DIFFUSION APPROXIMATION FOR FULLY COUPLED STOCHASTIC DYNAMICAL SYSTEMS

MICHAEL RÖCKNER AND LONGJIE XIE

Abstract. We consider a Poisson equation in $\mathbb{R}^d$ for the elliptic operator corresponding to an ergodic diffusion process. Optimal regularity and smoothness with respect to the parameter are obtained under mild conditions on the coefficients. The result is then applied to establish a general diffusion approximation for fully coupled multi-time-scales stochastic differential equations with only Hölder continuous coefficients. Four different limit equations as well as rates of convergence are obtained. Moreover, the convergence is shown to rely only on the regularities of the coefficients with respect to the slow variable, and does not depend on their regularities with respect to the fast component.

AMS 2010 Mathematics Subject Classification: 60H10, 60J60, 35B30.
Keywords and Phrases: Poisson equation; multi scale system; averaging principle; diffusion approximation; homogenization.

1. Introduction

1.1. **Poisson equation in the whole space.** The first topic of this paper is to study the following Poisson equation in $\mathbb{R}^{d_1}$ ($d_1 \geqslant 1$):

$$\mathscr{L}_0(x,y)u(x,y) = f(x,y), \quad x \in \mathbb{R}^{d_1}, \tag{1.1}$$

where $y \in \mathbb{R}^{d_2}$ ($d_2 \geqslant 1$) is a parameter, and

$$\mathscr{L}_0 := \mathscr{L}_0(x,y) := \sum_{i,j=1}^{d_1} a^{ij}(x,y)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d_1} b^i(x,y)\frac{\partial}{\partial x_i}. \tag{1.2}$$

The Poisson equation is one of the well-known equations in mathematical physics. When the above equation is formulated on a compact set, the corresponding theory is well known, see e.g. [21] or [16, Chapter 12]. However, equation (1.1) *with a parameter and in the whole space* $\mathbb{R}^{d_1}$ has been studied only relatively recently, and it turns out to be one of the key tools in the theory of stochastic averaging, homogenization and other limit theorems in probability theory (see e.g. [18, 23, 40, 41]). This understanding was the reason for a series of papers by Pardoux and Veretennikov [42, 43, 44], where equation (1.1) was first studied and then used

This work is supported by the DFG through CRC 1283, the Alexander-von-Humboldt foundation, NSFC (No. 11701233, 11931004) and NSF of Jiangsu (BK20170226).

to establish diffusion approximations for slow-fast stochastic differential equations (SDEs for short), see also [24, 25, 32, 47, 53] for further generalizations.

Since there is no boundary condition in equation (1.1), the solution u turns out to be defined up to an additive constant, which is quite natural due to $\mathscr{L}_0(x, y)1 \equiv 0$. To fix this constant, it is convenient to make the following "centering" assumption on the potential term f:

$$\int_{\mathbb{R}^{d_1}} f(x, y)\mu^y(\mathrm{d}x) = 0, \quad \forall y \in \mathbb{R}^{d_2}, \tag{1.3}$$

where $\mu^y(\mathrm{d}x)$ is the (unique) invariant measure of an ergodic Markov process X_t^y (see (1.7) below) with a generator $\mathscr{L}_0(\cdot, y)$ given y. Such kind of assumption is analogous to the centering in the standard central limit theorem (CLT for short). We mention that the main problem addressed in [42, 43, 44] concerning equation (1.1) is the second order regularity of the solution u with respect to x as well as the parameter y (which is more difficult !), which would suffice for the application of Itô formula with some diffusions plugged in for both variables, and [42] used mainly probability arguments while [43, 44] used essentially results from the partial differential equations (PDEs for short) theory. Later on, the results of [42, 43, 44] for the Poisson equation (1.1) have also been adopted to study the CLT, moderate and large derivations (see e.g. [9, 14, 38, 48]), spectral methods, averaging principle and homogenization for multi-scale systems (see e.g. [1, 7, 47]) as well as numerical approximation for time-averaging estimators and the invariant measure of SDEs or stochastic partial differential equations (SPDEs for short) (see e.g. [8, 25, 37, 39]).

One of our objectives in this paper is to further study the Poisson equation (1.1). We develop a robust method to study the regularities of the solution u, especially for the smoothness with respect to the parameter y, which leads to simplifications and extensions of the existing results. The main result in this direction is given by **Theorem 2.1**. Our argument is different from all those works mentioned above, and the assumptions on the coefficients are weaker. In fact, instead of establishing differentiablity of the corresponding semigroup with respect to the parameter y, we shall focus on the optimal regularity of the solution u with respect to the x variable. Then based on a key observation of a transfer formula, we use an induction argument to show that smoothness of the solution u with respect to the parameter y follows directly by the optimal a priori estimate with respect to the x variable, which is much simpler insofar. In addition, our method has at least three more advantages: first of all, we obtain any order of differentiablity (which is important for the asymptotic expansion analysis used in [27, 28]) as well as Hölder continuity (which will play a crucial role below for us to study diffusion approximations) of the solution with respect to the parameter y under explicit conditions on the coefficients; secondly, we provide explicit dependence on the norms of the coefficients a, b as well as the potential term f involved, which might be useful for numerical analysis and has been used essentially in [47] to study the rate of convergence in averaging principle for two-time-scales SDEs; thirdly, our argument can also be adopted to study equation (1.1) in Sobolev spaces with certain L^p conditions on the coefficients and the potential term, thus leading to only weak differentiablity of the solution u with respect to the parameter. Such problem has been posed in [53] and seems difficult to be handled by the arguments used in the previous publications. For the sake of simplicity, we do not deal with this setting in the present article and postpone it to further studies.

1.2. **Diffusion approximations.** The result of Theorem 2.1 will then be used to study the asymptotic problem for fully coupled multi-scale stochastic dynamical systems, which is the second topic of this paper. More precisely, consider the following in-homogeneous multi-time-scales SDE in $\mathbb{R}^{d_1+d_2}$:

$$\begin{cases} \mathrm{d}X_t^\varepsilon = \alpha_\varepsilon^{-2} b(X_t^\varepsilon, Y_t^\varepsilon)\mathrm{d}t + \beta_\varepsilon^{-1} c(X_t^\varepsilon, Y_t^\varepsilon)\mathrm{d}t + \alpha_\varepsilon^{-1}\sigma(X_t^\varepsilon, Y_t^\varepsilon)\mathrm{d}W_t^1, \\ \mathrm{d}Y_t^\varepsilon = F(t, X_t^\varepsilon, Y_t^\varepsilon)\mathrm{d}t + \gamma_\varepsilon^{-1} H(t, X_t^\varepsilon, Y_t^\varepsilon)\mathrm{d}t + G(t, X_t^\varepsilon, Y_t^\varepsilon)\mathrm{d}W_t^2, \\ X_0^\varepsilon = x \in \mathbb{R}^{d_1}, \quad Y_0^\varepsilon = y \in \mathbb{R}^{d_2}, \end{cases} \tag{1.4}$$

where the small parameters $\alpha_\varepsilon, \beta_\varepsilon, \gamma_\varepsilon \downarrow 0$ as $\varepsilon \to 0$, and without loss of generality, we may assume $\alpha_\varepsilon^2/\beta_\varepsilon \to 0$ as $\varepsilon \to 0$. Such model has wide applications in many real world dynamical systems including planetray motion, climate models (see e.g. [30, 36]), geophysical fluid flows (see e.g. [19]), intracellular biochemical reactions (see e.g. [5]), etc. We refer the interested readers to the books [31, 45] for a more comprehensive overview. Note that we are considering (1.4) in the whole space and not just on compact sets. Moreover, there exist two time-scales in the fast component X_t^ε and even the slow motion Y_t^ε has a fast varying term. Usually, the underlying system (1.4) is difficult to deal with due to the widely separated time-scales and the cross interactions of slow and fast modes. Hence a simplified equation which governs the evolution of the system for small ε is highly desirable. We also mention that the infinitesimal generator corresponding to $(X_t^\varepsilon, Y_t^\varepsilon)$ has the form

$$\mathscr{L}_\varepsilon := \alpha_\varepsilon^{-2}\mathscr{L}_0(x, y) + \beta_\varepsilon^{-1}\mathscr{L}_3(t, x, y) + \gamma_\varepsilon^{-1}\mathscr{L}_2(t, x, y) + \mathscr{L}_1(t, x, y),$$

where $\mathscr{L}_0(x, y)$ is given by (1.2) with $a(x, y) := \sigma\sigma^*(x, y)/2$, and

$$\begin{aligned} \mathscr{L}_3 &:= \mathscr{L}_3(x, y) := \sum_{i=1}^{d_1} c^i(x, y)\frac{\partial}{\partial x_i}, \\ \mathscr{L}_2 &:= \mathscr{L}_2(t, x, y) := \sum_{i=1}^{d_2} H^i(t, x, y)\frac{\partial}{\partial y_i}, \\ \mathscr{L}_1 &:= \mathscr{L}_1(t, x, y) := \sum_{i,j}^{d_2} \mathcal{G}^{ij}(t, x, y)\frac{\partial^2}{\partial y_i \partial y_j} + \sum_{i=1}^{d_2} F^i(t, x, y)\frac{\partial}{\partial y_i} \end{aligned} \tag{1.5}$$

with $\mathcal{G}(t,x,y) = GG^*(t,x,y)/2$. Thus the asymptotic behavior of SDE (1.4) as $\varepsilon \to 0$ is also closely related to the limit theorem for solutions of second order parabolic and elliptic equations with singularly perturbed terms, which has its own interest in the theory of PDEs, see e.g. [23, 41] and [17, Chapter IV].

When $c = H \equiv 0$, the celebrated theory of averaging principle asserts that the slow motion Y_t^ε will convergence in distribution as $\varepsilon \to 0$ to the solution $\bar{Y}_t$ of the following reduced equation in $\mathbb{R}^{d_2}$:

$$\mathrm{d}\bar{Y}_t = \bar{F}(t,\bar{Y}_t)\mathrm{d}t + \bar{G}(t,\bar{Y}_t)\mathrm{d}W_t^2, \quad \bar{Y}_0 = y, \tag{1.6}$$

where the new averaged coefficients are given by

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

and for each $y \in \mathbb{R}^{d_2}$, $\mu^y(\mathrm{d}x)$ is the unique invariant measure for X_t^y which satisfies the frozen equation

$$\mathrm{d}X_t^y = b(X_t^y, y)\mathrm{d}t + \sigma(X_t^y, y)\mathrm{d}W_t^1, \quad X_0^y = x \in \mathbb{R}^{d_1}. \tag{1.7}$$

The corresponding results are also known as the averaging principle of functional law of large numbers (LLN for short) type and have been intensively studied by the classical time discretisation method, see e.g. [2, 3, 13, 22, 29, 34, 46, 51], see also [10, 11, 12, 49] for similar results for SPDEs. Generalization to the general case that $\alpha_\varepsilon = \beta_\varepsilon = \gamma_\varepsilon$ was first carried out by Papanicolaou, Stroock and Varadhan [40] for a compact state space for the fast component and time-independent coefficients, see also [4] for a similar result in terms of PDEs. It was found that the limit distribution of the slow component will be obtained in terms of the solution of an auxiliary Poisson equation. Such result can be regarded as an averaging principle of functional CLT type and is often called diffusion approximation, which is important for applications in homogenization. Later a non-compact and homogeneous case with $c \equiv 0$ and $\alpha_\varepsilon = \gamma_\varepsilon$ was studied in [42, 43, 44] by using the method of martingale problem and in [28] by the asymptotic expansion approach, see also [45, Chapter 11]. We also mention that for numerical purposes, the existence of the effective system (1.6) is not enough, and the rate of convergence of the slow variable to its limit distribution has to be derived. The main motivation comes from the well-known Heterogeneous Multi-scale Methods used to approximate the slow component Y_t^ε, see e.g. [6, 7, 15, 26]. However, getting error bounds is significantly harder than just showing convergence, and many of the methods commonly employed to show distributional convergence only possibly yield a convergence rate after serious added effort. In this direction, there are many works devoted to study the convergence rate in the averaging principle of LLN type for SDE (1.4) when $c = H \equiv 0$, see e.g. [20, 35, 50, 54] and the references therein. As far as we know, there is still no result concerning error bounds for the CLT type convergence of system (1.4) in the general case.

We will study the asymptotic problem for system (1.4) systematically. The main result is given by **Theorem 2.3**. It turns out that, depending on the orders how $\alpha_\varepsilon, \beta_\varepsilon, \gamma_\varepsilon$ go to zero, we shall have four different regimes of interactions, which lead to four different asymptotic behaviors of system (1.4) as $\varepsilon \to 0$, i.e.,

$$\begin{cases} \lim\limits_{\varepsilon\to 0} \dfrac{\alpha_\varepsilon}{\gamma_\varepsilon} = 0 \quad \text{and} \quad \lim\limits_{\varepsilon\to 0} \dfrac{\alpha_\varepsilon^2}{\beta_\varepsilon\gamma_\varepsilon} = 0, & \text{Regime 1;} \\ \lim\limits_{\varepsilon\to 0} \dfrac{\alpha_\varepsilon}{\gamma_\varepsilon} = 0 \quad \text{and} \quad \alpha_\varepsilon^2 = \beta_\varepsilon\gamma_\varepsilon, & \text{Regime 2;} \\ \alpha_\varepsilon = \gamma_\varepsilon \quad \text{and} \quad \lim\limits_{\varepsilon\to 0} \dfrac{\alpha_\varepsilon}{\beta_\varepsilon} = 0, & \text{Regime 3;} \\ \alpha_\varepsilon = \beta_\varepsilon = \gamma_\varepsilon, & \text{Regime 4.} \end{cases} \tag{1.8}$$

If α_ε and α_ε^2 go to zero faster than γ_ε and $\beta_\varepsilon\gamma_\varepsilon$ respectively (Regime 1), we show that the limit behavior of system (1.4) coincides with SDE (1.6), i.e., the traditional case corresponding to $c = H \equiv 0$; if α_ε goes to zero faster than γ_ε while α_ε^2 and $\beta_\varepsilon\gamma_\varepsilon$ are of the same order (Regime 2), then the homogenization effect of term c will occur in the limit dynamics; whereas if α_ε and γ_ε are of the same order and α_ε goes to zero faster than β_ε (Regime 3), then the homogenization effect of term H appears; finally, when all the parameters are of the same order (Regime 4), then homogenization effects of term c and term H will occur together.

We shall handle Regime 1-4 in a robust and unified way. Our method relies only on the technique of Poisson equation (1.1), and does not involve extra time discretisation procedure (see e.g. [3, 12, 34, 51]), martingale problem (see [38, 40, 42, 43, 44, 48]) nor asymptotic expansion argument (see [27, 28]) and thus is quite simple. Moreover, the conditions on the coefficients are weaker (only Hölder continuous) than the known results in the literature, and rates of convergence are obtained as easy by-products of our argument, which we believe are rather sharp, see **Remark 2.4** and **Remark 2.5** for more explanations. We also point out that unlike the above mentioned results, where the second order regularity of the solution to the Poisson equation with respect to the parameter is commonly needed in the proof, we only use its Hölder continuity, which also simplifies the arguments used in [47] and appear intuitively natural, since the second order derivative of the solution to the Poisson equation with respect to the parameter does not appear in the final limit equation. Throughout our proof, two new fluctuation estimates of functional LLN type in Lemma 4.2 and functional CLT type in Lemma 4.4 will play important roles, which might be used to study other limit theorems and should be of independent interest.

The rest of this paper is organized as follows. In Section 2 we provide the assumptions and state our main results. Section 3 is devoted to the study of Possion equation (1.1) and we prove Theorem 2.1. In Section 4 we prepare two fluctuation lemmas, and then we give the proof of Theorem 2.3 in Section 5. Throughout our paper, we use the following convention: C and c with or without subscripts will denote positive constants, whose values may change in different places, and whose dependence on parameters can be traced from the calculations.

