Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2021, Vol. 57, No. 1, 547–576
https://doi.org/10.1214/20-AIHP1087

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES

Strong convergence order for slow–fast McKean–Vlasov stochastic differential equations

Michael Röckner[a], Xiaobin Sun[b,*] and Yingchao Xie[b,†]

[a]*Fakultät für Mathematik, Universität Bielefeld, D-33501 Bielefeld, Germany, and Academy of Mathematics and Systems Science, Chinese Academy of Sciences (CAS), Beijing, 100190, China. E-mail: roeckner@math.uni-bielefeld.de*
[b]*School of Mathematics and Statistics and Research Institute of Mathematical Science, Jiangsu Normal University, Xuzhou, 221116, China. E-mail: *xbsun@jsnu.edu.cn; †ycxie@jsnu.edu.cn*

Received 28 September 2019; revised 31 May 2020; accepted 16 July 2020

Abstract. In this paper, we consider the averaging principle for a class of McKean–Vlasov stochastic differential equations with slow and fast time-scales. Under some proper assumptions on the coefficients, we first prove that the slow component strongly converges to the solution of the corresponding averaged equation with convergence order $1/3$ using the approach of time discretization. Furthermore, under stronger regularity conditions on the coefficients, we use the technique of Poisson equation to improve the order to $1/2$, which is the optimal order of strong convergence in general.

Résumé. Dans cet article, nous considérons le principe de moyennisation pour une classe d'équations différentielles stochastiques de type McKean–Vlasov avec une échelle de temps lente et une échelle de temps rapide. Sous des hypothèses adéquates sur les coefficients, nous montrons d'abord que la composante lente converge vers la solution de l'équation moyennée correspondante avec un ordre de convergence $1/3$ en utilisant une approche par discrétisation en temps. D'autre part, sous des hypothèses de régularité plus fortes sur les coefficients, nous utilisons la technique de l'équation de Poisson pour améliorer l'ordre à $1/2$, ce qui est l'ordre optimal de la convergence forte en général.

MSC2020 subject classifications: Primary 60H10; secondary 34F05

Keywords: Averaging principle; McKean–Vlasov stochastic differential equations; Slow–fast; Poisson equation; Strong convergence order

1. Introduction

Let $\{W_t^1\}_{t\geq 0}$ and $\{W_t^2\}_{t\geq 0}$ be mutually independent d_1 and d_2 dimensional standard Brownian motions on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and $\{\mathscr{F}_t, t \geq 0\}$ be the natural filtration generated by W_t^1 and W_t^2. Let the following maps $b = b(t, x, \mu, y)$, $\sigma = \sigma(t, x, \mu)$, $f = f(t, x, \mu, y)$ and $g = g(t, x, \mu, y)$ be given:

$$b : [0, \infty) \times \mathbb{R}^n \times \mathscr{P}_2 \times \mathbb{R}^m \to \mathbb{R}^n;$$

$$\sigma : [0, \infty) \times \mathbb{R}^n \times \mathscr{P}_2 \to \mathbb{R}^{n\times d_1};$$

$$f : [0, \infty) \times \mathbb{R}^n \times \mathscr{P}_2 \times \mathbb{R}^m \to \mathbb{R}^m;$$

$$g : [0, \infty) \times \mathbb{R}^n \times \mathscr{P}_2 \times \mathbb{R}^m \to \mathbb{R}^{m\times d_2}$$

such that b, σ, f and g are continuous in $(t, x, \mu, y) \in [0, \infty) \times \mathbb{R}^n \times \mathscr{P}_2 \times \mathbb{R}^m$, where $\mathscr{P}_2$ is defined by

$$\mathscr{P}_2 := \left\{\mu \in \mathscr{P} : \mu\big(|\cdot|^2\big) := \int_{\mathbb{R}^n} |x|^2 \mu(dx) < \infty\right\},$$

where $\mathscr{P}$ is the set of all probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. Then $\mathscr{P}_2$ is a polish space under the L^2-Wasserstein distance, i.e.,

$$\mathbb{W}_2(\mu_1,\mu_2) := \inf_{\pi\in\mathcal{C}_{\mu_1,\mu_2}}\left[\int_{\mathbb{R}^n\times\mathbb{R}^n}|x-y|^2\pi(dx,dy)\right]^{1/2}, \quad \mu_1,\mu_2\in\mathscr{P}_2,$$

where $\mathcal{C}_{\mu_1,\mu_2}$ is the set of all couplings for μ_1 and μ_2.

In this paper, we consider the following slow–fast McKean–Vlasov stochastic differential equations (SDEs):

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

where $\mathscr{L}_{X_t^\epsilon}$ is the law of X_t^ϵ, ϵ is a small and positive parameter describing the ratio of the time scale between the slow component $X_t^\epsilon\in\mathbb{R}^n$ and fast component $Y_t^\epsilon\in\mathbb{R}^m$.

The averaging principle has a long and rich history in multiscale models, which have wide applications in material sciences, chemistry, fluid dynamics, biology, ecology, climate dynamics etc., see e.g., [1,13,14,23,29,41] and references therein. The averaging principle is essential to describe the asymptotic behavior of the slow component as $\epsilon\to 0$, i.e., the slow component will convergence to the so-called averaged equation. Bogoliubov and Mitropolsky [2] first studied the averaging principle for deterministic systems. The averaging principle for SDEs was first studied by Khasminskii in [24], see e.g., [20–22,25,28,37] for further developments. The averaging principle for slow–fast stochastic partial differential equations (SPDEs) was first investigated by Cerrai and Freidlin in [9], see e.g., [3,7,8,10–12,15–19,27,36,39,40] for further developments.

The McKean–Vlasov SDEs (also called distribution dependent SDEs) describe stochastic systems whose evolution is determined by both the microcosmic location and the macrocosmic distribution of the particle. The time marginal laws of the solution of such SDEs satisfies a nonlinear Fokker–Planck–Kolmogorov equation. The existence and uniqueness of weak and strong solutions have been studied intensively (see e.g., [30,38]), and see [35] for the case of singular drifts. Further properties, such as the Harnack inequality or the Bismut formula for the Lions Derivative have been investigated in [38] and [33] respectively. However, to the authors' knowledge, this paper is the first in which the averaging principle for two-time scale distribution dependent SDEs is considered.

As is well-known, the exponential ergodicity of the transition semigroup of the corresponding frozen equation plays an important role in the proof of strong averaging convergence. However it does not hold if the coefficients of equation (1.1) depend on the law of the fast component. So, we focus on the coefficients depending only on the law of the slow component here. In addition, it is worth pointing out that the assumption, that the diffusion coefficient σ of equation (1.1) does not depend on the fast component Y_t^ε, is necessary. Otherwise, the strong convergence will be incorrect (please see the counter-example in [25, Section 4.1]).

For numerical purposes, however, only studying the strong convergence of the slow component to the corresponding averaged equation is not enough, since in addition one needs to know the rate of convergence. Hence, the main purpose of our paper is to study the strong convergence rate for two-time scale distribution dependent SDEs. More precisely, one tries to find the largest possible $\alpha>0$ such that

$$\sup_{t\in[0,T]}\mathbb{E}\left|X_t^\epsilon-\bar{X}_t\right|^2\le C\epsilon^\alpha, \tag{1.2}$$

where C is a constant depending on T, $|x|$, $|y|$, and $\bar{X}$ is the solution of the corresponding averaged equation (see Eq. (2.18) below).

In the distribution-independent case, the strong convergence rate for two-time scale stochastic system has been studied in a number of papers (see e.g., [20,21,25,34] for the finite dimensional case, and [3,4] for the infinite dimensional case). The approach based on Khasminskii's technique of time discretization is often used to study the strong convergence rate (see [3,20,21,25]). Recently, the technique of Poisson equation has been used to study the strong convergence rate in [4,34], and the optimal convergence order was obtained in general. Motivated by this, in this paper we will use the techniques of time discretization and Poisson equation to study the strong convergence rate for two-time scale distribution dependent SDEs separately. More precisely, under some proper assumptions on the coefficients, we use the technique of time discretization to obtain the convergence order 1/3, which is however usually not the optimal order. It turns out that under some stronger assumptions on the coefficients, the optimal convergence order 1/2 can indeed be obtained by the method of Poisson equation.

If the technique of Poisson equation (see [31,32,34]) is applied to prove our main result, the main difficulty is to analyse the regularity of the solution $\Phi(t,x,\mu,y)$ of the corresponding Poisson equation with respect to (*w.r.t.*) the parameter μ. Indeed, this method highly depends on the regularity of Φ *w.r.t.* parameters. However, due to the coefficients dependence on the distribution, Φ will also depend on the distribution μ. Unlike as for classical SDEs, we have to apply Itô's formula to Φ composed with the process $(t, X_t^\epsilon, \mathscr{L}_{X_t^\epsilon}, Y_t^\epsilon)$, which in particular, means that we have to differentiate in the measure μ. As a consequence, some additional terms involving the Lions derivative of Φ appear, so we have to estimate the regularity of Φ *w.r.t.* the parameter μ carefully.

The paper is organized as follows. In the next section, we introduce some notation and assumptions that we use throughout the paper, and present out the main results. Sections 3 and 4 are devoted to proving the strong convergence rate by using the techniques of time discretization and Poisson equation respectively. We give an example in Section 5. In the Appendix, we give the detailed proof of the existence and uniqueness of solutions for our system and prove some important estimates.

We note that throughout this paper C and C_T denote positive constants which may change from line to line, where the subscript T is used to emphasize that the constant depends on T.

2. Notations and main results

Now, we first remind the reader of the definition of differentiability on the Wasserstein space. Following the idea in [6, Section 6], for $u : \mathscr{P}_2 \to \mathbb{R}$ we denote by U its "extension" to $L^2(\Omega, \mathbb{P}; \mathbb{R}^n)$ defined by

$$U(X) := u(\mathscr{L}_X), \quad X \in L^2\big(\Omega, \mathbb{P}; \mathbb{R}^n\big).$$

Then we say that u is differentiable at $\mu \in \mathscr{P}_2$ if there exists $X \in L^2(\Omega, \mathbb{P}; \mathbb{R}^n)$ such that $\mathscr{L}_X = \mu$ and U is Fréchet differentiable at X. By Riesz' theorem, the Fréchet derivative $DU(X)$, viewed as an element of $L^2(\Omega, \mathbb{P}; \mathbb{R}^n)$, can be represented as

$$DU(X) = \partial_\mu u(\mathscr{L}_X)(X),$$

where $\partial_\mu u(\mathscr{L}_X) : \mathbb{R}^n \to \mathbb{R}^n$, which is called Lions derivative of u at $\mu = \mathscr{L}_X$. Moreover, $\partial_\mu u(\mu) \in L^2(\mu; \mathbb{R}^n)$, for $\mu \in \mathscr{P}_2$. Furthermore, if $\partial_\mu u(\mu)(z) : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable at $z \in \mathbb{R}^n$, we denote its derivative by $\partial_z \partial_\mu u(\mu)(z) : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$.

