2016), 3519–3543.

[3] Bréhier C. E.: Strong and weak orders in averaging for SPDEs. *Stoch. Process. Appl.*, **122** (2012), 2553–2593.

[4] Bréhier C. E.: Analysis of an HMM time-discretization scheme for a system of stochastic PDEs. *SIAM J. Numer. Anal.*, **51** (2013), 1185–1210.

[5] Bréhier C. E.: Orders of convergence in the averaging principle for SPDEs: the case of a stochastically forced slow component. https://arxiv.org/abs/1810.06448.

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

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

[8] Chen Z., Hu E., Xie L. and Zhang X.: Heat kernels for non-symmetric diffusions operators with jumps. *J. Diff. Equations*, **263** (2017), 6576–6634.

[9] Chevyrev I., Friz P., Korepanov A. and Melbourne I.: Superdiffusive limits for deterministic fast-slow dynamical systems. arXiv:1907.04825v1.

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

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

[12] Friedlin M.: Functional integration and partial differential equations, Princeton Univ. Press, Princeton, N.J., 1985.

[13] Freidlin M. and Wentzell A.: Random perturbations of dynamical systems, Springer Science & Business Media, Berlin, Heidelberg, 2012.

[14] Fu H. and Duan J.: An averaging principle for two-scale stochastic partial differential equations. *Stoch. Dyn.*, **11** (2011), 353–367.

[15] Givon D.: Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems. *Multiscale Model. Simul.* **6** (2007), 577–594.

[16] 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.

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

[18] Khasminskii R. Z.: A limit theorem for the solutions of differential equations with random right-hand sides. *Theory Probab. Appl.*, **11** (1966), 390–406.

[19] Khasminskii R. Z.: On an averging principle for Itô stochastic differential equations. *Kibernetica*, **4** (1968), 260–279.

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

[21] Khasminskii R. Z and Yin G.: Limit behavior of two-time-scale diffusions revised. *J.Diff. Equ.*, **212** (2005), 85–113.

[22] Kifer Y.: Diffusion approximation for slow motion in fully coupled averaging. *Proba. Theor. Relat. Fields*, **129** (2004) 157–181.

[23] Kifer Y.: Another proof of the averaging principle for fully coupled dynamical systems with hyperbolic fast motions. *Discrete Contin. Dyn. Syst.*, **13** (2005), 1187–1201.

[24] Krylov N. V.: *Controlled diffusion processes.* Translated from the Russian by A.B. Aries. Applications of Mathematics, 14. Springer-Verlag, New York-Berlin, 1980.

[25] Krylov N. V. and Bogolyubov N.: Les proprietes ergodiques des suites des probabilites en chaine. *C. R. Acad. Sci. Paris*, **204** (1937), 1454–1456.

[26] Krylov N. V. and Röckner M.: Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Related Fields*, **131** (2005), 154–196.

[27] Kuehn C.: Multiple time scale dynamics, volume 191 of Applied Mathematical Sciences. Springer, 2015.

[28] Ladyženskaja O.A., Solonnikov V.A. and Ural'ceva N.N.: Linear and Quasi-linear Equations of Parabolic Type. Translated from Russian by S.Smmith. Amercian Mathematical Society, 1968.

[29] Li X. M.: An averaging principle for a completely integrable stochastic Hamiltonian system. *Nonlinearity*, **21** (2008), 803–822.

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

[31] Pardoux E. and Veretennikov A. Yu.: On the Poisson equation and diffusion approximation. I. *Ann. Prob.*, **29** (2001), 1061–1085.

[32] Pardoux E. and Veretennikov A. Yu.: On the Poisson equation and diffusion approximation 2. *Ann. Prob.*, **31** (2003), 1166–1192.

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

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

[35] Veretennikov A. Yu.: On the strong solutions of stochastic differential equations. *Theory Probab. Appl.*, **24** (1979), 354–366.

[36] Veretennikov A. Yu.: On the averaging principle for systems of stochastic differential equations. *Math. USSR Sborn.*, **69** (1991), 271–284.

[37] Veretennikov A. Yu.: On polynomial mixing bounds for stochastic differential equations. *Stoch. Proc. Appl.*, **70** (1997), 115–127.

[38] Veretennikov A. Yu.: On Sobolev solutions of poisson equations in $\mathbb{R}^d$ with a parameter. *J. Math. Sci.*, **179** (2011), 1–32.

[39] Wang W. and Roberts A. J.: Average and deviation for slow-fast stochastic partial differential equations. *J. Differential Equations*, **253** (2012), 1265–1286.
[40] Xie L. and Zhang X.: Sobolev differentiable flows of SDEs with local Sobolev and super-linear growth coefficients. *Ann. Prob.*, **44** (2016), 3661–3687.
[41] Xie L. and Zhang X.: Ergodicity of stochastic differential equations with jumps and singular coefficients. *Ann. Inst Henri Poincare-Pr.*, **56** (2020), 175-C229.
[42] Zhang B., Fu H., Wan L. and Liu J.: Weak order in averaging principle for stochastic differential equations with jumps. *Adv. Difference Equ.*, 2018, Paper No. 197, 20 pp.
[43] Zhang X.: Stochastic homemomorphism flows of SDEs with singular drifts and Sobolev diffusion coefficients. *Electron. J. Probab.*, **16** (2011), 1096–1116.
[44] Zvonkin A. K.: A transformation of the phase space of a diffusion process that removes the drift. *Mat. Sb.*, **135** (1974), 129–149.

Michael Röckner: Fakultät für Mathematik, Universität Bielefeld, D-33501 Bielefeld, Germany, and Academy of Mathematics and Systems Science, Chinese Academy of Sciences (CAS), Beijing, 100190, P.R.China, Email: roeckner@math.uni-bielefeld.de

Xiaobin Sun: School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, Jiangsu 221000, P.R.China, Email: xbsun@jsnu.edu.cn

Longjie Xie: School of Mathematics and Statistics, Jiangsu Normal University, Xuzhou, Jiangsu 221000, P.R.China, Email: longjiexie@jsnu.edu.cn