Notations: To end this section, we introduce some notations. Let $\mathbb{N}^* := \{1, 2 \cdots\}$. Given a function space, the subscript b will stand for boundness, while the subscript p stands for polynomial growth in x. More precisely, for a function $f(t, x, y) \in L_p^\infty := L_p^\infty(\mathbb{R}_+ \times \mathbb{R}^{d_1+d_2})$, we mean there exist constants $C, m > 0$ such that

$$|f(t, x, y)| \leqslant C(1 + |x|^m), \quad \forall t > 0, x \in \mathbb{R}^{d_1}, y \in \mathbb{R}^{d_2}.$$

For $0 < \delta \leqslant 1$, the space $C_p^{\delta,0} := C_p^{\delta,0}(\mathbb{R}^{d_1+d_2})$ consists of all functions which are local Hölder continuous and have at most polynomial growth in x uniformly with respect to y, i.e., there exist constants $C, m > 0$ such that for any $x_1, x_2 \in \mathbb{R}^{d_1}$,

$$|f(x_1, y) - f(x_2, y)| \leqslant C\big(|x_1 - x_2|^\delta \wedge 1\big)\big(1 + |x_1|^m + |x_2|^m\big), \quad \forall y \in \mathbb{R}^{d_2}.$$

We also define a quasi-norm for $C_p^{\delta,0}$ by

$$[f]_{C_p^{\delta,0}} := \sup_{y \in \mathbb{R}^{d_2}} \sup_{|x_1|, |x_2| \leqslant 1, |x_1 - x_2| \leqslant 1} |f(x_1, y) - f(x_2, y)|/|x_1 - x_2|^\delta.$$

For $0 < \vartheta < 1$, the space $C_p^{\delta,\vartheta} := C_p^{\delta,\vartheta}(\mathbb{R}^{d_1+d_2})$ consists of all functions that are δ-local Hölder continuous with polynomial growth in x and ϑ-Hölder continuous in y, i.e., there exist constants $C, m > 0$ such that for any $x_1, x_2 \in \mathbb{R}^{d_1}$ and $y_1, y_2 \in \mathbb{R}^{d_2}$,

$$|f(x_1, y_1) - f(x_2, y_2)| \leqslant C\Big[\big(|x_1 - x_2|^\delta \wedge 1\big) + \big(|y_1 - y_2|^\vartheta \wedge 1\big)\Big]\big(1 + |x_1|^m + |x_2|^m\big).$$

Similarly, we define a quasi-norm for $C_p^{\delta,\vartheta}$ by

$$[f]_{C_p^{\delta,\vartheta}} := \sup_{|y_1 - y_2| \leqslant 1} \sup_{|x_1|, |x_2| \leqslant 1, |x_1 - x_2| \leqslant 1} \frac{|f(x_1, y_1) - f(x_2, y_2)|}{|x_1 - x_2|^\delta + |y_1 - y_2|^\vartheta}.$$

When $\delta, \gamma \geqslant 1$, we use $C_p^{\delta,\vartheta} := C_p^{\delta,\vartheta}(\mathbb{R}^{d_1+d_2})$ to denote the space of all functions f satisfying $\partial_x^{[\delta]}\partial_y^{[\vartheta]} f \in C_p^{\delta-[\delta],\vartheta-[\vartheta]}$. Finally, $C_p^{\gamma,\delta,\vartheta} := C_p^{\gamma,\delta,\vartheta}(\mathbb{R}_+ \times \mathbb{R}^{d_1+d_2})$ with $0 < \gamma \leqslant 1$ denotes the space of all functions f such that for every fixed $t > 0$, $f(t, \cdot, \cdot) \in C_p^{\delta,\vartheta}$ and for every $(x, y) \in \mathbb{R}^{d_1+d_2}$, $f(, \cdot, x, y) \in C_b^\gamma(\mathbb{R}_+)$, where C_b^γ is the usual bounded Hölder space.

2. Assumptions and main results

To state our main results, we first introduce some basic assumptions. Throughout this paper, we shall always assume the following non-degeneracy conditions on the diffusion coefficients:

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

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

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

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

Note that the operator $\mathscr{L}_0$ in the Poisson equation (1.1) can be viewed as the infinitesimal generator of the frozen SDE (1.7). We make the following very weak recurrence assumption on the drift b to ensure the existence of an invariant measure $\mu^y(\mathrm{d}x)$ for X_t^y:

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

Our main result concerning the Poisson equation (1.1) is as follows.

Theorem 2.1. *Let* ($\mathbf{A}_\sigma$) *and* ($\mathbf{A}_b$) *hold. Assume that $a, b \in C_b^{\delta,\eta}$ with $0 < \delta \leqslant 1$ and $\eta \geqslant 0$. Then for every function $f \in C_p^{\delta,\eta}$ satisfying (1.3), there exists a unique solution $u \in C_p^{2+\delta,\eta}$ to equation (1.1).*

Moreover, there exist constants $m > 0$ and $C_0 > 0$ depending only on d_1, d_2 and $\|a\|_{C_b^{\delta,0}}, \|b\|_{C_b^{\delta,0}}, [f]_{C_p^{\delta,0}}$ such that:
(i) (Case $\eta = 0$) for any $x \in \mathbb{R}^{d_1}$ and $y \in \mathbb{R}^{d_2}$,

$$|u(x,y)| + |\nabla_x u(x,y)| + |\nabla_x^2 u(x,y)| \leqslant C_0(1 + |x|^m), \tag{2.1}$$

and for any $x_1, x_2 \in \mathbb{R}^{d_1}$,

$$|\nabla_x^2 u(x_1,y) - \nabla_x^2 u(x_2,y)| \leqslant C_0\big(|x_1 - x_2|^\delta \wedge 1\big)(1 + |x_1|^m + |x_2|^m); \tag{2.2}$$

(ii) (Case $\eta > 0$) for any $x \in \mathbb{R}^{d_1}$,

$$\|u(x,\cdot)\|_{C_b^\eta} \leqslant C_0 \mathcal{K}_\eta (1 + |x|^m), \tag{2.3}$$

where $\mathcal{K}_\eta > 0$ is a constant depending only on $\|a\|_{C_b^{\delta,\eta}}, \|b\|_{C_b^{\delta,\eta}}$ and $[f]_{C_p^{\delta,\eta}}$, which is defined recursively by (2.4) below.

Remark 2.2. *(i) Estimates (2.1) and (2.2) reflect the optimal regularity of the solution u with respect to the x variable, which are natural because we can get the two derivatives for free by virtue of the elliptic property of the operator. Note that we allow the potential term f to have polynomial growth in x. Estimate (2.3) means* *that we need both the coefficients and the right hand side f to be differentiable with respect to the y variable in order to guarantee the same regularity for u with respect to y, which is also reasonable and optimal since y is only a parameter in the equation. Note that only Hölder continuity of the coefficients with respect to the x variable is needed. The key point of our arguments is that we show that the smoothness of the solution with respect to the parameter y follows directly from its optimal regularity with respect to the x variable.*

(ii) The positive constant $\mathcal{K}_\eta$ in (2.3) can be given explicitly in terms of $\|a\|_{C_b^{\delta,\eta}}$, $\|b\|_{C_b^{\delta,\eta}}$ and $[f]_{C_p^{\delta,\eta}}$. In fact, let

$$\kappa_\eta := \|a\|_{C_b^{\delta,\eta}} + \|b\|_{C_b^{\delta,\eta}} \quad \textit{and} \quad \mathcal{K}_0 := 1.$$

Then when η is an integer, we have

$$\mathcal{K}_\eta := [f]_{C_p^{\delta,\eta}} + \sum_{\ell=1}^{\eta} C_\eta^\ell \cdot \kappa_\ell \cdot \mathcal{K}_{\eta-\ell}, \tag{2.4}$$

while for $\eta \in (0, \infty \setminus \mathbb{N}^*)$, we have

$$\mathcal{K}_\eta = \|f\|_{C_p^{\delta,\eta}} + \sum_{\ell=1}^{[\eta]} C_{[\eta]}^\ell \cdot \kappa_\ell \cdot \mathcal{K}_{\eta-\ell} + \sum_{\ell=0}^{[\eta]} C_{[\eta]}^{[\eta]-\ell} \cdot \kappa_{\eta-\ell} \cdot \mathcal{K}_\ell. \tag{2.5}$$

This will be useful for numerical purposes to approximate the solution of the Poisson equation with singular coefficients, see e.g. [47].

Now we turn to the multi-time-scales stochastic dynamical system (1.4). As shown in [40], the limit behavior for the slow component Y_t^ε will be given in terms of the solution of an auxiliary Poisson equation involving the drift H. Thus, it is necessary to make the following assumption:

($\mathbf{A}_H$): the drift H is centered, i.e.,

$$\int_{\mathbb{R}^{d_1}} H(t,x,y)\mu^y(\mathrm{d}x) = 0, \quad \forall (t,y) \in \mathbb{R}_+ \times \mathbb{R}^{d_2}, \tag{2.6}$$

where $\mu^y(\mathrm{d}x)$ is the invariant measure for SDE (1.7).

Note that the drift c is not involved in the frozen equation (1.7). We need the following additional condition for c to ensure the non-explosion of the solution X_t^ε: for $\varepsilon > 0$ small enough, it holds that

$$\lim_{|x|\to\infty} \sup_y \langle x, b(x,y) + \varepsilon c(x,y) \rangle = -\infty. \tag{2.7}$$

Under (2.6) and according to Theorem 2.1, there exists a unique solution $\Phi(t,x,y)$ to the following Poisson equation in $\mathbb{R}^{d_1}$:

$$\mathscr{L}_0(x,y)\Phi(t,x,y) = -H(t,x,y), \quad x \in \mathbb{R}^{d_1}, \tag{2.8}$$

where $(t,y)\in\mathbb{R}_+\times\mathbb{R}^{d_2}$ are regarded as parameters. We introduce the new averaged drift coefficients by

$$
\begin{aligned}
\hat{F}_1(t,y) &:= \int_{\mathbb{R}^{d_1}} F(t,x,y)\mu^y(\mathrm{d}x);\\
\hat{F}_2(t,y) &:= \int_{\mathbb{R}^{d_1}} \big[F(t,x,y)+c(x,y)\cdot\nabla_x\Phi(t,x,y)\big]\mu^y(\mathrm{d}x);\\
\hat{F}_3(t,y) &:= \int_{\mathbb{R}^{d_1}} \big[F(t,x,y)+H(t,x,y)\cdot\nabla_y\Phi(t,x,y)\big]\mu^y(\mathrm{d}x);\\
\hat{F}_4(t,y) &:= \int_{\mathbb{R}^{d_1}} \big[F(t,x,y)+c(x,y)\cdot\nabla_x\Phi(t,x,y)+H(t,x,y)\cdot\nabla_y\Phi(t,x,y)\big]\mu^y(\mathrm{d}x),
\end{aligned}
$$

which correspond to Regime 1-Regime 4 described in (1.8), and the new diffusion coefficients are given by

$$
\begin{aligned}
\hat{G}_1(t,y)=\hat{G}_2(t,y) &:= \sqrt{\int_{\mathbb{R}^{d_1}} GG^*(t,x,y)\mu^y(\mathrm{d}x)};\\
\hat{G}_3(t,y)=\hat{G}_4(t,y) &:= \sqrt{\int_{\mathbb{R}^{d_1}} \Big[GG^*(t,x,y)+H(t,x,y)\Phi^*(t,x,y)\Big]\mu^y(\mathrm{d}x)}.
\end{aligned}
$$

Hence, the precise formulation of the limit equation for SDE (1.4) will be of the following form: for $k=1,\cdots,4$,

$$
\mathrm{d}\hat{Y}^k_t=\hat{F}_k(t,\hat{Y}^k_t)\mathrm{d}t+\hat{G}_k(t,\hat{Y}^k_t)\mathrm{d}W^2_t,\quad \hat{Y}^k_0=y. \tag{2.9}
$$

The second main result of this paper is as follows.

Theorem 2.3. *Let* $(\mathbf{A}_\sigma)$-$(\mathbf{A}_b)$-$(\mathbf{A}_G)$-$(\mathbf{A}_H)$-*(2.7) hold,* $T>0$ *and* $\delta\in(0,1]$.
(i) (Regime 1) If $b,\sigma\in C^{\delta,\vartheta}_b$, $F,H,G\in C^{\vartheta/2,\delta,\vartheta}_p$ *with* $\vartheta\in(0,2]$, $c\in L^\infty_p$, *and assume further that* $\lim_{\varepsilon\to0}\alpha^\vartheta_\varepsilon/\gamma_\varepsilon=0$, *then for every* $\varphi\in C^{2+\vartheta}_b(\mathbb{R}^{d_2})$, *we have*

$$
\sup_{t\in[0,T]}\Big|\mathbb{E}[\varphi(Y^\varepsilon_t)]-\mathbb{E}[\varphi(\hat{Y}^1_t)]\Big|\leqslant C_T\left(\frac{\alpha^\vartheta_\varepsilon}{\gamma_\varepsilon}+\frac{\alpha^2_\varepsilon}{\gamma^2_\varepsilon}+\frac{\alpha^2_\varepsilon}{\beta_\varepsilon\gamma_\varepsilon}\right);
$$

(ii) (Regime 2) if $b,\sigma\in C^{\delta,\vartheta}_b$, $F,H,G\in C^{\vartheta/2,\delta,\vartheta}_p$ *and* $c\in C^{\delta,\vartheta}_p$ *with* $\vartheta\in(0,2]$, *and assume further that* $\lim_{\varepsilon\to0}\alpha^\vartheta_\varepsilon/\gamma_\varepsilon=0$, *then for every* $\varphi\in C^{2+\vartheta}_b(\mathbb{R}^{d_2})$, *we have*

$$
\sup_{t\in[0,T]}\Big|\mathbb{E}[\varphi(Y^\varepsilon_t)]-\mathbb{E}[\varphi(\hat{Y}^2_t)]\Big|\leqslant C_T\left(\frac{\alpha^\vartheta_\varepsilon}{\gamma_\varepsilon}+\frac{\alpha^2_\varepsilon}{\gamma^2_\varepsilon}+\frac{\alpha^2_\varepsilon}{\beta_\varepsilon}\right);
$$

(iii) (Regime 3) if $b,\sigma\in C_b^{\delta,1+\vartheta}$, $F,G\in C_p^{\vartheta/2,\delta,\vartheta}$, $H\in C_p^{(1+\vartheta)/2,\delta,1+\vartheta}$ with $\vartheta\in(0,1]$ and $c\in L_p^\infty$, then for every $\varphi\in C_b^{2+\vartheta}(\mathbb{R}^{d_2})$, we have

$$\sup_{t\in[0,T]}\Big|\mathbb{E}[\varphi(Y_t^\varepsilon)]-\mathbb{E}[\varphi(\hat{Y}_t^3)]\Big|\leqslant C_T\left(\alpha_\varepsilon^\vartheta+\frac{\alpha_\varepsilon}{\beta_\varepsilon}\right);$$

(iv) (Regime 4) if $b,\sigma\in C_b^{\delta,1+\vartheta}$, $F,G\in C_p^{\vartheta/2,\delta,\vartheta}$, $H\in C_p^{(1+\vartheta)/2,\delta,1+\vartheta}$ and $c\in C_p^{\delta,\vartheta}$ with $\vartheta\in(0,1]$, then for every $\varphi\in C_b^{2+\vartheta}(\mathbb{R}^{d_2})$, we have

$$\sup_{t\in[0,T]}\Big|\mathbb{E}[\varphi(Y_t^\varepsilon)]-\mathbb{E}[\varphi(\hat{Y}_t^4)]\Big|\leqslant C_T\,\alpha_\varepsilon^\vartheta,$$

where for $k=1,\cdots,4$, $\hat{Y}_t^k$ are the unique weak solutions for SDE (2.9), and $C_T>0$ is a constant independent of δ,ε.

Remark 2.4. *Note that in each case, the convergence rates do not depend on the index δ. This suggests that the convergence in the averaging principle for system (1.4) relies only on the regularity of the coefficients with respect to the y (slow) variable, and does not depend on their regularity with respect to the x (fast) variable.*

Let us give more explanations on the convergence rates as well as the assumptions on the coefficients in the above result, which we think are rather sharp.