Let $|\cdot|$ be the Euclidean vector norm, $\langle \cdot, \cdot \rangle$ be the Euclidean inner product and $\|\cdot\|$ be the matrix norm or the operator norm if there is no confusion possible. We call a vector-valued, or matrix-valued function $u(\mu) = (u_{ij}(\mu))$ differentiable at $\mu \in \mathscr{P}_2$, if all its components are differentiable at μ, and set $\partial_\mu u(\mu) := (\partial_\mu u_{ij}(\mu))$ and $\|\partial_\mu u(\mu)\|^2_{L^2(\mu)} := \sum_{i,j} \int_{\mathbb{R}^n} |\partial_\mu u_{ij}(\mu)(z)|^2 \mu(dz)$. Furthermore, we call $\partial_\mu u(\mu)(z)$ differentiable at $z \in \mathbb{R}^n$, if all its components are differentiable at z, and set $\partial_z \partial_\mu u(\mu)(z) := (\partial_z \partial_\mu u_{ij}(\mu)(z))$ and $\|\partial_z \partial_\mu u(\mu)\|^2_{L^2(\mu)} := \sum_{i,j} \int_{\mathbb{R}^n} \|\partial_z \partial_\mu u_{ij}(\mu)(z)\|^2 \mu(dz)$. For convenience, we write $u \in C^{1,1}(\mathscr{P}_2, \mathbb{R}^n)$, if the $\mathbb{R}^n$-valued map $\mu \mapsto u(\mu)$ is differentiable at any $\mu \in \mathscr{P}_2$, and $\partial_\mu u(\mu)(z) : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable at any $z \in \mathbb{R}^n$.

For a vector-valued or matrix-valued function $F(t, x, y)$ defined on $[0, \infty) \times \mathbb{R}^n \times \mathbb{R}^m$. For any $u, v \in \{t, x, y\}$, we use $\partial_u F$ to denote the first order partial derivative of F *w.r.t.* component u and $\partial^2_{uv} F$ to denote its second order partial derivatives of F *w.r.t.* components u and v. For convenience, we say an $\mathbb{R}^n$-valued F belongs to $C^{1,2,2}([0, \infty) \times \mathbb{R}^n \times \mathbb{R}^m, \mathbb{R}^n)$, if $\partial_t F(t, x, y)$, $\partial^2_{xx} F(t, x, y)$ and $\partial^2_{yy} F(t, x, y)$ exist for any $(t, x, y) \in [0, \infty) \times \mathbb{R}^n \times \mathbb{R}^m$.

We suppose that for any $T > 0$, there exist constants $C_T, \beta \in (0, \infty)$ and $\gamma_1, \gamma_2 \in (0, 1]$ such that the following conditions hold for all $t, t_1, t_2 \in [0, T]$, $x, x_1, x_2 \in \mathbb{R}^n$, $\mu, \mu_1, \mu_2 \in \mathscr{P}_2$, $y, y_1, y_2 \in \mathbb{R}^m$.

A 1 (Conditions on b, σ, f and g).

$$\begin{aligned} &\big|b(t_1, x_1, \mu_1, y_1) - b(t_2, x_2, \mu_2, y_2)\big| + \big\|\sigma(t_1, x_1, \mu_1) - \sigma(t_2, x_2, \mu_2)\big\| \\ &\quad \le C_T\big[|t_1 - t_2| + |x_1 - x_2| + |y_1 - y_2| + \mathbb{W}_2(\mu_1, \mu_2)\big]; \end{aligned} \tag{2.1}$$

$$\begin{aligned} &\big|f(t_1, x_1, \mu_1, y_1) - f(t_2, x_2, \mu_2, y_2)\big| + \big\|g(t_1, x_1, \mu_1, y_1) - g(t_2, x_2, \mu_2, y_2)\big\| \\ &\quad \le C_T\big[|t_1 - t_2| + |x_1 - x_2| + |y_1 - y_2| + \mathbb{W}_2(\mu_1, \mu_2)\big]; \end{aligned} \tag{2.2}$$

and

$$2\big\langle f(t, x, \mu, y_1) - f(t, x, \mu, y_2), y_1 - y_2\big\rangle + 3\big\|g(t, x, \mu, y_1) - g(t, x, \mu, y_2)\big\|^2 \le -\beta|y_1 - y_2|^2. \tag{2.3}$$

A 2 (Conditions on first-order partial derivatives). *The first-order partial derivatives* $\partial_t b(t,x,\mu,y)$, $\partial_x b(t,x,\mu,y)$, $\partial_\mu b(t,x,\mu,y)$ *and* $\partial_y b(t,x,\mu,y)$ *exist for any* $(t,x,y,\mu)\in[0,\infty)\times\mathbb{R}^n\times\mathbb{R}^m\times\mathscr{P}_2$. *Moreover,*

$$\sup_{t\in[0,T],x\in\mathbb{R}^n,\mu\in\mathscr{P}_2}\big|\partial_t b(t,x,\mu,y_1)-\partial_t b(t,x,\mu,y_2)\big|\le C_T|y_1-y_2|^{\gamma_1};\tag{2.4}$$

$$\sup_{t\in[0,T],x\in\mathbb{R}^n,\mu\in\mathscr{P}_2}\big\|\partial_x b(t,x,\mu,y_1)-\partial_x b(t,x,\mu,y_2)\big\|\le C_T|y_1-y_2|^{\gamma_1};\tag{2.5}$$

$$\sup_{t\in[0,T],x\in\mathbb{R}^n,\mu\in\mathscr{P}_2}\big\|\partial_\mu b(t,x,\mu,y_1)-\partial_\mu b(t,x,\mu,y_2)\big\|_{L^2(\mu)}\le C_T|y_1-y_2|^{\gamma_1};\tag{2.6}$$

$$\sup_{t\in[0,T],x\in\mathbb{R}^n,\mu\in\mathscr{P}_2}\big\|\partial_y b(t,x,\mu,y_1)-\partial_y b(t,x,\mu,y_2)\big\|\le C_T|y_1-y_2|^{\gamma_1}.\tag{2.7}$$

Furthermore, if b is replaced by f and g, the properties (2.4)–(2.7) also hold.

A 3 (Conditions on second-order partial derivatives). *The second-order partial derivatives* $\partial^2_{xx}b(t,x,\mu,y)$, $\partial^2_{xy}b(t,x,\mu,y)$ *and* $\partial^2_{yy}b(t,x,\mu,y)$ *exist* $(t,x,y,\mu)\in[0,\infty)\times\mathbb{R}^n\times\mathbb{R}^m\times\mathscr{P}_2$, *and* $b(t,x,\cdot,y)\in C^{1,1}(\mathscr{P}_2,\mathbb{R}^n)$. *Moreover,* $\partial^2_{xy}b(t,x,\mu,y)$, $\partial^2_{yy}b(t,x,\mu,y)$ *are uniformly bounded and*

$$\sup_{t\in[0,T],x\in\mathbb{R}^n,\mu\in\mathscr{P}_2}\big\|\partial^2_{xx}b(t,x,\mu,y_1)-\partial^2_{xx}b(t,x,\mu,y_2)\big\|\le C_T|y_1-y_2|^{\gamma_2};\tag{2.8}$$

$$\sup_{t\in[0,T],x\in\mathbb{R}^n,\mu\in\mathscr{P}_2}\big\|\partial^2_{xy}b(t,x,\mu,y_1)-\partial^2_{xy}b(t,x,\mu,y_2)\big\|\le C_T|y_1-y_2|^{\gamma_2};\tag{2.9}$$

$$\sup_{t\in[0,T],x\in\mathbb{R}^n,\mu\in\mathscr{P}_2}\big\|\partial^2_{yy}b(t,x,\mu,y_1)-\partial^2_{yy}b(t,x,\mu,y_2)\big\|\le C_T|y_1-y_2|^{\gamma_2};\tag{2.10}$$

$$\sup_{t\in[0,T],x\in\mathbb{R}^n,\mu\in\mathscr{P}_2}\big\|\partial_z\partial_\mu b(t,x,\mu,y_1)-\partial_z\partial_\mu b(t,x,\mu,y_2)\big\|_{L^2(\mu)}\le C_T|y_1-y_2|^{\gamma_2}.\tag{2.11}$$

Furthermore, if b is replaced by f and g, the properties (2.8)–(2.11) also hold, and

$$\begin{aligned}\sup_{t\in[0,T],x\in\mathbb{R}^n,\mu\in\mathscr{P}_2,y\in\mathbb{R}^m}&\max\big\{\big\|\partial^2_{xx}f(t,x,\mu,y)\big\|,\big\|\partial^2_{xy}f(t,x,\mu,y)\big\|,\big\|\partial^2_{yy}f(t,x,\mu,y)\big\|,\\&\big\|\partial^2_{xx}g(t,x,\mu,y)\big\|,\big\|\partial^2_{xy}g(t,x,\mu,y)\big\|,\big\|\partial^2_{yy}g(t,x,\mu,y)\big\|,\\&\big\|\partial_z\partial_\mu f(t,x,\mu,y)\big\|_{L^2(\mu)},\big\|\partial_z\partial_\mu g(t,x,\mu,y)\big\|_{L^2(\mu)}\big\}\le C_T.\end{aligned}$$

Remark 2.1. We here give some comments on the conditions above.

- Conditions (2.1) and (2.2) imply that for any $T>0$, there exists $C_T>0$ such that for any $x\in\mathbb{R}^n$, $y\in\mathbb{R}^m$, $\mu\in\mathscr{P}_2$, $t\in[0,T]$,

$$\big|b(t,x,\mu,y)\big|+\big\|\sigma(t,x,\mu)\big\|\le C_T\big\{1+|x|+|y|+\big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\}\tag{2.12}$$

  and

$$\big|f(t,x,\mu,y)\big|+\big\|g(t,x,\mu,y)\big\|\le C_T\big\{1+|x|+|y|+\big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\}.\tag{2.13}$$

- Conditions (2.2) and (2.3) imply that for any $T>0$, there exists $C_T>0$ such that for any $x\in\mathbb{R}^n$, $y\in\mathbb{R}^m$, $\mu\in\mathscr{P}_2$, $t\in[0,T]$,

$$2\big\langle f(t,x,\mu,y),y\big\rangle+3\big\|g(t,x,\mu,y)\big\|^2\le\frac{-\beta}{2}|y|^2+C_T\big[1+|x|^2+\mu\big(|\cdot|^2\big)\big].\tag{2.14}$$

- Condition (2.3) is used to guarantee the existence and uniqueness of an invariant measure for the frozen equation (see Eq. (2.19) below) and the solution of system (1.1) has finite fourth moment.
- Using the time discretization approach, to prove the strong convergence order we need assumptions A1 and A2. However, if using the technique of Poisson equation to prove the strong convergence order, we needs the assumption A3 additionally.

The following theorem is the existence and uniqueness of strong solutions for system (1.1), which can be obtained by using the result due to Wang in [38] and whose detailed proof will be presented in the Appendix.

Theorem 2.2. *Suppose that conditions* (2.1) *and* (2.2) *hold. For any* $\epsilon > 0$, *any given initial value* $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, *there exists a unique solution* $\{(X_t^\epsilon, Y_t^\epsilon), t \geq 0\}$ *to system* (1.1) *and for all* $T > 0$, $(X^\epsilon, Y^\epsilon) \in C([0,T];\mathbb{R}^n) \times C([0,T];\mathbb{R}^m)$, $\mathbb{P}$*-a.s. and*

$$\begin{cases} X_t^\epsilon = x + \int_0^t b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon)\,ds + \int_0^t \sigma(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon})\,dW_s^1, \\ Y_t^\epsilon = y + \frac{1}{\epsilon}\int_0^t f(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon)\,ds + \frac{1}{\sqrt{\epsilon}}\int_0^t g(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon)\,dW_s^2. \end{cases} \tag{2.15}$$

Now we formulate our first main result.

Theorem 2.3. *Suppose that assumptions* A1 *and* A2 *hold. Then for any* $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ *and* $T > 0$, *we have*

$$\sup_{t\in[0,T]} \mathbb{E}\big|X_t^\epsilon - \bar{X}_t\big|^2 \leq C\epsilon^{2/3}, \tag{2.16}$$

where C *is a constant depending on* T, $|x|$, $|y|$. *Furthermore, if there is no noise in the slow equation* (*i.e.*, $\sigma \equiv 0$), *we have*

$$\sup_{t\in[0,T]} \mathbb{E}\big|X_t^\epsilon - \bar{X}_t\big|^2 \leq C\epsilon. \tag{2.17}$$

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

$$\begin{cases} d\bar{X}_t = \bar{b}(t, \bar{X}_t, \mathscr{L}_{\bar{X}_t})\,dt + \sigma(t, \bar{X}_t, \mathscr{L}_{\bar{X}_t})\,dW_t^1, \\ \bar{X}_0 = x, \end{cases} \tag{2.18}$$

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

$$\begin{cases} dY_s = f(t,x,\mu,Y_s)\,ds + g(t,x,\mu,Y_s)\,d\tilde{W}_s^2, \\ Y_0 = y, \end{cases} \tag{2.19}$$

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

Remark 2.4. The estimates (2.16) and (2.17) imply that the slow component X_t^ϵ strongly converges to the solution $\bar{X}_t$ of the corresponding averaged equation with convergence order $\epsilon^{1/3}$ and $\epsilon^{1/2}$ respectively. Usually, the convergence order $\epsilon^{1/2}$ should be optimal. Hence, under more regularity conditions on the coefficients, we will use the technique of Poisson equation to obtain the optimal convergence order in the general case (i.e., $\sigma \neq 0$), which is stated in the following theorem.

Theorem 2.5. *Suppose that assumptions* A1–A3 *hold. Then for any* $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ *and* $T > 0$, *we have*

$$\sup_{t\in[0,T]} \mathbb{E}\big|X_t^\epsilon - \bar{X}_t\big|^2 \leq C\epsilon, \tag{2.20}$$

where C *is a constant depending on* T, $|x|$, $|y|$, *and* $\bar{X}$ *is the solution of the corresponding averaged equation* (2.18).

3. Proof of Theorem 2.3

In this section, we intend to use the approach of time discretization to get the strong convergence order. The proof consists of four parts, each of which is presented in the respective subsection below. In Section 3.1, we give some a-priori estimates of the solution $(X_t^\epsilon, Y_t^\epsilon)$. In Section 3.2, we introduce an auxiliary process $(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon)$, and obtain the convergence rate of the difference process $X_t^\epsilon - \hat{X}_t^\epsilon$. We study the frozen equation, and prove the exponential ergodicity of the corresponding semigroup in Section 3.3. In the final subsection, we prove a crucial estimate for $\sup_{t\in[0,T]}\mathbb{E}|\hat{X}_t^\epsilon - \bar{X}_t|$ which relies on somewhat delicate arguments. Note that we always assume conditions A1 and A2 to hold, and the initial values $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ are fixed in this section.

3.1. *Some a-priori estimates for* $(X_t^\epsilon, Y_t^\epsilon)$

Firstly, we prove some uniform bounds *w.r.t.* $\epsilon \in (0,1)$ for the 4th moment of the solution $(X_t^\epsilon, Y_t^\epsilon)$ to system (1.1).

Lemma 3.1. *For any* $T > 0$, *there exists a constant* $C_T > 0$ *such that*

$$\sup_{\epsilon\in(0,1)} \sup_{t\in[0,T]} \mathbb{E}\big|X_t^\epsilon\big|^4 \le C_T\big(1+|x|^4+|y|^4\big)$$

and

$$\sup_{\epsilon\in(0,1)} \sup_{t\in[0,T]} \mathbb{E}\big|Y_t^\epsilon\big|^4 \le C_T\big(1+|x|^4+|y|^4\big).$$

Proof. By Itô's formula and estimate (2.12), we obtain for any $t\in[0,T]$,

$$\begin{aligned}
\big|X_t^\epsilon\big|^4 &= |x|^4 + 4\int_0^t \big|X_s^\epsilon\big|^2\big\langle X_s^\epsilon, b\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\big\rangle\,ds + 4\int_0^t \big|X_s^\epsilon\big|^2\big\langle X_s^\epsilon, \sigma\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\,dW_s^1\big\rangle \\
&\quad + 4\int_0^t \big|\big\langle X_s^\epsilon, \sigma\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\big\rangle\big|^2\,ds + 2\int_0^t \big|X_s^\epsilon\big|^2\big\|\sigma\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\big\|^2\,ds \\
&\le |x|^4 + C_T\int_0^t \big(1+\big|X_s^\epsilon\big|^4+\big|Y_s^\epsilon\big|^4+\big[\mathcal{L}_{X_s^\epsilon}\big(|\cdot|^2\big)\big]^2\big)\,ds + 4\int_0^t \big|X_s^\epsilon\big|^2\big\langle X_s^\epsilon, \sigma\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\,dW_s^1\big\rangle.
\end{aligned}$$

Note that $\mathcal{L}_{X_s^\epsilon}(|\cdot|^2) = \mathbb{E}|X_s^\epsilon|^2$. Hence, we have

$$\sup_{t\in[0,T]} \mathbb{E}\big|X_t^\epsilon\big|^4 \le C_T\big(|x|^4+1\big) + C_T\int_0^T \mathbb{E}\big|Y_t^\epsilon\big|^4\,dt + C_T\int_0^T \mathbb{E}\big|X_t^\epsilon\big|^4\,dt. \tag{3.1}$$

Using Itô formula again and taking expectation, we get

$$\begin{aligned}
\mathbb{E}\big|Y_t^\epsilon\big|^4 &= |y|^4 + \frac{4}{\epsilon}\int_0^t \mathbb{E}\big[\big|Y_s^\epsilon\big|^2\big\langle f\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big), Y_s^\epsilon\big\rangle\big]\,ds \\
&\quad + \frac{2}{\epsilon}\int_0^t \mathbb{E}\big[\big|Y_s^\epsilon\big|^2\big\|g\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\big\|^2\big]\,ds + \frac{4}{\epsilon}\int_0^t \mathbb{E}\big|\big\langle Y_s^\epsilon, g\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\big\rangle\big|^2\,ds.
\end{aligned}$$

By (2.14), there exists $\beta > 0$ such that for any $t\in[0,T]$,

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}\big|Y_t^\epsilon\big|^4 &\le \frac{1}{\epsilon}\mathbb{E}\big[4\big|Y_t^\epsilon\big|^2\big\langle f\big(t, X_t^\epsilon, \mathcal{L}_{X_t^\epsilon}, Y_t^\epsilon\big), Y_t^\epsilon\big\rangle + 6\big|Y_t^\epsilon\big|^2\big\|g\big(t, X_t^\epsilon, \mathcal{L}_{X_t^\epsilon}, Y_t^\epsilon\big)\big\|^2\big] \\
&\le -\frac{\beta}{\epsilon}\mathbb{E}\big|Y_t^\epsilon\big|^4 + \frac{C_T}{\epsilon}\big(\mathbb{E}\big|X_t^\epsilon\big|^4+1\big).
\end{aligned}$$

The comparison theorem implies

$$\begin{aligned}
\mathbb{E}\big|Y_t^\epsilon\big|^4 &\le |y|^4 e^{-\frac{\beta t}{\epsilon}} + \frac{C_T}{\epsilon}\int_0^t e^{-\frac{\beta(t-s)}{\epsilon}}\big(\mathbb{E}\big|X_s^\epsilon\big|^4+1\big)\,ds \\
&\le |y|^4 + C_T\Big(\sup_{s\in[0,t]} \mathbb{E}\big|X_s^\epsilon\big|^4+1\Big).
\end{aligned} \tag{3.2}$$