Remark 2.5. *(i) Note that when $c=H\equiv0$, we can take $\gamma_\varepsilon=\beta_\varepsilon\equiv1$. In this particular case, the above result of Regime 1 simplifies to: if $b,\sigma\in C_b^{\delta,\vartheta}$ and $F,G\in C_p^{\vartheta/2,\delta,\vartheta}$ with $\vartheta\in(0,2]$, then for every $\varphi\in C_b^{2+\vartheta}(\mathbb{R}^{d_2})$, we have*

$$\sup_{t\in[0,T]}\Big|\mathbb{E}[\varphi(Y_t^\varepsilon)]-\mathbb{E}[\varphi(\hat{Y}_t^1)]\Big|\leqslant C_T\,\alpha_\varepsilon^\vartheta.$$

This is known to be optimal when $\vartheta=2$. In the general case, the rate of convergence will be dominated by the fast term $\gamma_\varepsilon^{-1}\int_0^t H(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s$ in the slow equation. Thus, the extra assumption that $\lim_{\varepsilon\to0}\alpha_\varepsilon^\vartheta/\gamma_\varepsilon=0$ in Regime 1-2 with only Hölder coefficients is necessary and will be automatically satisfied when $\vartheta\geqslant1$. It is also interesting to note that we may have $\lim_{\varepsilon\to0}\alpha_\varepsilon/\beta_\varepsilon=+\infty$ in these two regimes.

(ii) It will be clear from our proof that the convergence rates of Y_t^ε to $\hat{Y}_t^k$ ($k=1,\cdots,4$) depend only on the regularities of the averaged coefficients in the limit equations. Thus the regularities of the coefficients in the original equation with respect to the fast variable do not play any role, which appears to be intuitively natural. Our assumptions seem to be the weakest in order to get the desired result. Let us explain this with Regime 4 when the coefficients are time-independent. In order to get the $\alpha_\varepsilon^\vartheta$-order convergence of Y_t^ε to $\hat{Y}_t^4$, we shall need $\hat{F}_4,\hat{G}_4\in C_b^\vartheta$. Thanks to the assumption that $b,\sigma\in C_b^{\delta,1+\vartheta}$, $H\in C_p^{\delta,1+\vartheta}$ and Theorem 2.1, we can get that $\Phi\in C_p^{2+\delta,1+\vartheta}$, which in turn means that $\nabla_y\Phi\in C_p^{\delta,\vartheta}$. This together with the

assumptions $F, c \in C_p^{\delta,\vartheta}$ and Lemma 3.2 yields that $\hat{F}_4 \in C_b^{\vartheta}$. Similar computations hold for $\hat{G}_4$ as well as Regime 1-Regime 3.

3. Poisson equation in the whole space

This section is devoted to study the Poisson equation (1.1) in the whole space. Note that formally, the solution u should have the following probability representation:

$$u(x,y) = \int_0^\infty \mathbb{E} f\big(X_t^y(x), y\big)\mathrm{d}t, \tag{3.1}$$

where $X_t^y(x)$ is the unique solution for the frozen SDE (1.7). Let $p_t(x,x';y)$ be the density function of $X_t^y(x)$ (which is also the unique fundamental solution for the operator $\mathscr{L}_0$). For simplify, we denote by $T_t f(x,y)$ the semigroup corresponding to $X_t^y(x)$, i.e.,

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

Unlike the previous publications, we do not focus on the differentiablity of the semigroup $T_t f$ with respect to the parameter y. Instead, we shall first prove the optimal regularity estimates (2.1) and (2.2) for the solution u with respect to the x-variable, then we use an induction argument to show that the smoothness estimate (2.3) with respect to the parameter y follows directly. Throughout this section, we shall always assume that $(\mathbf{A}_\sigma)$-$(\mathbf{A}_b)$ hold.

In order to study the optimal regularity for equation (1.1), let us first collect some classical results concerning the properties for $p_t(x,x';y)$.

Lemma 3.1. *Assume $(\mathbf{A}_\sigma)$ holds and $T>0$. 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*

$$|\nabla_x^\ell p_t(x,x';y)| \leqslant C_T t^{-(d+\ell)/2} \exp\big(-c_0|x-x'|^2/t\big), \tag{3.2}$$

and

$$\left|\int_{\mathbb{R}^{d_1}} \nabla_x^2 p_t(x,x';y)\mathrm{d}x'\right| \leqslant C_T t^{-(2-\delta)/2}, \tag{3.3}$$

and for every $x_1, x_2 \in \mathbb{R}^{d_1}$,

$$\begin{aligned} |\nabla_x^2 p_t(x_1,x';y) - \nabla_x^2 p_t(x_2,x';y)| \leqslant C_T \Big[&\big(|x_1-x_2|\wedge 1\big) t^{-(d+3)/2} + \\ \big(|x_1-x_2|^\delta \wedge 1\big) t^{-(d+2)/2}\Big] &\Big(\exp\big(-c_0|x_1-x'|^2/t\big) + \exp\big(-c_0|x_2-x'|^2/t\big)\Big), \end{aligned} \tag{3.4}$$

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

If we further assume $(\mathbf{A}_b)$ *holds, then the limit*

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

exists, and for every $k,j \in \mathbb{R}_+$*, there exists a constant* $m > 0$ *such that for any* $t \geqslant 1$*,* $x, x' \in \mathbb{R}^{d_1}$ *and* $y \in \mathbb{R}^{d_2}$*,*

$$|p_\infty(x',y)| \leqslant \frac{C_0}{1+|x'|^j}, \tag{3.5}$$

and

$$|p_t(x,x';y) - p_\infty(x',y)| \leqslant C_0 \frac{1+|x|^m}{(1+t)^k(1+|x'|^j)}. \tag{3.6}$$

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

Proof. Estimate (3.2) is well-known, see e.g. [33, Chapter IV, §13, (13.1)], while estimate (3.3) can be found in [33, Chapter IV, §14, (14.2)]. When $|x_1 - x_2| \leqslant 1$, estimate (3.4) follows by [33, Chapter IV, §13, (13.2)]. While for $|x_1 - x_2| > 1$, the conclusion follows easily by (3.2). Finally, estimates (3.5) and (3.6) were given in [43, Proposition 3]. □

To shorten the notation, we will write for $\ell \in \mathbb{N}^*$,

$$\frac{\partial^\ell \mathscr{L}_0}{\partial y^\ell}(x,y) := \sum_{i,j=1}^{d_1} \partial_y^\ell a^{ij}(x,y)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d_1} \partial_y^\ell b^i(x,y)\frac{\partial}{\partial x_i}.$$

Given a function $h(x,y)$ on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, we shall denote by $\bar{h}(y)$ its average with respect to the measure $\mu^y(\mathrm{d}x)$, i.e.,

$$\bar{h}(y) := \int_{\mathbb{R}^{d_1}} h(x,y)\mu^y(\mathrm{d}x). \tag{3.7}$$

Then it is easy to see that

$$f(x,y) := h(x,y) - \bar{h}(y) \tag{3.8}$$

satisfies the centering condition (1.3). The following result will play an important role in the study of the smoothness of the solution to the Poisson equation with respect to the parameter y as well as the diffusion approximations for SDE (1.4).

Lemma 3.2. *Let* $(\mathbf{A}_\sigma)$ *and* $(\mathbf{A}_b)$ *hold,* $a, b \in C_b^{\delta,\eta}$ *with* $0 < \delta \leqslant 1$ *and* $\eta > 0$*. Given a function* $h \in C_p^{\delta,\eta}$*, let* f *be defined by (3.8). Assume that there exists a solution* $u \in C_p^{2+\delta,(\eta-1)\vee 0}$ *to the Poison equation (1.1). Then,*

(i) if $\eta \in \mathbb{N}^*$*, we have*

$$\partial_y^\eta \bar{h}(y) = \int_{\mathbb{R}^{d_1}} \Big[\partial_y^\eta h(x,y) - \sum_{\ell=1}^{\eta} C_\eta^\ell \cdot \frac{\partial^\ell \mathscr{L}_0}{\partial y^\ell}(x,y)\partial_y^{\eta-\ell}u(x,y)\Big]\mu^y(\mathrm{d}x); \tag{3.9}$$

(ii) if $\eta \in (0,\infty)\setminus \mathbb{N}^*$, we have that for any $y_1, y_2 \in \mathbb{R}^{d_2}$,

$$\begin{aligned}\partial_y^{[\eta]}\bar{h}(y_1) - \partial_y^{[\eta]}\bar{h}(y_2) = \int_{\mathbb{R}^{d_1}}\Bigg(&\big[\partial_y^{[\eta]}h(x,y_1) - \partial_y^{[\eta]}h(x,y_2)\big] \\ &- \sum_{\ell=1}^{[\eta]} C_{[\eta]}^{\ell}\cdot\frac{\partial^\ell \mathscr{L}_0}{\partial y^\ell}(x,y_2)\big[\partial_y^{[\eta]-\ell}u(x,y_1) - \partial_y^{[\eta]-\ell}u(x,y_2)\big] \\ &- \sum_{\ell=1}^{[\eta]} C_{[\eta]}^{\ell}\Big[\frac{\partial^\ell \mathscr{L}_0}{\partial y^\ell}(x,y_1) - \frac{\partial^\ell \mathscr{L}_0}{\partial y^\ell}(x,y_2)\Big]\partial_y^{[\eta]-\ell}u(x,y_1) \\ &- \big[\mathscr{L}_0(x,y_1) - \mathscr{L}_0(x,y_2)\big]\partial_y^{[\eta]}u(x,y_2)\Bigg)\mu^{y_1}(\mathrm{d}x).\end{aligned} \tag{3.10}$$

In particular, we have $\bar{h} \in C_b^{\eta}(\mathbb{R}^{d_2})$.

Remark 3.3. *(i) The above result together with Theorem 2.1 provide a useful tool to verify the regularity the averaged coefficients, which is a separate problem that will always encounter in the study of averaging principle, CLT, homogenization and other limit theorems. Thus, this lemma is of independent interest.*

(ii) Note that the left hand sides of the equality (3.9) and (3.10) involve η-order 'derivatives' of $\bar{h}$ with respect to the y-variable, while the right hand sides only involve at most $0\vee(\eta-1)$-order 'derivatives' of the solution u with respect to the parameter y. Hence these equality can be viewed as two transfer formulas, which transfer the smoothness with respect to the parameter y to the two derivatives with respect to the x variable. Let us explain this more clearly when $\eta = 1$. In this case, the above conclusion simplifies to: if we have $u \in C_p^{2+\delta,0}$, then

$$\partial_y\left(\int_{\mathbb{R}^{d_1}} h(x,y)\mu^y(\mathrm{d}x)\right) = \int_{\mathbb{R}^{d_1}}\Big[\partial_y h(x,y) - \frac{\partial \mathscr{L}_0}{\partial y}(x,y)u(x,y)\Big]\mu^y(\mathrm{d}x).$$

In order to derive the differentiablity of $\bar{h}(y)$, the left hand side of the above equality implies that we need to study the derivative of the invariant measure $\mu^y(\mathrm{d}x)$ with respect to the parameter y, which is usually difficult to obtain. However, the right hand side of the above equality only needs two derivatives of u with respect to the x variable (no derivative with respect to the parameter y is involved), which is quite classical due to the elliptic property of the operator $\mathscr{L}_0$.

Proof. We only need to prove the equality (3.9) and (3.10). Then the assertion that $\bar{h} \in C_b^{\eta}(\mathbb{R}^{d_2})$ follows directly. In fact, if (3.9) is true for $\eta \in \mathbb{N}^*$, we can derive by the assumptions $a, b \in C_b^{\delta,\eta}$, $h \in C_p^{\delta,\eta}$ and $u \in C_p^{2+\delta,(\eta-1)\vee 0}$ that for some $m > 0$,

$$\partial_y^{\eta}\bar{h}(y) \leqslant C_0\int_{\mathbb{R}^{d_1}}(1+|x|^m)\mu^y(\mathrm{d}x) \leqslant C_1 < \infty,$$

while if (3.10) is true for $\eta \in (0,\infty)\setminus\mathbb{N}^*$, we can derive that for $y_1, y_2 \in \mathbb{R}^{d_2}$,

$$\begin{aligned}\left|\partial_y^{[\eta]}\bar{h}(y_1)-\partial_y^{[\eta]}\bar{h}(y_2)\right| &\leqslant C_2\Big(|y_1-y_2|^{\eta-[\eta]}\wedge 1\Big)\int_{\mathbb{R}^{d_1}}(1+|x|^m)\mu^{y_1}(\mathrm{d}x)\\ &\leqslant C_3\Big(|y_1-y_2|^{\eta-[\eta]}\wedge 1\Big).\end{aligned}$$

Below, we divide the proof into two steps.
(i) Let us first prove (3.9) with $\eta = 1$. In fact, by the chain rule we can write

$$\begin{aligned}\partial_y\bar{h}(y) &= \partial_y\left(\int_{\mathbb{R}^{d_1}} h(x,y)\mu^y(\mathrm{d}x)\right)\\ &= \int_{\mathbb{R}^{d_1}}\partial_y h(x,y)\mu^y(\mathrm{d}x)+\int_{\mathbb{R}^{d_1}} h(x,y)\partial_y p_\infty(x,y)\mathrm{d}x.\end{aligned}$$

Then we use a formula established in [43, (28)] which yields

$$\begin{aligned}&\int_{\mathbb{R}^{d_1}} h(x,y)\partial_y p_\infty(x,y)\mathrm{d}x\\ &= -\int_{\mathbb{R}^{d_1}} h(x,y)\left(\int_0^\infty\int_{\mathbb{R}^{d_1}} p_\infty(x',y)\frac{\partial\mathscr{L}_0}{\partial y}(x',y)p_s(x',x;y)\mathrm{d}x'\mathrm{d}s\right)\mathrm{d}x.\end{aligned}$$

As a result, by (3.1), Fubini's theorem and the fact that $p_t(x',x;y)$ is a density function, we deduce that

$$\begin{aligned}&\int_{\mathbb{R}^{d_1}} h(x,y)\partial_y p_\infty(x,y)\mathrm{d}x\\ &= -\int_{\mathbb{R}^{d_1}}\frac{\partial\mathscr{L}_0}{\partial y}(x',y)\left(\int_0^\infty\int_{\mathbb{R}^{d_1}} p_s(x',x;y)h(x,y)\mathrm{d}x\mathrm{d}s\right)p_\infty(x',y)\mathrm{d}x'\\ &= -\int_{\mathbb{R}^{d_1}}\frac{\partial\mathscr{L}_0}{\partial y}(x',y)\left(\int_0^\infty\int_{\mathbb{R}^{d_1}} p_s(x',x;y)f(x,y)\mathrm{d}x\mathrm{d}s\right)p_\infty(x',y)\mathrm{d}x'\\ &= -\int_{\mathbb{R}^{d_1}}\frac{\partial\mathscr{L}_0}{\partial y}(x',y)\left(\int_0^\infty T_s f(x',y)\mathrm{d}s\right)p_\infty(x',y)\mathrm{d}x'\\ &= -\int_{\mathbb{R}^{d_1}}\frac{\partial\mathscr{L}_0}{\partial y}(x',y)u(x',y)\mu^y(\mathrm{d}x'),\end{aligned}$$

which in turn implies (3.9) is true for $\eta = 1$. The general case that $\eta \in \mathbb{N}^*$ can be proved by using formula [43, (34)] in stead of [43, (28)] and the induction argument, we omit the details here.
(ii) Now we prove (3.10) when $\eta \in (0,1)$. For any $y_1, y_2 \in \mathbb{R}^{d_2}$, we write

$$\bar{h}(y_1)-\bar{h}(y_2) = \int_{\mathbb{R}^{d_1}} h(x,y_1)\mu^{y_1}(\mathrm{d}x)-\int_{\mathbb{R}^{d_1}} h(x,y_2)\mu^{y_2}(\mathrm{d}x)$$

$$= \int_{\mathbb{R}^{d_1}} \big[h(x,y_1) - h(x,y_2)\big]\mu^{y_1}(\mathrm{d}x) + \int_{\mathbb{R}^{d_1}} h(x,y_2)\big[\mu^{y_1}(\mathrm{d}x) - \mu^{y_2}(\mathrm{d}x)\big].$$

Then we deduce by [47, Lemma 4.1] (which is similar in spirit of [43, (28)]) and the same argument as above that

$$\begin{aligned}
&\int_{\mathbb{R}^{d_1}} h(x,y_2)\big[\mu^{y_1}(\mathrm{d}x) - \mu^{y_2}(\mathrm{d}x)\big] \\
&= -\int_0^\infty \int_{\mathbb{R}^{d_1}} p_\infty(x';y_1)\big[\mathscr{L}_0(x',y_1) - \mathscr{L}_0(x',y_2)\big]\left(\int_{\mathbb{R}^{d_1}} p_s(x',x;y_2)h(x,y_2)\mathrm{d}x\right)\mathrm{d}x'\mathrm{d}s \\
&= -\int_0^\infty \int_{\mathbb{R}^{d_1}} p_\infty(x';y_1)\big[\mathscr{L}_0(x',y_1) - \mathscr{L}_0(x',y_2)\big]\left(\int_{\mathbb{R}^{d_1}} p_s(x',x;y)f(x,y_2)\mathrm{d}x\right)\mathrm{d}x'\mathrm{d}s \\
&= -\int_{\mathbb{R}^{d_1}} \big[\mathscr{L}_0(x',y_1) - \mathscr{L}_0(x',y_2)\big]u(x',y_2)\mu^{y_1}(\mathrm{d}x'),
\end{aligned}$$

where in the second equality we also used the fact that $p_t(x',x;y)$ is a density function. Thus, (3.10) is true for $\eta \in (0,1)$. The general case that $\eta \in (n,n+1)$ for some $n \in \mathbb{N}^*$ can be proved by using (3.9) with $\eta = n$ and the same arguments as before. The proof is finished. □

Now, we are in the position to give:

Proof of Theorem 2.1. We divide the proof into four steps.