This and (3.1) yield

$$\sup_{t\in[0,T]} \mathbb{E}\big|X_t^\epsilon\big|^4 \le C_T\big(|x|^4+|y|^4+1\big) + C_T\int_0^T \sup_{s\in[0,t]} \mathbb{E}\big|X_s^\epsilon\big|^4\,dt.$$

Then by Grownall's inequality, we finally obtain

$$\sup_{t\in[0,T]} \mathbb{E}\big|X_t^\epsilon\big|^4 \le C_T\big(|x|^4+|y|^4+1\big),$$

which also gives

$$\sup_{t\in[0,T]} \mathbb{E}\big|Y_t^\epsilon\big|^4 \le C_T\big(|x|^4+|y|^4+1\big).$$

The proof is complete. □

Lemma 3.2. *For any $T>0$, $0\le t\le t+h\le T$ and $\epsilon\in(0,1)$, there exists a constant $C_T>0$ such that*

$$\mathbb{E}\big|X_{t+h}^\epsilon - X_t^\epsilon\big|^2 \le C_T\big(1+|x|^2+|y|^2\big)h.$$

Proof. It is easy to see that

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

Then by estimate (2.12) and Lemma 3.1, we obtain

$$\begin{aligned}
\mathbb{E}\big|X_{t+h}^\epsilon - X_t^\epsilon\big|^2 &\le C\mathbb{E}\bigg|\int_t^{t+h} b\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\,ds\bigg|^2 + C\mathbb{E}\bigg|\int_t^{t+h} \sigma\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\,dW_s^1\bigg|^2 \\
&\le C\mathbb{E}\bigg|\int_t^{t+h} \big|b\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\big|\,ds\bigg|^2 + C\int_t^{t+h} \mathbb{E}\big\|\sigma\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\big\|^2\,ds \\
&\le C_T h\mathbb{E}\int_t^{t+h}\big(1+\big|X_s^\epsilon\big|^2+\big|Y_s^\epsilon\big|^2+\mathbb{E}\big|X_s^\epsilon\big|^2\big)\,ds + C_T\int_t^{t+h}\mathbb{E}\big(1+\big|X_s^\epsilon\big|^2+\mathbb{E}\big|X_s^\epsilon\big|^2\big)\,ds \\
&\le C_T\big(1+|x|^2+|y|^2\big)h.
\end{aligned}$$

The proof is complete. □

3.2. *Estimates for the auxiliary process $(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon)$*

Following the idea of Khasminskii in [24], we introduce an auxiliary process $(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon)\in\mathbb{R}^n\times\mathbb{R}^m$ and divide $[0,T]$ into intervals of size δ, where δ is a fixed positive number depending on ϵ, which will be chosen later. We construct a process $\hat{Y}_t^\epsilon$ with initial value $\hat{Y}_0^\epsilon = Y_0^\epsilon = y$ such that for $t\in[k\delta, \min((k+1)\delta, T)]$,

$$\hat{Y}_t^\epsilon = \hat{Y}_{k\delta}^\epsilon + \frac{1}{\epsilon}\int_{k\delta}^t f\big(k\delta, X_{k\delta}^\epsilon, \mathcal{L}_{X_{k\delta}^\epsilon}, \hat{Y}_s^\epsilon\big)\,ds + \frac{1}{\sqrt{\epsilon}}\int_{k\delta}^t g\big(k\delta, X_{k\delta}^\epsilon, \mathcal{L}_{X_{k\delta}^\epsilon}, \hat{Y}_s^\epsilon\big)\,dW_s^2,$$

i.e.,

$$\hat{Y}_t^\epsilon = y + \frac{1}{\epsilon}\int_0^t f\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon\big)\,ds + \frac{1}{\sqrt{\epsilon}}\int_0^t g\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon\big)\,dW_s^2,$$

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

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

By the construction of $\hat{Y}_t^\epsilon$ and by similar argument as in the proof of Lemma 3.1, it is easy to obtain the following estimate, whose proof is omitted here.

Lemma 3.3. *For any $T>0$, there exists a constant $C_T>0$ such that*

$$\sup_{\epsilon\in(0,1)}\sup_{t\in[0,T]} \mathbb{E}\big|\hat{Y}_t^\epsilon\big|^4 \le C_T\big(1+|x|^4+|y|^4\big).$$

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

Lemma 3.4. *For any $T > 0$, there exists a constant $C_T > 0$ such that*

$$\sup_{\epsilon\in(0,1)}\sup_{t\in[0,T]}\mathbb{E}\big|Y_t^\epsilon-\hat{Y}_t^\epsilon\big|^2\le C_T\big(1+|x|^2+|y|^2\big)\delta.$$

Proof. Note that

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

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

$$\begin{aligned}&\mathbb{E}\big|Y_t^\epsilon-\hat{Y}_t^\epsilon\big|^2\\&\quad=\frac{1}{\epsilon}\int_0^t\mathbb{E}\big[2\big\langle f\big(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon},Y_s^\epsilon\big)-f\big(s(\delta),X_{s(\delta)}^\epsilon,\mathscr{L}_{X_{s(\delta)}^\epsilon},\hat{Y}_s^\epsilon\big),Y_s^\epsilon-\hat{Y}_s^\epsilon\big\rangle\big]ds\\&\qquad+\frac{1}{\epsilon}\int_0^t\mathbb{E}\big\|g\big(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon},Y_s^\epsilon\big)-g\big(s(\delta),X_{s(\delta)}^\epsilon,\mathscr{L}_{X_{s(\delta)}^\epsilon},\hat{Y}_s^\epsilon\big)\big\|^2ds\\&\quad\le\frac{1}{\epsilon}\int_0^t\mathbb{E}\big[2\big\langle f\big(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon},Y_s^\epsilon\big)-f\big(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon},\hat{Y}_s^\epsilon\big),Y_s^\epsilon-\hat{Y}_s^\epsilon\big\rangle\\&\qquad+3\big\|g\big(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon},Y_s^\epsilon\big)-g\big(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon},\hat{Y}_s^\epsilon\big)\big\|^2\big]ds\\&\qquad+\frac{2}{\epsilon}\int_0^t\mathbb{E}\big[\big\langle f\big(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon},\hat{Y}_s^\epsilon\big)-f\big(s(\delta),X_{s(\delta)}^\epsilon,\mathscr{L}_{X_{s(\delta)}^\epsilon},\hat{Y}_s^\epsilon\big),Y_s^\epsilon-\hat{Y}_s^\epsilon\big\rangle\big]ds\\&\qquad+\frac{3}{\epsilon}\int_0^t\mathbb{E}\big\|g\big(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon},\hat{Y}_s^\epsilon\big)-g\big(s(\delta),X_{s(\delta)}^\epsilon,\mathscr{L}_{X_{s(\delta)}^\epsilon},\hat{Y}_s^\epsilon\big)\big\|^2ds.\end{aligned}$$

Then using the following estimate

$$\mathbb{W}_2(\mathscr{L}_{X_s^\epsilon},\mathscr{L}_{X_{s(\delta)}^\epsilon})^2\le\mathbb{E}\big|X_s^\epsilon-X_{s(\delta)}^\epsilon\big|^2$$

and the conditions (2.2), (2.3), there exists $\beta>0$ such that for any $t\in[0,T]$,

$$\begin{aligned}\frac{d}{dt}\mathbb{E}\big|Y_t^\epsilon-\hat{Y}_t^\epsilon\big|^2&\le\frac{-\beta}{\epsilon}\mathbb{E}\big|Y_t^\epsilon-\hat{Y}_t^\epsilon\big|^2+\frac{C_T}{\epsilon}\mathbb{E}\big[\delta^2+\big|X_t^\epsilon-X_{t(\delta)}^\epsilon\big|^2+\mathbb{W}_2(\mathscr{L}_{X_t^\epsilon},\mathscr{L}_{X_{t(\delta)}^\epsilon})^2\big]\\&\le-\frac{\beta}{\epsilon}\mathbb{E}\big|Y_t^\epsilon-\hat{Y}_t^\epsilon\big|^2+\frac{C_T}{\epsilon}\mathbb{E}\big|X_t^\epsilon-X_{t(\delta)}^\epsilon\big|^2+\frac{C_T\delta^2}{\epsilon}.\end{aligned}$$

Finally, the comparison theorem and Lemma 3.2 yield

$$\begin{aligned}\mathbb{E}\big|Y_t^\epsilon-\hat{Y}_t^\epsilon\big|^2&\le\frac{C_T}{\epsilon}\int_0^t e^{-\frac{\beta(t-s)}{\epsilon}}\mathbb{E}\big|X_s^\epsilon-X_{s(\delta)}^\epsilon\big|^2ds+\frac{C_T\delta^2}{\epsilon}\int_0^t e^{-\frac{\beta(t-s)}{\epsilon}}ds\\&\le C_T\big(1+|x|^2+|y|^2\big)\delta.\end{aligned}$$

The proof is complete. □

Lemma 3.5. *For any $T > 0$, there exists a constant $C_T > 0$ such that*

$$\sup_{t\in[0,T]}\mathbb{E}\big|X_t^\epsilon-\hat{X}_t^\epsilon\big|^2\le C_T\big(1+|x|^2+|y|^2\big)\delta.$$

Proof. Recall that

$$X_t^\epsilon=x+\int_0^t b\big(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon},Y_s^\epsilon\big)ds+\int_0^t\sigma\big(s,X_s^\epsilon,\mathscr{L}_{X_s^\epsilon}\big)dW_s^1$$

and that

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

Then we have

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

By Lemmas 3.2 and 3.4, we obtain

$$\begin{aligned}\sup_{t\in[0,T]} \mathbb{E}\big|X_t^\epsilon - \hat{X}_t^\epsilon\big|^2 &\le \mathbb{E}\Big[\int_0^T \big|b\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big) - b\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon\big)\big|\,ds\Big]^2 \\ &\le C_T \mathbb{E}\int_0^T \big(\delta^2 + \big|X_s^\epsilon - X_{s(\delta)}^\epsilon\big|^2 + \mathbb{W}_2(\mathcal{L}_{X_s^\epsilon}, \mathcal{L}_{X_{s(\delta)}^\epsilon})^2 + \big|Y_s^\epsilon - \hat{Y}_s^\epsilon\big|^2\big)\,ds \\ &\le C_T\big(1 + |x|^2 + |y|^2\big)\delta.\end{aligned}$$

The proof is complete. □

3.3. *The frozen equation*

We first introduce the frozen equation associated to the fast motion for fixed $t \ge 0$, $x \in \mathbb{R}^n$ and $\mu \in \mathcal{P}_2$,

$$\begin{cases} dY_s = f(t, x, \mu, Y_s)\,ds + g(t, x, \mu, Y_s)\,d\tilde{W}_s^2, \\ Y_0 = y, \end{cases} \tag{3.3}$$

where $\{\tilde{W}_s^2\}_{s\ge 0}$ is a d_2-dimensional Brownian motion on another complete probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ and $\{\tilde{\mathcal{F}}_t, t \ge 0\}$ is the natural filtration generated by $\tilde{W}_t^2$.

Under the conditions (2.2) and (2.3), it is easy to prove for any initial data $y \in \mathbb{R}^m$ that Eq. (3.3) has a unique strong solution $\{Y_s^{t,x,\mu,y}\}_{s\ge 0}$, which is a homogeneous Markov process. Moreover, for any $t \in [0, T]$, $\sup_{s\ge 0} \tilde{\mathbb{E}}|Y_s^{t,x,\mu,y}|^2 \le C_T[1 + |x|^2 + |y|^2 + \mu(|\cdot|^2)]$.

Let $\{P_s^{t,x,\mu}\}_{s\ge 0}$ be the transition semigroup of $Y_s^{t,x,\mu,y}$, i.e., for any bounded measurable function $\varphi : \mathbb{R}^m \to \mathbb{R}$,

$$P_s^{t,x,\mu}\varphi(y) := \tilde{\mathbb{E}}\varphi\big(Y_s^{t,x,\mu,y}\big), \quad y \in \mathbb{R}^m, s \ge 0,$$

where $\tilde{\mathbb{E}}$ is the expectation on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$. Then e.g. by [26, Theorem 4.3.9], under the assumption A1, it is easy to see that $P_s^{t,x,\mu}$ has a unique invariant measure $\nu^{t,x,\mu}$ satisfying

$$\sup_{t\in[0,T]} \int_{\mathbb{R}^m} |y|\nu^{t,x,\mu}(dy) \le C_T\big\{1 + |x| + \big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\}.$$

Lemma 3.6. *For any $T > 0$, $s > 0$, $t_i \in [0, T]$, $x_i \in \mathbb{R}^n$, $\mu_i \in \mathcal{P}_2$ and $y_i \in \mathbb{R}^m$, $i = 1, 2$, we have*

$$\tilde{\mathbb{E}}\big|Y_s^{t_1,x_1,\mu_1,y_1} - Y_s^{t_2,x_2,\mu_2,y_2}\big|^2 \le e^{-\beta s}|y_1 - y_2|^2 + C_T\big[|t_1 - t_2|^2 + |x_1 - x_2|^2 + \mathbb{W}_2(\mu_1, \mu_2)^2\big].$$

Proof. Note that

$$\begin{aligned}Y_s^{t_1,x_1,\mu_1,y_1} - Y_s^{t_2,x_2,\mu_2,y_2} = y_1 - y_2 &+ \int_0^s \big[f\big(t_1, x_1, \mu_1, Y_r^{t_1,x_1,\mu_1,y_1}\big) - f\big(t_2, x_2, \mu_2, Y_r^{t_2,x_2,\mu_2,y_2}\big)\big]\,dr \\ &+ \int_0^s \big[g\big(t_1, x_1, \mu_1, Y_r^{t_1,x_1,\mu_1,y_1}\big) - g\big(t_2, x_2, \mu_2, Y_r^{t_2,x_2,\mu_2,y_2}\big)\big]\,d\tilde{W}_r^2.\end{aligned}$$

By Itô's formula we have

$$\begin{aligned}
&\tilde{\mathbb{E}}\big|Y_s^{t_1,x_1,\mu_1,y_1}-Y_s^{t_2,x_2,\mu_2,y_2}\big|^2\\
&\quad=\int_0^s\tilde{\mathbb{E}}\big[2\big\langle f\big(t_1,x_1,\mu_1,Y_r^{t_1,x_1,\mu_1,y_1}\big)-f\big(t_2,x_2,\mu_2,Y_r^{t_2,x_2,\mu_2,y_2}\big),Y_r^{t_1,x_1,\mu_1,y_1}-Y_r^{t_2,x_2,\mu_2,y_2}\big\rangle\\
&\qquad+\big\|g\big(t_1,x_1,\mu_1,Y_r^{t_1,x_1,\mu_1,y_1}\big)-g\big(t_2,x_2,\mu_2,Y_r^{t_2,x_2,\mu_2,y_2}\big)\big\|^2\big]\,dr.
\end{aligned}$$

Then by Young's inequality and conditions (2.2) and (2.3), there exists $\beta>0$ such that

$$\begin{aligned}
&\frac{d}{ds}\tilde{\mathbb{E}}\big|Y_s^{t_1,x_1,\mu_1,y_1}-Y_s^{t_2,x_2,\mu_2,y_2}\big|^2\\
&\quad=\tilde{\mathbb{E}}\big[2\big\langle f\big(t_1,x_1,\mu_1,Y_s^{t_1,x_1,\mu_1,y_1}\big)-f\big(t_2,x_2,\mu_2,Y_s^{t_2,x_2,\mu_2,y_2}\big),Y_s^{t_1,x_1,\mu_1,y_1}-Y_s^{t_2,x_2,\mu_2,y_2}\big\rangle\\
&\qquad+\big\|g\big(t_1,x_1,\mu_1,Y_s^{t_1,x_1,\mu_1,y_1}\big)-g\big(t_2,x_2,\mu_2,Y_s^{t_2,x_2,\mu_2,y_2}\big)\big\|^2\big]\\
&\quad\le\tilde{\mathbb{E}}\big[2\big\langle f\big(t_1,x_1,\mu_1,Y_s^{t_1,x_1,\mu_1,y_1}\big)-f\big(t_1,x_1,\mu_1,Y_s^{t_2,x_2,\mu_2,y_2}\big),Y_s^{t_1,x_1,\mu_1,y_1}-Y_s^{t_2,x_2,\mu_2,y_2}\big\rangle\\
&\qquad+3\big\|g\big(t_1,x_1,\mu_1,Y_s^{t_1,x_1,\mu_1,y_1}\big)-g\big(t_1,x_1,\mu_1,Y_s^{t_2,x_2,\mu_2,y_2}\big)\big\|^2\big]\\
&\qquad+\tilde{\mathbb{E}}\big[2\big\langle f\big(t_1,x_1,\mu_1,Y_s^{t_2,x_2,\mu_2,y_2}\big)-f\big(t_2,x_2,\mu_2,Y_s^{t_2,x_2,\mu_2,y_2}\big),Y_s^{t_1,x_1,\mu_1,y_1}-Y_s^{t_2,x_2,\mu_2,y_2}\big\rangle\big]\\
&\qquad+\frac{1}{3}\tilde{\mathbb{E}}\big\|g\big(t_1,x_1,\mu_1,Y_s^{t_2,x_2,\mu_2,y_2}\big)-g\big(t_2,x_2,\mu_2,Y_s^{t_2,x_2,\mu_2,y_2}\big)\big\|^2\\
&\quad\le-\beta\tilde{\mathbb{E}}\big|Y_s^{t_1,x_1,\mu_1,y_1}-Y_s^{t_2,x_2,\mu_2,y_2}\big|^2+C_T\big[|t_1-t_2|^2+|x_1-x_2|^2+\mathbb{W}_2(\mu_1,\mu_2)^2\big].
\end{aligned}$$

Hence, the comparison theorem yields for any $s\ge 0$,

$$\tilde{\mathbb{E}}\big|Y_s^{t_1,x_1,\mu_1,y_1}-Y_s^{t_2,x_2,\mu_2,y_2}\big|^2\le e^{-\beta s}|y_1-y_2|^2+C_T\big[|t_1-t_2|^2+|x_1-x_2|^2+\mathbb{W}_2(\mu_1,\mu_2)^2\big].$$

The proof is complete. □

Proposition 3.7. *For any $T>0$, $t\in[0,T]$, $x\in\mathbb{R}^n$, $\mu\in\mathscr{P}_2$, $s\ge 0$ and $y\in\mathbb{R}^m$,*

$$\big|\tilde{\mathbb{E}}b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)-\bar{b}(t,x,\mu)\big|\le C_Te^{-\frac{\beta s}{2}}\big\{1+|x|+|y|+\big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\},\tag{3.4}$$

where $\bar{b}(t,x,\mu)=\int_{\mathbb{R}^m}b(t,x,\mu,z)\nu^{t,x,\mu}(dz)$.

Proof. By the definition of an invariant measure and Lemma 3.6, for any $s\ge 0$ we have

$$\begin{aligned}
\big|\tilde{\mathbb{E}}b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)-\bar{b}(t,x,\mu)\big|&=\Big|\tilde{\mathbb{E}}b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)-\int_{\mathbb{R}^m}b(t,x,\mu,z)\nu^{t,x,\mu}(dz)\Big|\\
&=\Big|\int_{\mathbb{R}^m}\big[\tilde{\mathbb{E}}b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)-\tilde{\mathbb{E}}b\big(t,x,\mu,Y_s^{t,x,\mu,z}\big)\big]\nu^{t,x,\mu}(dz)\Big|\\
&\le C_T\int_{\mathbb{R}^m}\tilde{\mathbb{E}}\big|Y_s^{t,x,\mu,y}-Y_s^{t,x,\mu,z}\big|\nu^{t,x,\mu}(dz)\\
&\le C_Te^{-\frac{\beta s}{2}}\int_{\mathbb{R}^m}|y-z|\nu^{t,x,\mu}(dz)\\
&\le C_Te^{-\frac{\beta s}{2}}\big\{1+|x|+|y|+\big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\}.
\end{aligned}$$

The proof is complete. □

3.4. *The averaged equation*

We can introduce the averaged equation as follows,

$$\begin{cases} d\bar{X}_t = \bar{b}(t, \bar{X}_t, \mathcal{L}_{\bar{X}_t})\,dt + \sigma(t, \bar{X}_t, \mathcal{L}_{\bar{X}_t})\,dW_t^1, \\ \bar{X}_0 = x \in \mathbb{R}^n, \end{cases} \tag{3.5}$$

with

$$\bar{b}(t, x, \mu) = \int_{\mathbb{R}^m} b(t, x, \mu, z)\nu^{t,x,\mu}(dz),$$

where $\nu^{t,x,\mu}$ is the unique invariant measure for Eq. (3.3).