Step 1. In this step we prove estimate (2.1). It suffices to consider the estimate for the second order derivative $\nabla_x^2 u$. To this end, we rewrite (3.1) as

$$u(x,y) = \int_0^2 T_t f(x,y)\mathrm{d}t + \int_2^\infty T_t f(x,y)\mathrm{d}t =: u_1(x,y) + u_2(x,y). \tag{3.11}$$

When $0 < t \leqslant 2$, using (3.2) and the fact that $p_t(x,x';y)$ is a density function, we deduce

$$\begin{aligned}
\nabla_x^2 T_t f(x,y) &= \int_{\mathbb{R}^{d_1}} \nabla_x^2 p_t(x,x';y)\big[f(x',y) - f(x,y)\big]\mathrm{d}x' \\
&\leqslant C_1[f]_{C_p^{\delta,0}} \int_{\mathbb{R}^{d_1}} \big(|x-x'|^\delta \wedge 1\big)\big(1 + |x|^m + |x'|^m\big) \\
&\qquad\qquad \times t^{-(d+2)/2} \exp\big(-c_0|x-x'|^2/t\big)\mathrm{d}x' \\
&\leqslant C_1[f]_{C_p^{\delta,0}}(1+|x|^m) \int_{\mathbb{R}^{d_1}} t^{-(d+2-\delta)/2} \exp\big(-c_1|x-x'|^2/t\big)\mathrm{d}x' \\
&\leqslant C_1[f]_{C_p^{\delta,0}}(1+|x|^m) \cdot t^{(\delta-2)/2},
\end{aligned}$$

while for $t > 2$, we have by the semigroup property that

$$
\begin{aligned}
\nabla_x^2 T_t f(x,y) &= \int_{\mathbb{R}^{d_1}} \int_{\mathbb{R}^{d_1}} \nabla_x^2 p_1(x,z;y) p_{t-1}(z,x';y) \mathrm{d}z f(x',y) \mathrm{d}x' \\
&= \int_{\mathbb{R}^{d_1}} \int_{\mathbb{R}^{d_1}} \nabla_x^2 p_1(x,z;y) \big[p_{t-1}(z,x';y) - p_\infty(x';y)\big] \mathrm{d}z f(x',y) \mathrm{d}x' \\
&\leqslant C_2 [f]_{C_p^{\delta,0}} \int_{\mathbb{R}^{d_1}} \int_{\mathbb{R}^{d_1}} \exp\big(-c_0|x-z|^2\big) \frac{(1+|z|^m)(1+|x'|^m)}{(1+t)^k(1+|x'|^j)} \mathrm{d}z \mathrm{d}x' \\
&\leqslant C_2 [f]_{C_p^{\delta,0}} \frac{(1+|x|^m)}{(1+t)^k} \int_{\mathbb{R}^{d_1}} \frac{(1+|x'|^m)}{(1+|x'|^j)} \mathrm{d}x' \\
&\leqslant C_2 [f]_{C_p^{\delta,0}} \frac{(1+|x|^m)}{(1+t)^k},
\end{aligned}
$$

where in the third inequality we have used (3.6), and we choose $j > d_1 + m$ in the last inequality. Now, taking these two estimates back into (3.11) gives

$$
\begin{aligned}
|\nabla_x^2 u(x,y)| &\leqslant C_3 [f]_{C_p^{\delta,0}} (1+|x|^m) \bigg(\int_0^2 t^{(\delta-2)/2} \mathrm{d}t + \int_2^\infty \frac{1}{1+t^k} \mathrm{d}t \bigg) \\
&\leqslant C_3 [f]_{C_p^{\delta,0}} (1+|x|^m).
\end{aligned}
$$

Thus estimate (2.1) is true.

Step 2. We proceed to prove estimate (2.2). Note that when $|x_1 - x_2| \geqslant 1$, the conclusion follows directly from (2.1). Below, we focus on the case where $|x_1 - x_2| < 1$. Let $\tilde{x}$ be one of the two points x_1 and x_2 which is nearer to x', and denote by $\mathcal{O}$ the ball with center $\tilde{x}$ and radius $2|x_1 - x_2|$. Without loss of generality, we may assume that $\tilde{x} = x_1$. We write

$$
\begin{aligned}
\nabla_x^2 u_1(x_1,y) - \nabla_x^2 u_1(x_2,y) &= \int_0^2 \int_{\mathcal{O}} \nabla_x^2 p_t(x_1,x';y) \big[f(x',y) - f(x_1,y)\big] \mathrm{d}x' \mathrm{d}t \\
&\quad - \int_0^2 \int_{\mathcal{O}} \nabla_x^2 p_t(x_2,x';y) \big[f(x',y) - f(x_2,y)\big] \mathrm{d}x' \mathrm{d}t \\
&\quad + \int_0^2 \int_{\mathbb{R}^{d_1}\setminus\mathcal{O}} \big[\nabla_x^2 p_t(x_1,x';y) - \nabla_x^2 p_t(x_2,x';y)\big] \big[f(x',y) - f(x_1,y)\big] \mathrm{d}x' \mathrm{d}t \\
&\quad + \big[f(x_2,y) - f(x_1,y)\big] \int_0^2 \int_{\mathbb{R}^{d_1}\setminus\mathcal{O}} \nabla_x^2 p_t(x_2,x';y) \mathrm{d}x' \mathrm{d}t =: \sum_{i=1}^4 \mathcal{I}_i.
\end{aligned}
$$

For the first term, we have by (3.2) that

$$
\mathcal{I}_1 \leqslant C_1 [f]_{C_p^{\delta,0}} \int_0^2 \int_{\mathcal{O}} t^{-(d+2)/2} \exp\big(-c_0|x_1 - x'|^2/t\big) |x' - x_1|^\delta \big(1 + |x_1|^m + |x'|^m\big) \mathrm{d}x' \mathrm{d}t
$$

$$\leqslant C_1[f]_{C_p^{\delta,0}}(1+|x_1|^m)\int_{\mathcal{O}}\int_0^2 t^{-(d+2-\delta)/2}\exp\big(-c_1|x_1-x'|^2/t\big)\mathrm{d}t\mathrm{d}x'$$
$$\leqslant C_1[f]_{C_p^{\delta,0}}(1+|x_1|^m)\int_{\mathcal{O}}|x_1-x'|^{-d+\delta}\mathrm{d}x' \leqslant C_1[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta(1+|x_1|^m).$$

In completely the same way we get

$$\mathcal{I}_2 \leqslant C_2[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta(1+|x_2|^m).$$

To control the third term, note that for $x' \in \mathbb{R}^{d_1}\setminus\mathcal{O}$, we have

$$|x_2-x'|/2 \leqslant |x_1-x'| \leqslant 3|x_2-x'|/2.$$

As a result, by (3.4) we deduce that

$$\begin{aligned}\mathcal{I}_3 \leqslant{}& C_3[f]_{C_p^{\delta,0}}|x_1-x_2|\int_0^2\int_{\mathbb{R}^{d_1}\setminus\mathcal{O}} t^{-(d+3)/2}\exp\big(-c_0|x_1-x'|^2/t\big)\\ &\qquad\times|x_1-x'|^\delta\big(1+|x_1|^m+|x'|^m\big)\mathrm{d}x'\mathrm{d}t\\ &+C_3[f]_{C_p^{\delta,0}}|x_1-x_2|^\beta\int_0^2\int_{\mathbb{R}^{d_1}\setminus\mathcal{O}} t^{-(d+2)/2}\exp\big(-c_0|x_1-x'|^2/t\big)\\ &\qquad\times|x_1-x'|^\delta\big(1+|x_1|^m+|x'|^m\big)\mathrm{d}x'\mathrm{d}t =: \mathcal{I}_{31}+\mathcal{I}_{32}.\end{aligned}$$

We further control $\mathcal{I}_{31}$ by

$$\begin{aligned}\mathcal{I}_{31} \leqslant{}& C_3[f]_{C_p^{\delta,0}}|x_1-x_2|(1+|x_1|^m)\\ &\times\int_{\mathbb{R}^{d_1}\setminus\mathcal{O}}\bigg(\int_0^2 t^{-(d+3)/2}\exp\big(-c_1|x_1-x'|^2/t\big)\mathrm{d}t\bigg)|x_1-x'|^\delta\mathrm{d}x'\\ \leqslant{}& C_3[f]_{C_p^{\delta,0}}|x_1-x_2|(1+|x_1|^m)\int_{\mathbb{R}^{d_1}\setminus\mathcal{O}}|x-x'|^{-d-1+\delta}\mathrm{d}x'\\ \leqslant{}& C_3[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta(1+|x_1|^m),\end{aligned}$$

and it is easy to check that

$$\begin{aligned}\mathcal{I}_{32} \leqslant{}& C_3[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta(1+|x_1|^m)\\ &\times\int_0^2\int_{\mathbb{R}^{d_1}} t^{-(d+2)/2}\exp\big(-c_2|x_1-x'|^2/t\big)|x_1-x'|^\delta\mathrm{d}x'\mathrm{d}t\\ \leqslant{}& C_3[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta(1+|x_1|^m)\int_0^2 t^{\delta/2-1}\mathrm{d}t \leqslant C_3[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta(1+|x_1|^m).\end{aligned}$$

Finally, thanks to (3.3), we have

$$\mathcal{I}_4 \leqslant C_4[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta(1+|x_1|^m+|x_2|^m)\int_0^2 t^{\delta/2-1}\mathrm{d}t$$

$$\leqslant C_4[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta(1+|x_1|^m+|x_2|^m).$$

Combining the above computations, we arrive at

$$|\nabla_x^2 u_1(x_1,y)-\nabla_x^2 u_1(x_2,y)|\leqslant C_5[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta(1+|x_1|^m+|x_2|^m).$$

On the other hand, we derive as in Step 1 that for $t\geqslant 2$,

$$\begin{aligned}\nabla_x^2 T_t f(x_1,y)-\nabla_x^2 T_t f(x_2,y)&=\int_{\mathbb{R}^{d_1}}\int_{\mathbb{R}^{d_1}}\big[\nabla_x^2 p_1(x_1,z;y)\\&\qquad-\nabla_x^2 p_1(x_2,z;y)\big]p_{t-1}(z,x';y)\mathrm{d}zf(x',y)\mathrm{d}x'\\
&=\int_{\mathbb{R}^{d_1}}\int_{\mathbb{R}^{d_1}}\big[\nabla_x^2 p_1(x_1,z;y)-\nabla_x^2 p_1(x_2,z;y)\big]\big[p_{t-1}(z,x';y)-p_\infty(x';y)\big]\mathrm{d}zf(x',y)\mathrm{d}x'\\
&\leqslant C_6[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta\int_{\mathbb{R}^{d_1}}\int_{\mathbb{R}^{d_1}}\exp\big(-c_0|x_1-z|^2\big)\frac{(1+|z|^m)(1+|x'|^m)}{(1+t)^k(1+|x'|^j)}\mathrm{d}z\mathrm{d}x'\\
&\leqslant C_6[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta\frac{(1+|x_1|^m)}{(1+t)^k},\end{aligned}$$

which in turn yields

$$\begin{aligned}|\nabla_x^2 u_2(x_1,y)-\nabla_x^2 u_2(x_2,y)|&\leqslant C_6[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta\int_2^\infty\frac{(1+|x_1|^m)}{(1+t)^k}\mathrm{d}t\\&\leqslant C_6[f]_{C_p^{\delta,0}}|x_1-x_2|^\delta(1+|x_1|^m).\end{aligned}$$

Now using (3.11) again, we get (2.2).

Step 3. In this step, we prove estimate (2.3) when $\eta\in\mathbb{N}^*$. We shall only focus on the a priori estimates. Let us first consider the case $\eta=1$. We start form the equation itself, i.e., u is a classical solution to

$$\mathscr{L}_0(x,y)u(x,y)=f(x,y).$$

Taking partial derivative with respect to the y variable from both sides of the equation, we get that

$$\mathscr{L}_0(x,y)\partial_y u(x,y)=\partial_y f(x,y)-\frac{\partial\mathscr{L}_0}{\partial y}(x,y)u(x,y).\tag{3.12}$$

According to (3.9), the right hand side of (3.12) satisfies the centering condition (1.3), i.e.,

$$\int_{\mathbb{R}^{d_1}}\Big[\partial_y f(x,y)-\frac{\partial\mathscr{L}_0}{\partial y}(x,y)u(x,y)\Big]\mu^y(\mathrm{d}x)=\partial_y\left(\int_{\mathbb{R}^{d_1}}f(x,y)\mu^y(\mathrm{d}x)\right)=0.$$

Moreover, by assumption we have that $\partial_y f\in C_p^{\delta,0}$. Meanwhile, due to the assumptions $a,b\in C_b^{\delta,1}$ and in view of the a priori estimates (2.1)-(2.2), it is easily checked that for any $x_1, x_2 \in \mathbb{R}^{d_1}$ and $y \in \mathbb{R}^{d_2}$, there exists a constant $C_1 > 0$ such that

$$
\begin{aligned}
&\Big|\frac{\partial \mathscr{L}_0}{\partial y}(x_1, y)u(x_1, y) - \frac{\partial \mathscr{L}_0}{\partial y}(x_2, y)u(x_2, y)\Big| \\
&\leqslant C_1\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)\|f\|_{C_p^{\delta,0}}\big(|x_1 - x_2|^\delta \wedge 1\big)(1 + |x_1|^m + |x_2|^m).
\end{aligned}
$$

As a result, we have $\frac{\partial \mathscr{L}_0}{\partial y} u \in C_p^{\delta,0}$ with

$$
\big[\tfrac{\partial \mathscr{L}_0}{\partial y} u\big]_{C_p^{\delta,0}} \leqslant C_1\Big(\|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)\|f\|_{C_p^{\delta,0}}.
$$

Thus, using the conclusions proved in Step 1 for the function $\partial_y u$, we get that $\partial_y u \in C_p^{2+\delta,0}$ and by the a priori estimate (2.1), we have

$$
\begin{aligned}
|\partial_y u(x, y)| &\leqslant C_2\Big([\partial_y f]_{C_p^{\delta,0}} + \big[\tfrac{\partial \mathscr{L}_0}{\partial y} u\big]_{C_p^{\delta,0}}\Big)(1 + |x|^m), \\
&\leqslant C_2\Big([f]_{C_p^{\delta,1}} + \|a\|_{C_b^{\delta,1}} + \|b\|_{C_b^{\delta,1}}\Big)(1 + |x|^m),
\end{aligned}
$$

which in particular yields (2.3) for $\eta = 1$. Suppose that (2.3) holds for some $\eta = n$ with constant $\mathcal{K}_\eta$ given by (2.4). By induction we find that for $\eta = n + 1$,

$$
\mathscr{L}_0(x, y)\partial_y^{n+1} u(x, y) = \partial_y^{n+1} f(x, y) - \sum_{\ell=1}^{n+1} C_{n+1}^\ell \cdot \frac{\partial^\ell \mathscr{L}_0}{\partial y^\ell}(x, y)\partial_y^{n+1-\ell} u(x, y).
$$

According to (3.9), the right hand side of the above equality satisfies the centering condition (1.3), i.e.,

$$
\begin{aligned}
&\int_{\mathbb{R}^{d_1}} \Big[\partial_y^{n+1} f(x, y) - \sum_{\ell=1}^{n+1} C_{n+1}^\ell \cdot \frac{\partial^\ell \mathscr{L}_0}{\partial y^\ell}(x, y)\partial_y^{n+1-\ell} u(x, y)\Big]\mu^y(\mathrm{d}x) \\
&= \partial_y^{n+1}\left(\int_{\mathbb{R}^{d_1}} f(x, y)\mu^y(\mathrm{d}x)\right) = 0.
\end{aligned}
$$

It then follows by using (2.1) again for the function $\partial_y^{n+1} u$ and the induction assumption that

$$
\begin{aligned}
|\partial_y^{n+1} u(x, y)| &\leqslant C_3\Big([\partial_y^{n+1} f]_{C_p^{\delta,0}} + \sum_{\ell=1}^{n+1} C_{n+1}^\ell \cdot \kappa_\ell\big[\partial_y^{n+1-\ell} u\big]_{C_p^{\delta,0}}\Big)(1 + |x|^m) \\
&\leqslant C_3\Big([f]_{C_p^{\delta,n+1}} + \sum_{\ell=1}^{n+1} C_{n+1}^\ell \cdot \kappa_\ell \cdot \mathcal{K}_{n+1-\ell}\Big)(1 + |x|^m),
\end{aligned}
$$

which means (2.3) is true for $\eta = n + 1$.