The following lemma gives the existence, uniqueness and uniformly estimates for the solution of Eq. (3.5), whose proof will be presented in the Appendix.

Lemma 3.8. *For any $x \in \mathbb{R}^n$, Eq. (3.5) has a unique solution $\bar{X}_t$. Moreover, for any $T > 0$, there exists a constant $C_T > 0$ such that*

$$\sup_{t\in[0,T]} \mathbb{E}|\bar{X}_t|^2 \le C_T\big(1 + |x|^2\big). \tag{3.6}$$

Now, we estimate the error between the auxiliary process $\hat{X}_t^\epsilon$ and the solution $\bar{X}_t$ of the averaged equation.

Lemma 3.9. *For any $T > 0$, there exists a constant $C_T > 0$ such that*

$$\sup_{t\in[0,T]} \mathbb{E}\big|\hat{X}_t^\epsilon - \bar{X}_t\big|^2 \le C_T\big(1 + |x|^3 + |y|^3\big)\bigg(\frac{\epsilon}{\delta^{1/2}} + \epsilon + \frac{\epsilon^2}{\delta} + \delta\bigg).$$

Proof. We will divide the proof into three steps.

Step 1. Recall that

$$\begin{aligned}
\hat{X}_t^\epsilon - \bar{X}_t &= \int_0^t \big[b\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon\big) - \bar{b}(s, \bar{X}_s, \mathcal{L}_{\bar{X}_s})\big]\,ds \\
&\quad + \int_0^t \big[\sigma\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big) - \sigma(s, \bar{X}_s, \mathcal{L}_{\bar{X}_s})\big]\,dW_s^1 \\
&= \int_0^t \big[b\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon\big) - \bar{b}\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}\big)\big]\,ds \\
&\quad + \int_0^t \big[\bar{b}\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}\big) - \bar{b}\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\big]\,ds \\
&\quad + \int_0^t \big[\bar{b}\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big) - \bar{b}(s, \bar{X}_s, \mathcal{L}_{\bar{X}_s})\big]\,ds \\
&\quad + \int_0^t \big[\sigma\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big) - \sigma(s, \bar{X}_s, \mathcal{L}_{\bar{X}_s})\big]\,dW_s^1.
\end{aligned}$$

Then it is esay to see that for any $t \in [0, T]$, we have

$$\begin{aligned}
\mathbb{E}\big|\hat{X}_t^\epsilon - \bar{X}_t\big|^2 &\le C\mathbb{E}\bigg|\int_0^t \big[b\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon\big) - \bar{b}\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}\big)\big]\,ds\bigg|^2 \\
&\quad + C_T\mathbb{E}\int_0^t \big|\bar{b}\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}\big) - \bar{b}\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\big|^2\,ds \\
&\quad + C_T\mathbb{E}\int_0^t \big|\bar{b}\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big) - \bar{b}(s, \bar{X}_s, \mathcal{L}_{\bar{X}_s})\big|^2\,ds
\end{aligned}$$

$$+ C\mathbb{E}\int_0^t \big\|\sigma\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big) - \sigma(s, \bar{X}_s, \mathcal{L}_{\bar{X}_s})\big\|^2 ds$$

$$:= \sum_{i=1}^4 I_i(t). \tag{3.7}$$

For $I_2(t)$ we have by the Lipschitz property of $\bar{b}(\cdot,\cdot,\cdot)$ (see (A.2) below) that

$$\begin{aligned}\sup_{t\in[0,T]} I_2(t) &\le C_T\delta^2 + \mathbb{E}\int_0^T \big|X_{s(\delta)}^\epsilon - X_s^\epsilon\big|^2 ds\\ &\le C_T\big(1+|x|^2+|y|^2\big)\delta.\end{aligned} \tag{3.8}$$

For $I_i(t)$, $i = 3, 4$, Lemma 3.5 implies

$$\begin{aligned}\sup_{t\in[0,T]} I_3(t) &\le C_T\int_0^T \mathbb{E}\big|X_t^\epsilon - \bar{X}_t\big|^2 dt\\ &\le C_T\int_0^T \mathbb{E}\big|X_t^\epsilon - \hat{X}_t^\epsilon\big|^2 dt + C_T\int_0^T \mathbb{E}\big|\hat{X}_t^\epsilon - \bar{X}_t\big|^2 dt\\ &\le C_T\big(1+|x|^2+|y|^2\big)\delta + C_T\int_0^T \mathbb{E}\big|\hat{X}_t^\epsilon - \bar{X}_t\big|^2 dt.\end{aligned} \tag{3.9}$$

Similarly, by condition (2.1),

$$\sup_{t\in[0,T]} I_4(t) \le C_T\big(1+|x|^2+|y|^2\big)\delta + C_T\int_0^T \mathbb{E}\big|\hat{X}_t^\epsilon - \bar{X}_t\big|^2 dt. \tag{3.10}$$

Therefore, (3.7)–(3.10) yield

$$\begin{aligned}\sup_{t\in[0,T]} \mathbb{E}\big|\hat{X}_t^\epsilon - \bar{X}_t\big|^2 \le{}& \sup_{t\in[0,T]} I_1(t) + C_T\big(1+|x|^2+|y|^2\big)\delta\\ &+ C_T\int_0^T \mathbb{E}\big|\hat{X}_t^\epsilon - \bar{X}_t\big|^2 dt.\end{aligned} \tag{3.11}$$

Then combining this with the following estimate of $I_1(t)$,

$$\sup_{t\in[0,T]} I_1(t) \le C_T\big(1+|x|^3+|y|^3\big)\left(\frac{\epsilon}{\delta^{1/2}} + \epsilon + \frac{\epsilon^2}{\delta} + \delta^2\right), \tag{3.12}$$

which will be proved in Step 2, we obtain

$$\sup_{t\in[0,T]} \mathbb{E}\big|\hat{X}_t^\epsilon - \bar{X}_t\big|^2 \le C_T\big(1+|x|^3+|y|^3\big)\left(\frac{\epsilon}{\delta^{1/2}} + \epsilon + \frac{\epsilon^2}{\delta} + \delta\right) + \int_0^T \mathbb{E}\big|\hat{X}_t^\epsilon - \bar{X}_t\big|^2 dt.$$

Hence, the Grownall's inequality yields

$$\sup_{t\in[0,T]} \mathbb{E}\big|\hat{X}_t^\epsilon - \bar{X}_t\big|^2 \le C_T\big(1+|x|^3+|y|^3\big)\left(\frac{\epsilon}{\delta^{1/2}} + \epsilon + \frac{\epsilon^2}{\delta} + \delta\right),$$

which completes the proof.

Step 2. In this step, we intend to prove estimate (3.12). Note that

$$\begin{aligned}&\left|\int_0^t \big[b\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon\big) - \bar{b}\big(s(\delta), X_{s(\delta)}^\epsilon, \mathcal{L}_{X_{s(\delta)}^\epsilon}\big)\big] ds\right|^2\\ &\quad\le 2\left|\sum_{k=0}^{[t/\delta]-1}\int_{k\delta}^{(k+1)\delta}\big[b\big(k\delta, X_{k\delta}^\epsilon, \mathcal{L}_{X_{k\delta}^\epsilon}, \hat{Y}_s^\epsilon\big) - \bar{b}\big(k\delta, X_{k\delta}^\epsilon, \mathcal{L}_{X_{k\delta}^\epsilon}\big)\big] ds\right|^2\end{aligned}$$

$$+2\left|\int_{t(\delta)}^{t}\left[b\left(t(\delta),X^{\epsilon}_{t(\delta)},\mathcal{L}_{X^{\epsilon}_{t(\delta)}}\hat{Y}^{\epsilon}_{s}\right)-\bar{b}\left(t(\delta),X^{\epsilon}_{t(\delta)},\mathcal{L}_{X^{\epsilon}_{t(\delta)}}\right)\right]ds\right|^{2}$$

$$=2\sum_{k=0}^{[t/\delta]-1}\left|\int_{k\delta}^{(k+1)\delta}\left[b\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}},\hat{Y}^{\epsilon}_{s}\right)-\bar{b}\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}}\right)\right]ds\right|^{2}$$

$$+4\sum_{0\le i<j\le[t/\delta]-1}\left\langle\int_{i\delta}^{(i+1)\delta}\left[b\left(i\delta,X^{\epsilon}_{i\delta},\mathcal{L}_{X^{\epsilon}_{i\delta}},\hat{Y}^{\epsilon}_{s}\right)-\bar{b}\left(i\delta,X^{\epsilon}_{i\delta},\mathcal{L}_{X^{\epsilon}_{i\delta}}\right)\right]ds,\right.$$

$$\left.\int_{j\delta}^{(j+1)\delta}\left[b\left(j\delta,X^{\epsilon}_{j\delta},\mathcal{L}_{X^{\epsilon}_{j\delta}},\hat{Y}^{\epsilon}_{s}\right)-\bar{b}\left(j\delta,X^{\epsilon}_{j\delta},\mathcal{L}_{X^{\epsilon}_{j\delta}}\right)\right]ds\right\rangle$$

$$+2\left|\int_{t(\delta)}^{t}\left[b\left(t(\delta),X^{\epsilon}_{t(\delta)},\mathcal{L}_{X^{\epsilon}_{t(\delta)}},\hat{Y}^{\epsilon}_{s}\right)-\bar{b}\left(t(\delta),X^{\epsilon}_{t(\delta)},\mathcal{L}_{X^{\epsilon}_{t(\delta)}}\right)\right]ds\right|^{2}$$

$$:=\sum_{i=1}^{3}I_{1i}(t). \tag{3.13}$$

For $I_{13}(t)$, by estimate (A.3) below, Lemmas 3.1 and 3.3, it is easy to prove that

$$\sup_{t\in[0,T]}\mathbb{E}I_{13}(t)\le C_T\delta\mathbb{E}\int_{t(\delta)}^{t}\left[1+\left|X^{\epsilon}_{t(\delta)}\right|^{2}+\mathbb{E}\left|X^{\epsilon}_{t(\delta)}\right|^{2}+\left|\hat{Y}^{\epsilon}_{s}\right|^{2}\right]ds$$
$$\le C_T\left(1+|x|^2+|y|^2\right)\delta^2. \tag{3.14}$$

For the term $I_{11}(t)$, we have

$$\mathbb{E}I_{11}(t)=2\sum_{k=0}^{[t/\delta]-1}\mathbb{E}\left|\int_{k\delta}^{(k+1)\delta}\left[b\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}},\hat{Y}^{\epsilon}_{s}\right)-\bar{b}\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}}\right)\right]ds\right|^{2}$$
$$=2\sum_{k=0}^{[t/\delta]-1}\mathbb{E}\left|\int_{0}^{\delta/\epsilon}\left[b\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}},\hat{Y}^{\epsilon}_{s\epsilon+k\delta}\right)-\bar{b}\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}}\right)\right]ds\right|^{2}$$
$$=2\epsilon^2\sum_{k=0}^{[t/\delta]-1}\int_{0}^{\delta/\epsilon}\int_{r}^{\delta/\epsilon}\Psi_k(s,r)\,ds\,dr,$$

where for any $0\le r\le s\le\delta/\epsilon$,

$$\Psi_k(s,r):=\mathbb{E}\left[\left\langle b\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}},\hat{Y}^{\epsilon}_{s\epsilon+k\delta}\right)-\bar{b}\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}}\right),\right.\right.$$
$$\left.\left.b\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}},\hat{Y}^{\epsilon}_{r\epsilon+k\delta}\right)-\bar{b}\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}}\right)\right\rangle\right].$$

For any $s>0$, $\mu\in\mathscr{P}_2$ and random variables $x,y\in\mathscr{F}_s$, we consider the following equation

$$\tilde{Y}^{\epsilon,s,x,\mu,y}_{t}=y+\frac{1}{\epsilon}\int_{s}^{t}f\left(s,x,\mu,\tilde{Y}^{\epsilon,s,x,\mu,y}_{r}\right)dr+\frac{1}{\sqrt{\epsilon}}\int_{s}^{t}g\left(s,x,\mu,\tilde{Y}^{\epsilon,s,x,\mu,y}_{r}\right)dW^2_r,\quad t\ge s.$$

Then by the construction of $\hat{Y}^{\epsilon}_{t}$, for any $k\in\mathbb{N}_*$, we have

$$\hat{Y}^{\epsilon}_{t}=\tilde{Y}^{\epsilon,k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}},\hat{Y}^{\epsilon}_{k\delta}}_{t},\quad t\in\left[k\delta,(k+1)\delta\right],$$

which implies

$$\Psi_k(s,r)=\mathbb{E}\left[\left\langle b\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}},\tilde{Y}^{\epsilon,k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}},\hat{Y}^{\epsilon}_{k\delta}}_{s\epsilon+k\delta}\right)-\bar{b}\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}}\right),\right.\right.$$
$$\left.\left.b\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}},\tilde{Y}^{\epsilon,k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}},\hat{Y}^{\epsilon}_{k\delta}}_{r\epsilon+k\delta}\right)-\bar{b}\left(k\delta,X^{\epsilon}_{k\delta},\mathcal{L}_{X^{\epsilon}_{k\delta}}\right)\right\rangle\right].$$

Note that since for any fixed $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $\tilde{Y}_{s\epsilon+k\delta}^{\epsilon,k\delta,x,\mu,y}$ is independent of $\mathcal{F}_{k\delta}$, and $X_{k\delta}^\epsilon$, $\hat{Y}_{k\delta}^\epsilon$ are $\mathcal{F}_{k\delta}$-measurable, we have

$$\begin{aligned}\Psi_k(s,r) &= \mathbb{E}\big\{\mathbb{E}\big[\big\langle b\big(k\delta, X_{k\delta}^\epsilon, \mathcal{L}_{X_{k\delta}^\epsilon}, \tilde{Y}_{s\epsilon+k\delta}^{\epsilon,k\delta,X_{k\delta}^\epsilon,\mathcal{L}_{X_{k\delta}^\epsilon},\hat{Y}_{k\delta}^\epsilon}\big) - \bar{b}\big(k\delta, X_{k\delta}^\epsilon, \mathcal{L}_{X_{k\delta}^\epsilon}\big),\\ &\qquad b\big(k\delta, X_{k\delta}^\epsilon, \mathcal{L}_{X_{k\delta}^\epsilon}, \tilde{Y}_{r\epsilon+k\delta}^{\epsilon,k\delta,X_{k\delta}^\epsilon,\mathcal{L}_{X_{k\delta}^\epsilon},\hat{Y}_{k\delta}^\epsilon}\big) - \bar{b}\big(k\delta, X_{k\delta}^\epsilon, \mathcal{L}_{X_{k\delta}^\epsilon}\big)\big\rangle \mid \mathcal{F}_{k\delta}\big](\omega)\big\}\\ &= \mathbb{E}\big\{\mathbb{E}\big[\big\langle b\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}, \tilde{Y}_{s\epsilon+k\delta}^{\epsilon,k\delta,X_{k\delta}^\epsilon(\omega),\mathcal{L}_{X_{k\delta}^\epsilon},\hat{Y}_{k\delta}^\epsilon(\omega)}\big) - \bar{b}\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}\big),\\ &\qquad b\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}, \tilde{Y}_{r\epsilon+k\delta}^{\epsilon,k\delta,X_{k\delta}^\epsilon(\omega),\mathcal{L}_{X_{k\delta}^\epsilon},\hat{Y}_{k\delta}^\epsilon(\omega)}\big) - \bar{b}\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}\big)\big\rangle\big]\big\}.\end{aligned}$$

By the definition of the process $\{\tilde{Y}_t^{\epsilon,s,x,\mu,y}\}_{t\geq 0}$, it is easy to see that

$$\begin{aligned}\tilde{Y}_{s\epsilon+k\delta}^{\epsilon,k\delta,x,\mu,y} &= y + \frac{1}{\epsilon}\int_{k\delta}^{s\epsilon+k\delta} f\big(k\delta, x, \mu, \tilde{Y}_r^{\epsilon,k\delta,x,\mu,y}\big)\,dr + \frac{1}{\sqrt{\epsilon}}\int_{k\delta}^{s\epsilon+k\delta} g\big(k\delta, x, \mu, \tilde{Y}_r^{\epsilon,k\delta,x,\mu,y}\big)\,dW_r^2\\ &= y + \frac{1}{\epsilon}\int_0^{s\epsilon} f\big(k\delta, x, \mu, \tilde{Y}_{r+k\delta}^{\epsilon,k\delta,x,\mu,y}\big)\,dr + \frac{1}{\sqrt{\epsilon}}\int_0^{s\epsilon} g\big(k\delta, x, \mu, \tilde{Y}_{r+k\delta}^{\epsilon,k\delta,x,\mu,y}\big)\,dW_r^{2,k\delta}\\ &= y + \int_0^s f\big(k\delta, x, \mu, \tilde{Y}_{r\epsilon+k\delta}^{\epsilon,k\delta,x,\mu,y}\big)\,dr + \int_0^s g\big(k\delta, x, \mu, \tilde{Y}_{r\epsilon+k\delta}^{\epsilon,k\delta,x,\mu,y}\big)\,d\hat{W}_r^{2,k\delta},\end{aligned} \tag{3.15}$$

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

$$Y_s^{k\delta,x,\mu,y} = y + \int_0^s f\big(k\delta, x, \mu, Y_r^{k\delta,x,\mu,y}\big)\,dr + \int_0^s g\big(k\delta, x, \mu, Y_r^{k\delta,x,\mu,y}\big)\,d\tilde{W}_r^2. \tag{3.16}$$

The uniqueness of the solutions of Eq. (3.15) and Eq. (3.16) implies that the distribution of $\{\tilde{Y}_{s\epsilon+k\delta}^{\epsilon,k\delta,x,\mu,y}\}_{0\leq s\leq\delta/\epsilon}$ coincides with the distribution of $\{Y_s^{k\delta,x,\mu,y}\}_{0\leq s\leq\delta/\epsilon}$. Then by Proposition 3.7, we have

$$\begin{aligned}\Psi_k(s,r) &= \mathbb{E}\big[\tilde{\mathbb{E}}\big\langle b\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}, Y_s^{k\delta,X_{k\delta}^\epsilon(\omega),\mathcal{L}_{X_{k\delta}^\epsilon},\hat{Y}_{k\delta}^\epsilon(\omega)}\big) - \bar{b}\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}\big),\\ &\qquad b\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}, Y_r^{k\delta,X_{k\delta}^\epsilon(\omega),\mathcal{L}_{X_{k\delta}^\epsilon},\hat{Y}_{k\delta}^\epsilon(\omega)}\big) - \bar{b}\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}\big)\big\rangle\big]\\ &= \mathbb{E}\big[\tilde{\mathbb{E}}\big\langle\tilde{\mathbb{E}}\big[b\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}, Y_s^{k\delta,X_{k\delta}^\epsilon(\omega),\mathcal{L}_{X_{k\delta}^\epsilon},\hat{Y}_{k\delta}^\epsilon(\omega)}\big) \mid \tilde{\mathcal{F}}_r\big](\tilde{\omega}) - \bar{b}\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}\big),\\ &\qquad b\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}, Y_r^{k\delta,X_{k\delta}^\epsilon(\omega),\mathcal{L}_{X_{k\delta}^\epsilon},\hat{Y}_{k\delta}^\epsilon(\omega)}(\tilde{\omega})\big) - \bar{b}\big(k\delta, X_{k\delta}^\epsilon(\omega), \mathcal{L}_{X_{k\delta}^\epsilon}\big)\big\rangle\big]\\ &\leq C_T\mathbb{E}\big\{\tilde{\mathbb{E}}\big[1 + \big|X_{k\delta}^\epsilon(\omega)\big|^2 + \big|Y_r^{k\delta,X_{k\delta}^\epsilon(\omega),\mathcal{L}_{X_{k\delta}^\epsilon},\hat{Y}_{k\delta}^\epsilon(\omega)}(\tilde{\omega})\big|^2 + \mathcal{L}_{X_{k\delta}^\epsilon}\big(|\cdot|^2\big)\big]e^{-[(s-r)\beta]/2}\big\}\\ &\leq C_T\mathbb{E}\big(1 + \big|X_{k\delta}^\epsilon\big|^2 + \big|\hat{Y}_{k\delta}^\epsilon\big|^2 + \mathbb{E}\big|X_{k\delta}^\epsilon\big|^2\big)e^{-[(s-r)\beta]/2}\\ &\leq C_T\big(1 + |x|^2 + |y|^2\big)e^{-[(s-r)\beta]2},\end{aligned}$$

where the last inequality is consequence of Lemmas 3.1 and 3.3. Hence we have

$$\begin{aligned}\sup_{t\in[0,T]}\mathbb{E}I_{11}(t) &\leq C_T\big(1 + |x|^2 + |y|^2\big)\frac{\epsilon^2}{\delta}\int_0^{\delta/\epsilon}\int_r^{\frac{\delta}{\epsilon}} e^{-[(s-r)\beta]/2}\,ds\,dr\\ &= C_T\big(1 + |x|^2 + |y|^2\big)\frac{\epsilon^2}{\delta}\bigg(\frac{\delta}{\beta\epsilon} - \frac{1}{\beta^2} + \frac{1}{\beta^2}e^{-\frac{\beta\delta}{\epsilon}}\bigg)\\ &\leq C_T\big(1 + |x|^2 + |y|^2\big)\bigg(\epsilon + \frac{\epsilon^2}{\delta}\bigg).\end{aligned} \tag{3.17}$$