Step 4. Finally, we prove (2.3) when $\eta \in (0,\infty)\setminus\mathbb{N}^*$. Let us first consider the case $\eta\in(0,1)$. By the equation (2.8), we write for any $x\in\mathbb{R}^{d_1}$ and $y_1,y_2\in\mathbb{R}^{d_2}$

$$\begin{aligned}\mathscr{L}_0(x,y_1)\big[u(x,y_1)-u(x,y_2)\big] &= \big[f(x,y_1)-f(x,y_2)\big]\\ &\quad+\big[\mathscr{L}_0(x,y_2)-\mathscr{L}_0(x,y_1)\big]u(x,y_2). \end{aligned}\tag{3.13}$$

According to (3.10), the right hand side of (3.13) satisfies the centering condition (1.3), i.e.,

$$\begin{aligned}&\int_{\mathbb{R}^{d_1}}\Big(\big[f(x,y_1)-f(x,y_2)\big]+\big[\mathscr{L}_0(x,y_2)-\mathscr{L}_0(x,y_1)\big]u(x,y_2)\Big)\mu^{y_1}(\mathrm{d}x)\\ &=\int_{\mathbb{R}^{d_1}}f(x,y_1)\mu^{y_1}(\mathrm{d}x)-\int_{\mathbb{R}^{d_1}}f(x,y_2)\mu^{y_2}(\mathrm{d}x)=0.\end{aligned}$$

As a result, we can use (2.1) for the function $u(x,y_1)-u(x,y_2)$ and by the same argument as above to get that

$$\big|u(x,y_1)-u(x,y_2)\big|\leqslant C_1\big(|y_1-y_2|^\eta\wedge 1\big)\Big([f]_{C_p^{\delta,\eta}}+\|a\|_{C_b^{\delta,\eta}}+\|b\|_{C_b^{\delta,\eta}}\Big)(1+|x|^m),$$

which means that $u(x,\cdot)\in C_b^\eta$ and (2.3) is true. Assume that (2.3) holds for some $\eta\in(n,n+1)$ with constant $\mathcal{K}_\eta$ given by (2.5). Then for $\eta\in(n+1,n+2)$, by the conclusion proved in (ii), for any $x\in\mathbb{R}^{d_1}$ and $y_1,y_2\in\mathbb{R}^{d_2}$ we obtain

$$\begin{aligned}\mathscr{L}_0(x,y_1)\big[\partial_y^{n+1}u(x,y_1)-\partial_y^{n+1}u(x,y_2)\big] &= \big[\partial_y^{n+1}f(x,y_1)-\partial_y^{n+1}f(x,y_2)\big]\\ &\quad-\sum_{\ell=1}^{n+1}C_{n+1}^\ell\cdot\frac{\partial^\ell\mathscr{L}_0}{\partial y^\ell}(x,y_2)\big[\partial_y^{n+1-\ell}u(x,y_1)-\partial_y^{n+1-\ell}u(x,y_2)\big]\\ &\quad-\sum_{\ell=1}^{n+1}C_{n+1}^\ell\Big[\frac{\partial^\ell\mathscr{L}_0}{\partial y^\ell}(x,y_1)-\frac{\partial^\ell\mathscr{L}_0}{\partial y^\ell}(x,y_2)\Big]\partial_y^{n+1-\ell}u(x,y_1)\\ &\quad-\big[\mathscr{L}_0(x,y_1)-\mathscr{L}_0(x,y_2)\big]\partial_y^{n+1}u(x,y_2).\end{aligned}\tag{3.14}$$

Using (3.10) again we can see that the right hand side of (3.14) satisfies the centering condition. Consequently, we have

$$\begin{aligned}\big|\partial_y^{n+1}u(x,y_1)-\partial_y^{n+1}u(x,y_2)\big| &\leqslant C_2\big(|y_1-y_2|^\eta\wedge1\big)\Big([f]_{C_p^{\delta,\eta}}+\kappa_{\eta-n-1}[\partial_y^{n+1}u]_{C_p^{\delta,0}}\\ &\quad+\sum_{\ell=1}^{n+1}C_{n+1}^\ell\cdot\kappa_\ell[\partial_y^{\eta-\ell}u]_{C_p^{\delta,0}}+\sum_{\ell=1}^{n+1}C_{n+1}^\ell\cdot\kappa_{\eta+\ell-n-1}[\partial_y^{n+1-\ell}u]_{C_p^{\delta,0}}\Big)(1+|x|^m)\\ &\leqslant C_2\big(|y_1-y_2|^\eta\wedge1\big)\Big([f]_{C_p^{\delta,\eta}}+\kappa_{\eta-n-1}\mathcal{K}_{n+1}\\ &\quad+\sum_{\ell=1}^{n+1}C_{n+1}^\ell\cdot\kappa_\ell\mathcal{K}_{\eta-\ell}+\sum_{\ell=1}^{n+1}C_{n+1}^\ell\cdot\kappa_{\eta+\ell-n-1}\mathcal{K}_{n+1-\ell}\Big)(1+|x|^m)\end{aligned}$$

$$= C_2\mathcal{K}_\eta\big(|y_1-y_2|^\eta\wedge 1\big)(1+|x|^m),$$

which means (2.3) is true for $\eta\in(n+1,n+2)$. So, the proof is finished. $\square$

4. Fluctuation estimates

We shall first prepare some results concerning the mollifying approximation of functions in Subsection 4.1. Then we derive two new fluctuation estimates using the Poisson equation (1.1): one of functional LLN type in Subsection 4.2 and one of functional CLT type in Subsection 4.3.

4.1. **Mollifying approximation.** We need some mollification arguments due to our low regularity assumptions on the coefficients. To this end, 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 $k\geqslant 1$, there exist constants $C_k>0$ such that $|\nabla^k\rho_1(r)|\leqslant C_k\rho_1(r)$ and $|\nabla^k\rho_2(y)|\leqslant C_k\rho_2(y)$. For every $n\in\mathbb{N}^*$, set

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

Note that the scaling speed of ρ_2^n is n, while ρ_1^n has the speed n^2 in mollifying. Given a function $f(t,x,y)$, define the mollifying approximations of f in t and y variables by

$$f_n(t,x,y):=f*\rho_2^n*\rho_1^n:=\int_{\mathbb{R}^{d_2+1}}f(t-s,x,y-z)\rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s. \tag{4.1}$$

We have the following easy result.

Lemma 4.1. *Let $f\in C_p^{\vartheta/2,\delta,\vartheta}$ with $0<\vartheta\leqslant 2$, $0<\delta\leqslant 1$, and define f_n by (4.1). Then we have*

$$\|f(\cdot,x,\cdot)-f_n(\cdot,x,\cdot)\|_\infty\leqslant C_0n^{-\vartheta}(1+|x|^m), \tag{4.2}$$

and

$$\|\partial_tf_n(\cdot,x,\cdot)\|_\infty+\|\nabla_y^2f_n(\cdot,x,\cdot)\|_\infty\leqslant C_0n^{2-\vartheta}(1+|x|^m), \tag{4.3}$$

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

Proof. We divide the proof into two cases.
(i) (Case $0<\vartheta\leqslant 1$). By definition and a change of variable, there exists a constant $m>0$ such that

$$\begin{aligned}|f(t,x,y)-f_n(t,x,y)|&\leqslant\int_{\mathbb{R}^{d_2+1}}\big|f(t,x,y)-f(t-s,x,y-z)\big|\cdot\rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s\\&\leqslant C_0\int_{\mathbb{R}^{d_2+1}}\big(s^{\vartheta/2}+|z|^\vartheta\big)(1+|x|^m)\rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s\\&\leqslant C_0n^{-\vartheta}(1+|x|^m).\end{aligned}$$

Furthermore,

$$
\begin{aligned}
|\partial_t f_n(t,x,y)| &\leqslant \int_{\mathbb{R}^{d_2+1}} \big|f(t-s,x,y-z)-f(t,x,y-z)\big|\rho_2^n(z)|\partial_s\rho_1^n(s)|\mathrm{d}z\mathrm{d}s\\
&\leqslant C_0 n^2\int_{\mathbb{R}^{d_2+1}} s^{\vartheta/2}(1+|x|^m)\rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s \leqslant C_0 n^{2-\vartheta}(1+|x|^m),
\end{aligned}
$$

and

$$
\begin{aligned}
|\nabla_y^2 f_n(t,x,y)| &\leqslant \int_{\mathbb{R}^{d_2+1}} \big|f(t-s,x,y-z)-f(t-s,x,y)\big|\cdot|\nabla_z^2\rho_2^n(z)|\rho_1^n(s)\mathrm{d}z\mathrm{d}s\\
&\leqslant C_0 n^2\int_{\mathbb{R}^{d_2+1}} |z|^{\vartheta}(1+|x|^m)\rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s \leqslant C_0 n^{2-\vartheta}(1+|x|^m).
\end{aligned}
$$

(ii) (Case $1<\vartheta\leqslant 2$). By the symmetric property of ρ_1 and ρ_2, we can use the second order difference to deduce that

$$
\begin{aligned}
|f(t,x,y)-f_n(t,x,y)| &\leqslant \int_{\mathbb{R}^{d_2+1}} \big|f(t-s,x,y-z)+f(t-s,x,y+z)\\
&\qquad\qquad -2f(t,x,y)\big|\cdot\rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s\\
&\leqslant C_1\int_{\mathbb{R}^{d_2+1}} \big(s^{\vartheta/2}+|z|^{\vartheta}\big)(1+|x|^m)\cdot\rho_2^n(z)\rho_1^n(s)\mathrm{d}z\mathrm{d}s\\
&\leqslant C_1 n^{-\vartheta}(1+|x|^m),
\end{aligned}
$$

and

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

The estimate concerning the derivative with respect to the t variable can be proved similarly. The proof is finished. $\square$

4.2. **Fluctuation estimate - LLN type.** Given a function $h(t,x,y)$, recall that $f(t,x,y) := h(t,x,y)-\bar{h}(t,y)$ satisfies the centering condition (1.3), where $\bar{h}$ is defined by (3.7). The following result establishes the behavior of the fluctuation between $h(s,X_s^\varepsilon,Y_s^\varepsilon)$ and $\bar{h}(s,Y_s^\varepsilon)$ over the time interval $[0,t]$, which will play an important role below.

Lemma 4.2. *Let* $(\mathbf{A}_\sigma)$, $(\mathbf{A}_b)$ *and (2.7) hold true. Assume that* $b,\sigma\in C_b^{\delta,\vartheta}$ *with* $0<\delta,\vartheta\leqslant 2$, *and* $c,F,H,G\in L_p^\infty$. *Then for every* $f\in C_p^{\vartheta/2,\delta,\vartheta}$ *satisfying (1.3), we have*

$$
\mathbb{E}\left(\int_0^t f(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right)\leqslant C_t\Big(\alpha_\varepsilon^{\vartheta}+\alpha_\varepsilon^{\vartheta\wedge 1}\cdot\frac{\alpha_\varepsilon}{\gamma_\varepsilon}+\frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big),
$$

where $C_t > 0$ is a constant independent of δ, ε.

Remark 4.3. *If $0 < \vartheta \leqslant 1$, then the fluctuation is controlled by $\alpha_\varepsilon^\vartheta + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}$; and if $\vartheta = 2$, the error bound can be controlled by $\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}$; while when $1 < \vartheta \leqslant 2$, the fluctuation will depend on the balance between α_ε and γ_ε. In particular, in the case of Regime 3 and Regime 4, the bound simplifies to $\alpha_\varepsilon^{\vartheta \wedge 1}$.*

Proof. Since f satisfies (1.3), by Theorem 2.1 and in view of (3.1), there exists a unique solution $\Phi^f(t,x,y) \in C_p^{\vartheta/2, 2+\delta, \vartheta}$ to the following Poisson equation in $\mathbb{R}^{d_1}$:

$$\mathscr{L}_0(x,y)\Phi^f(t,x,y) = -f(t,x,y), \tag{4.4}$$

where $(t,y) \in \mathbb{R}_+ \times \mathbb{R}^{d_2}$ are parameters. Let Φ_n^f be the mollifyer of Φ^f defined as in (4.1). By Itô's formula, we get

$$\begin{aligned}\Phi_n^f(t, X_t^\varepsilon, Y_t^\varepsilon) = \Phi_n^f(0,x,y) + \int_0^t \Big(\partial_s + \beta_\varepsilon^{-1}\mathscr{L}_3 + \gamma_\varepsilon^{-1}\mathscr{L}_2 + \mathscr{L}_1\Big)\Phi_n^f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s \\ + \frac{1}{\alpha_\varepsilon^2}\int_0^t \mathscr{L}_0 \Phi_n^f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s + \frac{1}{\alpha_\varepsilon} M_n^1(t) + M_n^2(t),\end{aligned}$$

where $\mathscr{L}_3, \mathscr{L}_2$ and $\mathscr{L}_1$ are defined by (1.5), and for $i = 1,2$, $M_n^i(t)$ are martingales given by

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

and

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

This in turn yields that

$$\begin{aligned}\int_0^t f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s = \alpha_\varepsilon^2 \Phi_n^f(0,x,y) - \alpha_\varepsilon^2 \Phi_n^f(t, X_t^\varepsilon, Y_t^\varepsilon) + \alpha_\varepsilon M_n^1(t) + \alpha_\varepsilon^2 M_n^2(t) \\ + \alpha_\varepsilon^2 \int_0^t \big(\partial_s + \mathscr{L}_1\big)\Phi_n^f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s \\ + \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon}\int_0^t \mathscr{L}_2 \Phi_n^f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\int_0^t \mathscr{L}_3 \Phi_n^f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s \\ + \int_0^t \big(\mathscr{L}_0 \Phi_n^f - \mathscr{L}_0 \Phi^f\big)(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s.\end{aligned} \tag{4.5}$$

As a result, we have

$$\mathcal{Q}(\varepsilon) := \mathbb{E}\left(\int_0^t f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\right) \leqslant 2\alpha_\varepsilon^2 \mathbb{E}\|\Phi_n^f(\cdot, X_t^\varepsilon, \cdot)\|_\infty$$

$$
\begin{aligned}
&+\alpha_\varepsilon^2\mathbb{E}\left|\int_0^t \big(\partial_s+\mathscr{L}_1\big)\Phi_n^f(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right|\\
&+\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon}\mathbb{E}\left|\int_0^t \mathscr{L}_2\Phi_n^f(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right|+\frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\mathbb{E}\left|\int_0^t \mathscr{L}_3\Phi_n^f(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right|\\
&+\mathbb{E}\left|\int_0^t \big(\mathscr{L}_0\Phi_n^f-\mathscr{L}_0\Phi^f\big)(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right| =: \sum_{i=1}^5 \mathcal{Q}_i(\varepsilon).
\end{aligned}
$$

Note that the assumptions $(\mathbf{A}_\sigma)$ and (2.7) hold uniformly in y. Hence it follows by [52, Lemma 1] (see also [43, Lemma 2] or [48]) that for any $m>0$ and $\varepsilon\in(0,1/2)$,

$$
\mathbb{E}|X_t^\varepsilon|^m \leqslant C_0(1+|x|^m), \tag{4.6}
$$

where C_0 is a positive constant independent of ε. Hence, it follows by (2.1) that there exists a constant $C_1>0$ such that

$$
\mathcal{Q}_1(\varepsilon)\leqslant C_1\alpha_\varepsilon^2\mathbb{E}\|\Phi^f(\cdot,X_t^\varepsilon,\cdot)\|_\infty \leqslant C_1\alpha_\varepsilon^2\mathbb{E}\big(1+|X_t^\varepsilon|^m\big)\leqslant C_1\alpha_\varepsilon^2.
$$

To control the second term, we have by (4.3) and the assumption $F,G\in L_p^\infty$ that

$$
\begin{aligned}
\|\big(\partial_s+\mathscr{L}_1\big)\Phi_n^f(\cdot,x,\cdot)\|_\infty &\leqslant C_2\big(1+|x|^m\big)\\
&\times\left(\|\partial_s\Phi_n^f(\cdot,x,\cdot)\|_\infty+\sum_{\ell=1,2}\big\|\nabla_y^\ell\Phi_n^f(\cdot,x,\cdot)\big\|_\infty\right)\leqslant C_2 n^{2-\vartheta}(1+|x|^{2m}).
\end{aligned}
$$

Consequently, by (4.6)

$$
\mathcal{Q}_2(\varepsilon)\leqslant C_2\alpha_\varepsilon^2 n^{2-\vartheta}\mathbb{E}\left(\int_0^t\big(1+|X_s^\varepsilon|^{2m}\big)\mathrm{d}s\right)\leqslant C_2\alpha_\varepsilon^2 n^{2-\vartheta}.
$$

Following the same argument as above, we get that when $0<\vartheta\leqslant 1$,

$$
\|\mathscr{L}_2\Phi_n^f(\cdot,x,\cdot)\|_\infty\leqslant C_3\|\nabla_y\Phi_n^f(\cdot,x,\cdot)\|_\infty(1+|x|^m)\leqslant C_3 n^{1-\vartheta}(1+|x|^{2m}),
$$

while for $1<\vartheta\leqslant 2$, we have

$$
\|\mathscr{L}_2\Phi_n^f(\cdot,x,\cdot)\|_\infty\leqslant C_3\|\nabla_y\Phi^f(\cdot,x,\cdot)\|_\infty(1+|x|^m)\leqslant C_3(1+|x|^{2m}).
$$

Thus, we get

$$
\mathcal{Q}_3(\varepsilon)\leqslant C_3\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon}n^{1-(\vartheta\wedge 1)}\mathbb{E}\left(\int_0^t\big(1+|X_s^\varepsilon|^m\big)\mathrm{d}s\right)\leqslant C_3\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon}n^{1-(\vartheta\wedge 1)}.
$$

Meanwhile, since $c\in L_p^\infty$, it follows easily that for some $m>0$,

$$
\mathcal{Q}_4(\varepsilon)\leqslant C_4\frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\mathbb{E}\left(\int_0^t\big(1+|X_s^\varepsilon|^m\big)\mathrm{d}s\right)\leqslant C_4\frac{\alpha_\varepsilon^2}{\beta_\varepsilon}.
$$

Finally, since $\nabla_x^2\Phi^f\in C_p^{\vartheta/2,\delta,\vartheta}$ and by the fact that

$$
\nabla_x^2(\Phi_n^f)=(\nabla_x^2\Phi^f)*\rho_1^n*\rho_2^n,
$$

we can derive by (4.2) that

$$\begin{aligned}\mathcal{Q}_5(\varepsilon) &\leqslant C_5\big(\|a\|_\infty+\|b\|_\infty\big)\cdot\mathbb{E}\left(\int_0^t\sum_{\ell=1,2}\big\|(\nabla_x^\ell\Phi_n^f-\nabla_x^\ell\Phi^f)(\cdot,X_s^\varepsilon,\cdot)\big\|_\infty\mathrm{d}s\right)\\ &\leqslant C_5 n^{-\vartheta}\mathbb{E}\left(\int_0^t\big(1+|X_s^\varepsilon|^m\big)\mathrm{d}s\right)\leqslant C_5 n^{-\vartheta}.\end{aligned}$$

Combining the above computations, we arrive at

$$\mathcal{Q}(\varepsilon)\leqslant C_6\Big(\alpha_\varepsilon^2+\alpha_\varepsilon^2 n^{2-\vartheta}+\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon}n^{1-(\vartheta\wedge1)}+n^{-\vartheta}+\frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

Taking $n=\alpha_\varepsilon^{-1}$, we thus get

$$\mathcal{Q}(\varepsilon)\leqslant C_6\Big(\alpha_\varepsilon^\vartheta+\frac{\alpha_\varepsilon^{1+(\vartheta\wedge1)}}{\gamma_\varepsilon}+\frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

The proof is finished. □

4.3. **Fluctuation estimate - CLT type.** Now, we derive a CLT type fluctuation estimate for $f(s,X_s^\varepsilon,Y_s^\varepsilon)$ over the time interval $[0,t]$. This will depend on the orders how $\alpha_\varepsilon,\beta_\varepsilon,\gamma_\varepsilon$ go to zero described in Regime 1-Regime 4 of (1.8), and the limit behavior will involve the solution of the auxiliary Poisson equation.

Recall that Φ^f is the solution to the Poisson equation (4.4). For simplify, we denote by

$$\begin{aligned}\overline{c\cdot\nabla_x\Phi^f}(t,y)&:=\int_{\mathbb{R}^{d_1}}c(x,y)\cdot\nabla_x\Phi^f(t,x,y)\mu^y(\mathrm{d}x),\\ \overline{H\cdot\nabla_y\Phi^f}(t,y)&:=\int_{\mathbb{R}^{d_1}}H(t,x,y)\cdot\nabla_y\Phi^f(t,x,y)\mu^y(\mathrm{d}x).\end{aligned}$$

The following is the main result of this subsection, which will play a crucial role in the proof of Theorem 2.3.

Lemma 4.4. *Let* $(\mathbf{A}_\sigma)$, $(\mathbf{A}_b)$, *(2.7) hold and* $\delta\in(0,1]$. *For any function* f *satisfying (1.3), we have:*

(i) (Regime 1) if $b,\sigma\in C_b^{\delta,\vartheta}$ *with* $\vartheta\in(0,2]$, $c\in L_p^\infty$, $F,H,G\in L_p^\infty$ *and* $f\in C_p^{\vartheta/2,\delta,\vartheta}$, *then*

$$\mathbb{E}\left(\frac{1}{\gamma_\varepsilon}\int_0^t f(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right)\leqslant C_t\Big(\frac{\alpha_\varepsilon^\vartheta}{\gamma_\varepsilon}+\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2}+\frac{\alpha_\varepsilon^2}{\beta_\varepsilon\gamma_\varepsilon}\Big);$$

(ii) (Regime 2) if $b,\sigma\in C_b^{\delta,\vartheta}$ *with* $\vartheta\in(0,2]$, $c\in C_p^{\delta,\vartheta}$, $F,H,G\in L_p^\infty$ *and* $f\in C_p^{\vartheta/2,\delta,\vartheta}$, *then*

$$\mathbb{E}\left(\frac{1}{\gamma_\varepsilon}\int_0^t f(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right)-\mathbb{E}\left(\int_0^t\overline{c\cdot\nabla_x\Phi^f}(s,Y_s^\varepsilon)\mathrm{d}s\right)\leqslant C_t\Big(\frac{\alpha_\varepsilon^\vartheta}{\gamma_\varepsilon}+\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2}+\frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big);$$

(iii) (Regime 3) If $b, \sigma \in C_b^{\delta,1+\vartheta}$ with $\vartheta \in (0,1]$, $c \in L_p^\infty$, $F, G \in L_p^\infty$, $H \in C_p^{\vartheta/2,\delta,\vartheta}$ and $f \in C_p^{(1+\vartheta)/2,\delta,1+\vartheta}$, then

$$\mathbb{E}\left(\frac{1}{\gamma_\varepsilon}\int_0^t f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\right) - \mathbb{E}\left(\int_0^t \overline{H \cdot \nabla_y \Phi^f}(s, Y_s^\varepsilon)\mathrm{d}s\right) \leqslant C_t\Big(\alpha_\varepsilon^\vartheta + \frac{\alpha_\varepsilon}{\beta_\varepsilon}\Big),$$

(iv) (Regime 4) If $b, \sigma \in C_b^{\delta,1+\vartheta}$ with $\vartheta \in (0,1]$, $c \in C_p^{\delta,\vartheta}$, $F, G \in L_p^\infty$, $H \in C_p^{\vartheta/2,\delta,\vartheta}$ and $f \in C_p^{(1+\vartheta)/2,\delta,1+\vartheta}$, then

$$\mathbb{E}\left(\frac{1}{\gamma_\varepsilon}\int_0^t f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\right) - \mathbb{E}\left(\int_0^t \Big[\overline{c \cdot \nabla_x \Phi^f} + \overline{H \cdot \nabla_y \Phi^f}\Big](s, Y_s^\varepsilon)\mathrm{d}s\right) \leqslant C_t \alpha_\varepsilon^\vartheta,$$

where $C_t > 0$ is a constant independent on δ, ε.

Proof. As in the proof of Lemma 4.2, by (4.5) we have that

$$\begin{aligned}
\hat{\mathscr{Q}}(\varepsilon) &:= \mathbb{E}\left(\frac{1}{\gamma_\varepsilon}\int_0^t f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\right) = \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon}\mathbb{E}\big[\Phi_n^f(0,x,y) - \Phi_n^f(t, X_t^\varepsilon, Y_t^\varepsilon)\big] \\
&\quad + \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon}\mathbb{E}\left(\int_0^t \big(\partial_s + \mathscr{L}_1\big)\Phi_n^f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\right) \\
&\quad + \frac{1}{\gamma_\varepsilon}\mathbb{E}\left(\int_0^t \big(\mathscr{L}_0\Phi_n^f - \mathscr{L}_0\Phi^f\big)(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\right) \\
&\quad + \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2}\mathbb{E}\left(\int_0^t \mathscr{L}_2\Phi_n^f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\right) \\
&\quad + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon\gamma_\varepsilon}\mathbb{E}\left(\int_0^t \mathscr{L}_3\Phi_n^f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\right) =: \sum_{i=1}^5 \hat{\mathscr{Q}}_i(\varepsilon),
\end{aligned}$$

where Φ_n^f is the mollifyer of Φ^f defined as in (4.1). Below, we first prove the most general case *(iv)*, and then provide the proof of the other cases with sight changes.

(iv) (Regime 4) In this case, according to Theorem 2.1, (3.1) and by the assumptions on b, σ and f, we have $\Phi^f \in C_p^{(1+\vartheta)/2,2+\delta,1+\vartheta}$ with $\vartheta \in (0,1]$. Following exactly the same arguments as in Lemma 4.2, we get

$$\hat{\mathscr{Q}}_1(\varepsilon) \leqslant C_1\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon},$$

and by (4.3),

$$\hat{\mathscr{Q}}_2(\varepsilon) \leqslant C_2\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon}n^{2-(1+\vartheta)}\mathbb{E}\left(\int_0^t \big(1 + |X_s^\varepsilon|^{2m}\big)\mathrm{d}s\right) \leqslant C_2\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon}n^{1-\vartheta}.$$

Meanwhile, we have by (4.2) that

$$\hat{\mathscr{Q}}_3(\varepsilon) \leqslant C_3 \frac{1}{\gamma_\varepsilon} n^{-1-\vartheta} \mathbb{E}\left(\int_0^t \big(1+|X_s^\varepsilon|^{2m}\big)\mathrm{d}s\right) \leqslant C_3 \frac{1}{\gamma_\varepsilon} n^{-1-\vartheta}.$$

To control the last two terms, recall that we have $\alpha_\varepsilon = \beta_\varepsilon = \gamma_\varepsilon$ in this case. Thus, by the definition of $\mathscr{L}_2$, we can write

$$\begin{aligned}
&\hat{\mathscr{Q}}_4(\varepsilon) - \mathbb{E}\left(\int_0^t \overline{H \cdot \nabla_y \Phi^f}(s, Y_s^\varepsilon)\mathrm{d}s\right)\\
&\leqslant \mathbb{E}\left(\int_0^t H(s, X_s^\varepsilon, Y_s^\varepsilon)\cdot \nabla_y \Phi_n^f(s, X_s^\varepsilon, Y_s^\varepsilon) - H(s, X_s^\varepsilon, Y_s^\varepsilon)\cdot \nabla_y \Phi^f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\right)\\
&\quad + \mathbb{E}\left(\int_0^t H(s, X_s^\varepsilon, Y_s^\varepsilon)\cdot \nabla_y \Phi^f(s, X_s^\varepsilon, Y_s^\varepsilon) - \overline{H \cdot \nabla_y \Phi^f}(s, Y_s^\varepsilon)\mathrm{d}s\right) =: \hat{\mathscr{Q}}_{41}(\varepsilon) + \hat{\mathscr{Q}}_{42}(\varepsilon).
\end{aligned}$$

Since $\nabla_y \Phi^f \in C_p^{(1+\vartheta)/2, 2+\delta, \vartheta}$ and by the fact

$$\nabla_y(\Phi_n^f) = (\nabla_y \Phi^f) * \rho_1^n * \rho_2^n,$$

we can deduce by using (4.2) again that

$$\begin{aligned}
\hat{\mathscr{Q}}_{41}(\varepsilon) &\leqslant C_4 \mathbb{E}\left(\int_0^t \big\|\nabla_y \Phi_n^f(\cdot, X_s^\varepsilon, \cdot) - \nabla_y \Phi^f(\cdot, X_s^\varepsilon, \cdot)\big\|_\infty \big(1+|X_s^\varepsilon|^m\big)\mathrm{d}s\right)\\
&\leqslant C_4 n^{-\vartheta} \mathbb{E}\left(\int_0^t \big(1+|X_s^\varepsilon|^{2m}\big)\mathrm{d}s\right) \leqslant C_4 n^{-\vartheta}.
\end{aligned}$$

On the other hand, by the assumption that $H \in C_p^{\vartheta/2, \delta, \vartheta}$, it is easy to check that $H \cdot \nabla_y \Phi^f \in C_p^{\vartheta/2, \delta, \vartheta}$. Thus $\overline{H \cdot \nabla_y \Phi^f} \in C_b^{\vartheta/2, \vartheta}$ by Lemma 3.2. Note that the function

$$H(t, x, y)\cdot \nabla_y \Phi^f(t, x, y) - \overline{H \cdot \nabla_y \Phi^f}(t, y)$$

satisfies (1.3). Hence by applying Lemma 4.2 with $\theta \in (0, 1]$ it follows that

$$\hat{\mathscr{Q}}_{42}(\varepsilon) \leqslant C_4 \Big(\alpha_\varepsilon^\vartheta + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

Finally, due to the assumption that $c \in C_p^{\delta, \vartheta}$ and by the same idea as above, one can check that

$$\hat{\mathscr{Q}}_5(\varepsilon) - \mathbb{E}\left(\int_0^t \overline{c \cdot \nabla_x \Phi^f}(s, Y_s^\varepsilon)\mathrm{d}s\right) \leqslant C_5\big(n^{-1-\vartheta} + \alpha_\varepsilon^\vartheta\big).$$

Combining the above computations, we arrive at

$$\hat{\mathscr{Q}}(\varepsilon) \leqslant C_6\Big(\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon} n^{1-\vartheta} + \frac{1}{\gamma_\varepsilon} n^{-1-\vartheta} + n^{-\vartheta} + \alpha_\varepsilon^\vartheta + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

Taking $n = \alpha_\varepsilon^{-1}$, we get the desired result.

(iii) (Regime 3) The only difference to *(iv)* is the last term. Recall that we have $\alpha_\varepsilon = \gamma_\varepsilon$ in this case. Thus, as above we have

$$\hat{\mathcal{Q}}_1(\varepsilon) + \hat{\mathcal{Q}}_2(\varepsilon) + \hat{\mathcal{Q}}_3(\varepsilon) + \hat{\mathcal{Q}}_4(\varepsilon) \leqslant C_1\Big(\alpha_\varepsilon n^{1-\vartheta} + \frac{1}{\alpha_\varepsilon} n^{-1-\vartheta} + n^{-\vartheta} + \alpha_\varepsilon^\vartheta + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

To control $\hat{\mathcal{Q}}_5(\varepsilon)$, we can simply use the growth condition on c to get

$$\hat{\mathcal{Q}}_5(\varepsilon) \leqslant C_2 \frac{\alpha_\varepsilon^2}{\beta_\varepsilon \gamma_\varepsilon} \mathbb{E}\left(\int_0^t \big(1 + |X_s^\varepsilon|^m\big)\mathrm{d}s\right) \leqslant C_2 \frac{\alpha_\varepsilon}{\beta_\varepsilon}.$$

Consequently, we arrive at

$$\hat{\mathcal{Q}}(\varepsilon) \leqslant C_3\Big(\frac{1}{\alpha_\varepsilon} n^{-1-\vartheta} + \alpha_\varepsilon n^{1-\vartheta} + n^{-\vartheta} + \alpha_\varepsilon^\vartheta + \frac{\alpha_\varepsilon}{\beta_\varepsilon}\Big)$$

Taking $n = \alpha_\varepsilon^{-1}$ again, we get the desired result.

(ii) (Regime 2) Now we need to pay attention to the assumption that $\vartheta \in (0, 2]$. By Theorem 2.1, (3.1) and the assumptions on b, σ and f, we have $\Phi^f \in C_p^{\vartheta/2, 2+\delta, \vartheta}$. Arguing as before, we have

$$\hat{\mathcal{Q}}_1(\varepsilon) + \hat{\mathcal{Q}}_2(\varepsilon) + \hat{\mathcal{Q}}_3(\varepsilon) \leqslant C_1\Big(\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon} n^{2-\vartheta} + \frac{1}{\gamma_\varepsilon} n^{-\vartheta}\Big),$$

and for $\hat{\mathcal{Q}}_4(\varepsilon)$, as in the proof of Lemma 4.2, we have

$$\hat{\mathcal{Q}}_4(\varepsilon) \leqslant C_2 \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2} n^{1-(\vartheta\wedge 1)} \mathbb{E}\left(\int_0^t \big(1 + |X_s^\varepsilon|^{2m}\big)\mathrm{d}s\right) \leqslant C_2 \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2} n^{1-(\vartheta\wedge 1)}.$$

For the last term, recall that $\alpha_\varepsilon^2 = \beta_\varepsilon\gamma_\varepsilon$ in this case. We use Lemma 4.2 with $\vartheta \in (0, 2]$ to deduce that

$$\begin{aligned}
&\hat{\mathcal{Q}}_5(\varepsilon) - \mathbb{E}\left(\int_0^t \overline{c \cdot \nabla_x \Phi^f}(s, Y_s^\varepsilon)\mathrm{d}s\right) \\
&\leqslant \mathbb{E}\left(\int_0^t c(X_s^\varepsilon, Y_s^\varepsilon) \cdot \big[\nabla_x \Phi_n^f(s, X_s^\varepsilon, Y_s^\varepsilon) - \nabla_x \Phi^f(s, X_s^\varepsilon, Y_s^\varepsilon)\big]\mathrm{d}s\right) \\
&\quad + \mathbb{E}\left(\int_0^t c(X_s^\varepsilon, Y_s^\varepsilon) \cdot \nabla_x \Phi^f(s, X_s^\varepsilon, Y_s^\varepsilon) - \overline{c \cdot \nabla_x \Phi^f}(s, Y_s^\varepsilon)\mathrm{d}s\right) \\
&\leqslant C_3\Big(n^{-\vartheta} + \alpha_\varepsilon^\vartheta + \alpha_\varepsilon^{\vartheta\wedge 1} \cdot \frac{\alpha_\varepsilon}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).
\end{aligned}$$

Consequently, we arrive at

$$\hat{\mathcal{Q}}(\varepsilon) \leqslant C_4\Big(\frac{\alpha_\varepsilon^2}{\gamma_\varepsilon} n^{2-\vartheta} + \frac{1}{\gamma_\varepsilon} n^{-\vartheta} + \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2} n^{1-(\vartheta\wedge 1)} + \alpha_\varepsilon^\vartheta + \alpha_\varepsilon^{\vartheta\wedge 1} \cdot \frac{\alpha_\varepsilon}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

Still, we can take $n = \alpha_\varepsilon^{-1}$ to get that for $\vartheta \in (0, 1]$,

$$\hat{\mathcal{Q}}(\varepsilon) \leqslant C_4\Big(\frac{\alpha_\varepsilon^\vartheta}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^{1+\vartheta}}{\gamma_\varepsilon^2} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big) \leqslant C_4\Big(\frac{\alpha_\varepsilon^\vartheta}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

and for $\vartheta \in (1, 2]$, we have

$$\hat{\mathcal{Q}}(\varepsilon) \leqslant C_4\Big(\frac{\alpha_\varepsilon^\vartheta}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big),$$

which in turn yields the desired result.

(i) (Regime 1) Finally, we can use Lemma 4.2 directly to get that

$$\mathbb{E}\left(\frac{1}{\gamma_\varepsilon}\int_0^t f(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\right) \leqslant C_1\Big(\frac{\alpha_\varepsilon^\vartheta}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^{1+(\vartheta\wedge 1)}}{\gamma_\varepsilon^2} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon\gamma_\varepsilon}\Big) \leqslant C_1\Big(\frac{\alpha_\varepsilon^\vartheta}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon\gamma_\varepsilon}\Big).$$

The whole proof is finished. □

5. Diffusion approximations

Recall that $\Phi(t, x, y)$ is the solution to Poisson equation (2.8), and the limit effective system is given by (2.9). For $k = 1, \cdots, 4$, denote by $\hat{\mathcal{L}}_k$ the infinitesimal operator of $\hat{Y}_t^k$, i.e.,

$$\hat{\mathcal{L}}_k := \sum_{i,j=1}^{d_2} \hat{\mathcal{G}}_k^{ij}(t, y)\frac{\partial^2}{\partial y_i \partial y_j} + \sum_{i=1}^{d_2} \hat{F}_k^i(t, y)\frac{\partial}{\partial y_i},$$

where $\hat{\mathcal{G}}_k(t, y) := \hat{G}_k\hat{G}_k^*(t, y)/2$. The following properties for the averaged coefficients follow by Lemma 3.2.

Lemma 5.1. *Under the assumptions in Theorem 2.3, for every $k = 1, \cdots, 4$, $\hat{\mathcal{G}}_k$ is non-degenerate in y uniformly with respect to t. Moreover, we have $\hat{F}_k, \hat{\mathcal{G}}_k \in C_b^{\vartheta/2,\vartheta}$.*

Proof. The non-degeneracy of $\hat{\mathcal{G}}_1 = \hat{\mathcal{G}}_2$ follows directly by $(\mathbf{A}_G)$ and the definition. Furthermore, by (3.1) we have

$$\int_{\mathbb{R}^{d_1}} H(t, x, y)\Phi^*(t, x, y)\mu^y(\mathrm{d}x) = \mathbb{E}\left(\int_0^\infty \int_{\mathbb{R}^{d_1}} H(t, x, y)H^*(t, X_t^y, y)\mu^y(\mathrm{d}x)\mathrm{d}t\right),$$

which is non-negative by the homogeneity of X_t^y. Thus $\hat{\mathcal{G}}_3 = \hat{\mathcal{G}}_4$ are also non-degenerate. Let us prove the regularity for the most general case $\hat{F}_4$. The conclusions for $\hat{F}_2, \hat{F}_3, \hat{F}_4$ and $\hat{G}_k$ $(k = 1, \cdots, 4)$ can be proved by the same argument. In this case, we have by the assumption that $b, \sigma \in C_b^{\delta,1+\vartheta}$, $H \in C_p^{(1+\vartheta)/2,\delta,1+\vartheta}$ and Theorem 2.1 that $\Phi(t, x, y) \in C_p^{(1+\vartheta)/2,2+\delta,1+\vartheta}$, which together with the assumptions on F and c implies that

$$F(t, x, y) + c(x, y)\cdot\nabla_x\Phi(t, x, y) + H(t, x, y)\cdot\nabla_y\Phi(t, x, y) \in C_p^{\vartheta/2,\delta,\vartheta}.$$

Thus the conclusion follows by Lemma 3.2 directly. □

According to the above result, there exists a unique weak solution $\hat{Y}_t^k$ for SDE (2.9) for every $k = 1, \cdots, 4$. To prove the weak convergence of Y_t^ε to $\hat{Y}_t^k$, we need to consider the following Cauchy problem on $[0,T] \times \mathbb{R}^{d_2}$:

$$\begin{cases} \partial_t u_k(t,y) - \hat{\mathcal{L}}_k u_k(t,y) = 0, & t \in [0,T), \\ u_k(0,y) = \varphi(y), \end{cases} \tag{5.1}$$

where $T > 0$ and φ is a function on $\mathbb{R}^{d_2}$. As a direct consequence of Lemma 5.1, we have the following result for the Cauchy problem (5.1), which is well-known in the theory of PDEs, see e.g. [33, Chapter IV, Section 5].

Lemma 5.2. *Assume that $\varphi \in C_b^{2+\vartheta}$. Then for every $k = 1, \cdots, 4$, there exists a unique solution $u_k \in C_b^{(2+\vartheta)/2, 2+\vartheta}$ to equation (5.1) which is given by*

$$u_k(t,y) = \mathbb{E}\varphi(\hat{Y}_t^k(y)). \tag{5.2}$$

Moreover, we also have $\nabla_y u_k \in C_b^{(1+\vartheta)/2, 1+\vartheta}$ and $\nabla_y^2 u \in C_b^{\vartheta/2, \vartheta}$.

Now, we are in the position to give:

Proof of Theorem 2.3. Given $T > 0$ and $\varphi \in C_b^{2+\vartheta}$, let u_k be the unique solution to equation (5.1). Define

$$\tilde{u}_k(t,y) := u_k(T-t, y).$$

Then it is obvious that $\tilde{u}_k(T,y) = u_k(0,y) = \varphi(y)$. As a result, we can deduce by (5.2) and the Itô's formula that

$$\begin{aligned} \mathcal{R}_k(\varepsilon) &:= \mathbb{E}[\varphi(Y_T^\varepsilon)] - \mathbb{E}[\varphi(\hat{Y}_T^k)] = \mathbb{E}[\tilde{u}_k(T, Y_T^\varepsilon) - \tilde{u}_k(0,y)] \\ &= \mathbb{E}\bigg(\int_0^T \big(\partial_s + \mathscr{L}_1\big) \tilde{u}_k(s, Y_s^\varepsilon) \mathrm{d}s \bigg) + \mathbb{E}\left(\frac{1}{\gamma_\varepsilon} \int_0^T \mathscr{L}_2 \tilde{u}_k(s, Y_s^\varepsilon) \mathrm{d}s \right) \\ &= \mathbb{E}\bigg(\int_0^T \big(\mathscr{L}_1 - \hat{\mathcal{L}}_k\big) \tilde{u}_k(s, Y_s^\varepsilon) \mathrm{d}s \bigg) + \frac{1}{\gamma_\varepsilon} \mathbb{E}\left(\int_0^T H(s, X_s^\varepsilon, Y_s^\varepsilon) \cdot \nabla_y \tilde{u}_k(s, Y_s^\varepsilon) \mathrm{d}s \right), \end{aligned}$$

where $\mathscr{L}_2$ and $\mathscr{L}_1$ are defined by (1.5). Below, we divide the proof into four parts, which correspond to the four regimes in (1.8).

(i) (Regime 1) In this case, we have

$$\begin{aligned} \big(\mathscr{L}_1 - \hat{\mathcal{L}}_1\big) \tilde{u}_1(t,y) &= \big[\mathcal{G}(t,x,y) - \hat{\mathcal{G}}_1(t,y)\big] \cdot \nabla_y^2 \tilde{u}_1(t,y) \\ &\quad + \big[F(t,x,y) - \hat{F}_1(t,y)\big] \cdot \nabla_y \tilde{u}_1(t,y). \end{aligned}$$

Note that the function $\big(\mathscr{L}_1 - \hat{\mathcal{L}}_1\big)\tilde{u}_1(t,y)$ satisfies the centering condition (1.3). Moreover, by the assumption that $F, \mathcal{G} \in C_p^{\vartheta/2, \delta, \vartheta}$, Lemma 5.1 and Lemma 5.2, we have $\big(\mathscr{L}_1 - \hat{\mathcal{L}}_1\big)\tilde{u}_1(t,y) \in C_p^{\vartheta/2,\delta,\vartheta}$. Thus, applying Lemma 4.2 with $\vartheta \in (0,2]$ we can get

$$\mathcal{R}_{11}(\varepsilon) := \mathbb{E}\left(\int_0^T \big(\mathscr{L}_1 - \hat{\mathcal{L}}_1\big)\tilde{u}_1(s,Y_s^\varepsilon)\mathrm{d}s\right) \leqslant C_1\Big(\alpha_\varepsilon^\vartheta + \alpha_\varepsilon^{\vartheta\wedge 1}\cdot\frac{\alpha_\varepsilon}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

On the other hand, note that $H(t,x,y)\cdot\nabla_y\tilde{u}_1(t,y) \in C_p^{\vartheta/2,\delta,\vartheta}$ also satisfies (1.3). Hence we have by Lemma 4.4 *(i)* that

$$\mathcal{R}_{12}(\varepsilon) := \frac{1}{\gamma_\varepsilon}\mathbb{E}\left(\int_0^T H(s,X_s^\varepsilon,Y_s^\varepsilon)\cdot\nabla_y\tilde{u}_1(s,Y_s^\varepsilon)\mathrm{d}s\right) \leqslant C_1\Big(\frac{\alpha_\varepsilon^\vartheta}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon\gamma_\varepsilon}\Big).$$

Consequently,

$$\mathcal{R}_1(\varepsilon) = \mathcal{R}_{11}(\varepsilon) + \mathcal{R}_{12}(\varepsilon) \leqslant C_1\Big(\frac{\alpha_\varepsilon^\vartheta}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon\gamma_\varepsilon}\Big).$$

(ii) (Regime 2) In this case, note that

$$\hat{\mathcal{L}}_2 = \hat{\mathcal{L}}_1 + \overline{c\cdot\nabla_x\Phi}(t,y)\cdot\nabla_y.$$

Thus, we can write

$$\begin{aligned}\mathcal{R}_2(\varepsilon) = \mathbb{E}&\left(\int_0^T \big(\mathscr{L}_1 - \hat{\mathcal{L}}_1\big)\tilde{u}_2(s,Y_s^\varepsilon)\mathrm{d}s\right)\\ &+ \left[\frac{1}{\gamma_\varepsilon}\mathbb{E}\left(\int_0^T H(s,X_s^\varepsilon,Y_s^\varepsilon)\cdot\nabla_y\tilde{u}_2(s,Y_s^\varepsilon)\mathrm{d}s\right)\right.\\ &\qquad \left. - \mathbb{E}\left(\int_0^T \overline{c\cdot\nabla_x\Phi}(s,Y_s^\varepsilon)\cdot\nabla_y\tilde{u}_2(s,Y_s^\varepsilon)\mathrm{d}s\right)\right] =: \mathcal{R}_{21}(\varepsilon) + \mathcal{R}_{22}(\varepsilon).\end{aligned}$$

Entirely similar as above, we can control the first term by

$$\mathcal{R}_{21}(\varepsilon) \leqslant C_2\Big(\alpha_\varepsilon^\vartheta + \alpha_\varepsilon^{\vartheta\wedge 1}\cdot\frac{\alpha_\varepsilon}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

On the other hand, let Ψ_2 be the solution to

$$\mathscr{L}_0(x,y)\Psi_2(t,x,y) = -H(t,x,y)\cdot\nabla_y\tilde{u}_2(t,y) \in C_p^{\vartheta/2,\delta,\vartheta}.$$

Then, it is easy to see that

$$\Psi_2(t,x,y) = \Phi(t,x,y)\cdot\nabla_y\tilde{u}_2(t,y),$$

where Φ satisfies (2.8). Furthermore,

$$\int_{\mathbb{R}^{d_1}} c(x,y)\cdot\nabla_x\Psi_2(t,x,y)\mu^y(\mathrm{d}x) = \overline{c\cdot\nabla_x\Phi}(t,y)\cdot\nabla_y\tilde{u}_2(t,y).$$

Consequently, we can apply Lemma 4.4 *(ii)* with $f = H \cdot \nabla_y \tilde{u}_2$ to get that

$$\mathcal{R}_{22}(\varepsilon) \leqslant C_2\Big(\frac{\alpha_\varepsilon^\vartheta}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

Thus we arrive at

$$\mathcal{R}_2(\varepsilon) = \mathcal{R}_{21}(\varepsilon) + \mathcal{R}_{22}(\varepsilon) \leqslant C_2\Big(\frac{\alpha_\varepsilon^\vartheta}{\gamma_\varepsilon} + \frac{\alpha_\varepsilon^2}{\gamma_\varepsilon^2} + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

(iii) (Regime 3) In this case, we have

$$\hat{\mathcal{L}}_3 = \hat{\mathcal{L}}_1 + \overline{H \cdot \nabla_y \Phi}(t, y) \cdot \nabla_y + \overline{H \cdot \Phi^*}(t, y) \cdot \nabla_y^2,$$

where the extra diffusion coefficient is given by

$$\overline{H \cdot \Phi^*}(t, y) := \int_{\mathbb{R}^{d_1}} H(t, x, y)\Phi^*(t, x, y)\mu^y(\mathrm{d}x).$$

Thus we can write

$$\begin{aligned}
\mathcal{R}_3(\varepsilon) = \mathbb{E}\left(\int_0^T \big(\mathscr{L}_1 - \hat{\mathcal{L}}_1\big)\tilde{u}_3(s, Y_s^\varepsilon)\mathrm{d}s\right)& \\
+ \Bigg[\frac{1}{\gamma_\varepsilon}\mathbb{E}\left(\int_0^T H(s, X_s^\varepsilon, Y_s^\varepsilon) \cdot \nabla_y \tilde{u}_3(s, Y_s^\varepsilon)\mathrm{d}s\right)& \\
- \mathbb{E}\left(\int_0^T \overline{H \cdot \nabla_y \Phi}(s, Y_s^\varepsilon) \cdot \nabla_y \tilde{u}_3(s, Y_s^\varepsilon) + \overline{H \cdot \Phi^*}(s, Y_s^\varepsilon) \cdot \nabla_y^2 \tilde{u}_3(s, Y_s^\varepsilon)\mathrm{d}s\right)\Bigg]& \\
=: \mathcal{R}_{31}(\varepsilon) + \mathcal{R}_{32}(\varepsilon).&
\end{aligned}$$

Following the same idea as in (i) and using Lemma 4.2 with $\vartheta \in (0, 1]$, we can control the first term by

$$\mathcal{R}_{31}(\varepsilon) \leqslant C_3\Big(\alpha_\varepsilon^\vartheta + \frac{\alpha_\varepsilon^2}{\beta_\varepsilon}\Big).$$

On the other hand, let Ψ_3 be the solution to

$$\mathscr{L}_0(x, y)\Psi_3(t, x, y) = -H(t, x, y) \cdot \nabla_y \tilde{u}_3(t, y) \in C_p^{(1+\vartheta)/2, \delta, 1+\vartheta},$$

which is given by

$$\Psi_3(t, x, y) = \Phi(t, x, y) \cdot \nabla_y \tilde{u}_3(t, y).$$

One can check that

$$\begin{aligned}
&\int_{\mathbb{R}^{d_1}} H(t, x, y) \cdot \nabla_y \Psi_3(t, x, y)\mu^y(\mathrm{d}x) \\
&= \int_{\mathbb{R}^{d_1}} H(t, x, y) \cdot \nabla_y \Phi(t, x, y)\mu^y(\mathrm{d}x) \cdot \nabla_y \tilde{u}_3(t, y) \\
&\quad + \int_{\mathbb{R}^{d_1}} H(t, x, y) \cdot \Phi^*(t, x, y)\mu^y(\mathrm{d}x) \cdot \nabla_y^2 \tilde{u}_3(t, y)
\end{aligned}$$

$$= \overline{H \cdot \nabla_y \Phi}(t,y) \cdot \nabla_y \tilde{u}_3(t,y) + \overline{H \cdot \Phi^*}(t,y) \cdot \nabla_y^2 \tilde{u}_3(t,y).$$

Consequently, by applying Lemma 4.4 *(iii)* with $f = H \cdot \nabla_y \tilde{u}_3$ we have that

$$\mathcal{R}_{32}(\varepsilon) \leqslant C_3 \Big(\alpha_\varepsilon^\vartheta + \frac{\alpha_\varepsilon}{\beta_\varepsilon} \Big),$$

which in turn yields the desired result.

(iv) (Regime 4) Finally, we have

$$\hat{\mathcal{L}}_4 = \hat{\mathcal{L}}_1 + \overline{c \cdot \nabla_x \Phi}(t,y) \cdot \nabla_y + \overline{H \cdot \nabla_y \Phi}(t,y) \cdot \nabla_y + \overline{H \cdot \Phi^*}(t,y) \cdot \nabla_y^2.$$

We thus have

$$\begin{aligned}
\mathcal{R}_4(\varepsilon) = \mathbb{E} &\left(\int_0^T \big(\mathscr{L}_1 - \hat{\mathcal{L}}_1 \big) \tilde{u}_4(s, Y_s^\varepsilon) \mathrm{d}s \right) \\
&+ \left[\frac{1}{\gamma_\varepsilon} \mathbb{E} \left(\int_0^T H(s, X_s^\varepsilon, Y_s^\varepsilon) \cdot \nabla_y \tilde{u}_4(s, Y_s^\varepsilon) \mathrm{d}s \right) \right. \\
&- \mathbb{E} \Big(\int_0^T \Big[\overline{c \cdot \nabla_x \Phi} + \overline{H \cdot \nabla_y \Phi} \Big] (s, Y_s^\varepsilon) \cdot \nabla_y \tilde{u}_4(s, Y_s^\varepsilon) \\
&\qquad \left. + \overline{H \cdot \Phi^*}(s, Y_s^\varepsilon) \cdot \nabla_y^2 \tilde{u}_4(s, Y_s^\varepsilon) \mathrm{d}s \Big) \right] =: \mathcal{R}_{41}(\varepsilon) + \mathcal{R}_{42}(\varepsilon).
\end{aligned}$$

As above, we can control the first term by

$$\mathcal{R}_{41}(\varepsilon) \leqslant C_4 \alpha_\varepsilon^\vartheta.$$

Let Ψ_4 be the solution to

$$\mathscr{L}_0(x,y) \Psi_4(t,x,y) = -H(t,x,y) \cdot \nabla_y \tilde{u}_4(t,y) \in C_p^{(1+\vartheta)/2, \delta, 1+\vartheta},$$

which is given by

$$\Psi_4(t,x,y) = \Phi(t,x,y) \cdot \nabla_y \tilde{u}_4(t,y).$$

One can check that

$$\begin{aligned}
&\int_{\mathbb{R}^{d_1}} c(x,y) \cdot \nabla_x \Psi_4(t,x,y) \mu^y(\mathrm{d}x) + \int_{\mathbb{R}^{d_1}} H(t,x,y) \cdot \nabla_y \Psi_4(t,x,y) \mu^y(\mathrm{d}x) \\
&= \Big[\overline{c \cdot \nabla_x \Phi} + \overline{H \cdot \nabla_y \Phi} \Big] (t,y) \cdot \nabla_y \tilde{u}_4(t,y) + \overline{H \cdot \Phi^*}(t,y) \cdot \nabla_y^2 \tilde{u}_4(t,y).
\end{aligned}$$

Consequently, by applying Lemma 4.4 *(iv)* with $f = H \cdot \nabla_y \tilde{u}_4$ we have that

$$\mathcal{R}_{42}(\varepsilon) \leqslant C_4 \alpha_\varepsilon^\vartheta.$$

The whole proof is finished. $\square$

References

[1] Abdulle A., Pavliotis G. A. and Vaes U.: Spectral methods for multiscale stochastic differential equations. *SIAM/ASA J. Uncertain. Quantif.*, **5** (2017), 720–761.

[2] Abourashchia N. and Veretennikov A. Yu.: On stochastic averaging and mixing. *Theory Stoch. Process.*, **16** (2010), 111–129.

[3] Bakhtin V. and Kifer Y.: Diffusion approximation for slow motion in fully coupled averaging. *Probab. Theory Rel. Fields*, **129** (2004), 157–181.

[4] Bakhvalov N. S.: Averaging of partial differential equations with rapidly oscillating coefficients. *Sov. Math., Dokl.*, **16** (1975), 351–355.

[5] Ball K., Kurtz T. G., Rempala G. and Popovic L.: Asymptotic analysis of multiscale approximations to reaction networks. *Ann. Appl. Prob.*, **16** (2005), 1925–1961.

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

[7] 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.

[8] Bréhier C. E. and Kopec M.: Approximation of the invariant law of SPDEs: error analysis using a Poisson equation for a full-discretization scheme. *SIMA J. Numer. Anal.*, **37** (2017), 1375–1410.

[9] Cattiaux P., Chafaï D. and Guillin A.: Central limit theorems for additive functionals of ergodic Markov diffusions processes. *ALEA, Lat. Am. J. Probab. Math. Stat.*, **9** (2012), 337–382.

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

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

[12] Cerrai S. and Lunardi A.: Averaging principle for nonautonomous slow-fast systems of stochastic reaction-diffusion equations: the almost periodic case. *SIAM J. Math. Anal.*, **49** (2017), 2843–2884.

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

[14] Dupuis P. and Spiliopoulos K.: Large deviations for multiscale problems via weak convergence methods. *Stoch. Proc. Appl.*, **122** (2012), 1947–1987.

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

[16] Ethier S. N. and Kurtz T. G.: Markov Processes. Characterization and Convergence. Wiley, New York, 1986.

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

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

[19] Gao H. and Duan J.: Dynamics of quasi-geostrophic fluid motion with rapidly oscillating Coriolis force. *Nonl. Anal.: Real World Appl.*, **4** (2003), 127–138.

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

[21] Glynn P. W. and Meyn S. P.: A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.*, **24** (1996), 916–931.

[22] Gonzales-Gargate I. I. and Ruffino P. R.: An averaging principle for diffusions in foliated spaces. *Ann. Probab.*, **44** (2016) 567–588.

[23] Hairer M. and Pardoux E.: Homogenization of periodic linear degenerate PDEs. *J. Func. Anal.*, **255** (2008), 2462–2487.

[24] Hashemi S. N. and Heunis A. J.: On the Poisson equation for singular diffusions. *Stochastics: International J. Probab. Stoch. Proc.*, **77** (2005), 155–189.

[25] Honoré I., Menozzi S. and Pagès G.: Non-asymptotic Gaussian estimates for the recursive approximation of the invariant measure of a diffusion. https://arxiv.org/pdf/1605.08525.pdf.

[26] Kelly D. and Vanden-Eijnden E.: Fluctuations in the heterogeneous multiscale methods for fastslow systems. *Res. Math. Sci.*, **4** (2017), DOI: 10.1186/s40687-017-0112-2.

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

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

[29] Kifer Y.: Averaging in dynamical systems and large deviations. *Invent. Math.*, **110** (1992), 337–370.

[30] Kifer Y.: Averaging and climate models. Stochastic Climate Models. (Progress in Probability, 49). Birkhäuser, Basel, 2001.

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

[32] Kulik A. M. and Veretennikov A. Yu.: Extended Poisson equation for weakly ergodic Markov processes. *Theor. Probability and Math. Statist.*, **85** (2012), 23–39.

[33] 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.

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

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

[36] Majda A., Timofeyev I. and Eijnden E. V.: A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.*, **54** (2001), 891–974.

[37] Mattingly J. C., Stuart A. M. and Tretyakov M. V.: Convergence of numerical time-averaging and stationary measures via Poisson equations. *SIAM J. Numer. Anal.*, **48** (2010), 552–577.

[38] Morse M. R. and Spiliopoulos K.: Moderate deviations principle for systems of slow-fast diffusions. https://arxiv.org/abs/1611.05903.

[39] Pagès G. and Panloup F.: Ergodic approximation of the distribution of a stationary diffusion: rate of convergence. *Ann. Appl. Probab.*, **22** (2012), 1059–1100.

[40] Papanicolaou G. C., Stroock D. and Varadhan S. R. S.: Martingale approach to some limit theorems. Proceedings of the 1976 Duke University Conference on Turbulence, Durham, NC, 1976.

[41] Pardoux E.: Homogenization of linear and semilinear second order parabolic PDEs with periodic coefficients: a probabilistic approach. *J. Func. Anal.*, **167** (1999), 498–520.

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

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

[44] Pardoux E. and Veretennikov A. Yu.: On the Poisson equation and diffusion approximation 3. *Ann. Prob.*, **33** (2005), 1111–1133.

[45] Pavliotis G. A. and Stuart A. M.: Multiscale methods: averaging and homogenization. Texts Appl. Math., 53, Springer, New York, 2008.
[46] Puhalskii A. A.: On large deviations of coupled diffusions with time scale separation. *Ann. Prob.*, **44** (2016), 3111–3186.
[47] Röckner M., Sun X. and Xie L.: Strong and weak convergence in averaging principle for SDEs with Hölder coefficients. https://arxiv.org/pdf/1907.09256.pdf.
[48] Spiliopoulos K.: Fluctuation analysis and short time asymptotics for multiple scales diffusion processes. *Stochastics and Dynamics*, **14** (2014), pp. 1350026.
[49] Wang W. and Roberts A. J.: Average and deviation for slow-fast stochastic partial differential equations. *J. Diff. Equ.*, **253** (2012), 1265–1286.
[50] Vanden-Eijnden E: Numerical techniques for multi-scale dynamical systems with stochastic effects. *Commun Math Sci.*, **1** (2003), 377–384.
[51] Veretennikov A. Yu.: On the averaging principle for systems of stochastic differential equations. *Math. USSR Sborn.*, **69** (1991), 271–284.
[52] Veretennikov A. Yu.: On polynomial mixing bounds for stochastic differential equations. *Stoch. Processes Appl.*, **70** (1997), 115–127.
[53] Veretennikov A. Yu.: On Sobolev solutions of poisson equations in $\mathbb{R}^d$ with a parameter. *J. Math. Sci.*, **179** (2011), 1–32.
[54] Zhang B., Fu H., Wan L., Liu J.: Weak order in averaging principle for stochastic differential equations with jumps. *Adv. Difference Equ.*, 2018, Paper No. 197, 20 pp.

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

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