For the term $I_{12}(t)$, in Step 3 we will prove the following estimate:

$$\sup_{t\in[0,T]} \mathbb{E}I_{12}(t) \le C_T\big(1+|x|^3+|y|^3\big)\bigg(\frac{\epsilon}{\delta^{1/2}}+\epsilon\bigg). \tag{3.18}$$

As a consequence, estimates (3.13), (3.14), (3.17) and (3.18) imply (3.12).

Step 3. In this step, we intend to prove estimate (3.18). For convenience, for any $i\in\mathbb{N}$, setting $Z^\epsilon_{i,t} := \tilde{Y}_t^{\epsilon,i\delta,X^\epsilon_{i\delta},\mathcal{L}_{X^\epsilon_{i\delta}},\hat{Y}^\epsilon_{i\delta}}$ with $i\delta\le t$, we obtain that

$$\begin{cases} dZ^\epsilon_{i,t} = \frac{1}{\epsilon}f(i\delta, X^\epsilon_{i\delta}, \mathcal{L}_{X^\epsilon_{i\delta}} Z^\epsilon_{i,t})\,dt + \frac{1}{\sqrt{\epsilon}}g(i\delta, X^\epsilon_{i\delta}, \mathcal{L}_{X^\epsilon_{i\delta}}, Z^\epsilon_{i,t})\,dW^2_t, \\ Z^\epsilon_{i,i\delta} = \hat{Y}^\epsilon_{i\delta}. \end{cases} \tag{3.19}$$

By the definition above, it is easy to see that

$$Z^\epsilon_{k,t} = \hat{Y}^\epsilon_t, \quad t\in\big[k\delta,(k+1)\delta\big]$$

and continuity implies that

$$Z^\epsilon_{k,(k+1)\delta} = Z^\epsilon_{k+1,(k+1)\delta} = \hat{Y}^\epsilon_{(k+1)\delta}.$$

Let $\mathbb{E}_s$ be the conditional expectation *w.r.t.* $\mathcal{F}_s$, $s\ge 0$. Then for any $0\le i<j\le[t/\delta]-1$,

$$\begin{aligned}
&\mathbb{E}\bigg\langle \int_{i\delta}^{(i+1)\delta}\big[b\big(i\delta, X^\epsilon_{i\delta}, \mathcal{L}_{X^\epsilon_{i\delta}}, \hat{Y}^\epsilon_s\big) - \bar{b}\big(i\delta, X^\epsilon_{i\delta}, \mathcal{L}_{X^\epsilon_{i\delta}}\big)\big]\,ds, \\
&\qquad \int_{j\delta}^{(j+1)\delta}\big[b\big(j\delta, X^\epsilon_{j\delta}, \mathcal{L}_{X^\epsilon_{j\delta}}, \hat{Y}^\epsilon_s\big) - \bar{b}\big(j\delta, X^\epsilon_{j\delta}, \mathcal{L}_{X^\epsilon_{j\delta}}\big)\big]\,ds\bigg\rangle \\
&\quad = \int_{i\delta}^{(i+1)\delta}\int_{j\delta}^{(j+1)\delta} \mathbb{E}\big\langle b\big(i\delta, X^\epsilon_{i\delta}, \mathcal{L}_{X^\epsilon_{i\delta}}, \hat{Y}^\epsilon_s\big) - \bar{b}\big(i\delta, X^\epsilon_{i\delta}, \mathcal{L}_{X^\epsilon_{i\delta}}\big), \\
&\qquad b\big(j\delta, X^\epsilon_{j\delta}, \mathcal{L}_{X^\epsilon_{j\delta}}, \hat{Y}^\epsilon_t\big) - \bar{b}\big(j\delta, X^\epsilon_{j\delta}, \mathcal{L}_{X^\epsilon_{j\delta}}\big)\big\rangle\,ds\,dt \\
&\quad \le \int_{i\delta}^{(i+1)\delta}\int_{j\delta}^{(j+1)\delta} \mathbb{E}\big\{\big|b\big(i\delta, X^\epsilon_{i\delta}, \mathcal{L}_{X^\epsilon_{i\delta}}, \hat{Y}^\epsilon_s\big) - \bar{b}\big(i\delta, X^\epsilon_{i\delta}, \mathcal{L}_{X^\epsilon_{i\delta}}\big)\big| \\
&\qquad \cdot \big|\mathbb{E}_{(i+1)\delta}\big[b\big(j\delta, X^\epsilon_{j\delta}, \mathcal{L}_{X^\epsilon_{j\delta}} \hat{Y}^\epsilon_t\big) - \bar{b}\big(j\delta, X^\epsilon_{j\delta}, \mathcal{L}_{X^\epsilon_{j\delta}}\big)\big]\big|\big\}\,ds\,dt \\
&\quad \le C_T\int_{i\delta}^{(i+1)\delta}\int_{j\delta}^{(j+1)\delta} \mathbb{E}\big\{\big(1+\big|X^\epsilon_{i\delta}\big|+\big|\hat{Y}^\epsilon_s\big|\big)\big|\mathbb{E}_{(i+1)\delta}\big[\big(b\big(j\delta, X^\epsilon_{j\delta}, \mathcal{L}_{X^\epsilon_{j\delta}}, \hat{Y}^\epsilon_t\big) - \bar{b}\big(j\delta, X^\epsilon_{j\delta}, \mathcal{L}_{X^\epsilon_{j\delta}}\big)\big) \\
&\qquad - \big(b\big((i+1)\delta, X^\epsilon_{(i+1)\delta}, \mathcal{L}_{X^\epsilon_{(i+1)\delta}}, Z^\epsilon_{i+1,t}\big) - \bar{b}\big((i+1)\delta, X^\epsilon_{(i+1)\delta}, \mathcal{L}_{X^\epsilon_{(i+1)\delta}}\big)\big)\big]\big|\big\}\,ds\,dt \\
&\qquad + C_T\int_{i\delta}^{(i+1)\delta}\int_{j\delta}^{(j+1)\delta} \mathbb{E}\big\{\big(1+\big|X^\epsilon_{i\delta}\big|+\big|\hat{Y}^\epsilon_s\big|\big) \\
&\qquad \cdot \big|\mathbb{E}_{(i+1)\delta}\big[b\big((i+1)\delta, X^\epsilon_{(i+1)\delta}, \mathcal{L}_{X^\epsilon_{(i+1)\delta}}, Z^\epsilon_{i+1,t}\big) - \bar{b}\big((i+1)\delta, X^\epsilon_{(i+1)\delta}, \mathcal{L}_{X^\epsilon_{(i+1)\delta}}\big)\big]\big|\big\}\,ds\,dt \\
&\quad := B_1 + B_2.
\end{aligned} \tag{3.20}$$

On one hand, by a similar argument for $I_{11}(t)$, we obtain

$$\begin{aligned}
B_2 &\le C_T\int_{i\delta}^{(i+1)\delta}\int_{j\delta}^{(j+1)\delta} \mathbb{E}\big[\big(1+\big|X^\epsilon_{i\delta}\big|+\big|\hat{Y}^\epsilon_s\big|\big)\big(1+\big|X^\epsilon_{(i+1)\delta}\big|+\big|\hat{Y}^\epsilon_{(i+1)\delta}\big|\big)\big]e^{\frac{-\beta[t-(i+1)\delta]}{2\epsilon}}\,ds\,dt \\
&\le C_T\big(1+|x|^2+|y|^2\big)\int_{i\delta}^{(i+1)\delta}\int_{j\delta}^{(j+1)\delta} e^{\frac{-\beta[t-(i+1)\delta]}{2\epsilon}}\,ds\,dt \\
&\le C_T\big(1+|x|^2+|y|^2\big)\epsilon\delta e^{\frac{-\beta(j-i)\delta}{2\epsilon}}\big(1-e^{\frac{-\beta\delta}{2\epsilon}}\big).
\end{aligned} \tag{3.21}$$

On the other hand,

$$
\begin{aligned}
B_1 = {} & C_T \int_{i\delta}^{(i+1)\delta} \int_{j\delta}^{(j+1)\delta} \sum_{k=i+1}^{j-1} \mathbb{E}\big\{\big(1+\big|X_{i\delta}^{\epsilon}\big|+\big|\hat{Y}_s^{\epsilon}\big|\big)\big|\mathbb{E}_{(i+1)\delta}\big[\big(b\big((k+1)\delta, X_{(k+1)\delta}^{\epsilon}, \mathcal{L}_{X_{(k+1)\delta}^{\epsilon}}, Z_{k+1,t}^{\epsilon}\big) \\
& - \bar{b}\big((k+1)\delta, X_{(k+1)\delta}^{\epsilon}, \mathcal{L}_{X_{(k+1)\delta}^{\epsilon}}\big)\big) - \big(b\big(k\delta, X_{k\delta}^{\epsilon}, \mathcal{L}_{X_{k\delta}^{\epsilon}}, Z_{k,t}^{\epsilon}\big) - \bar{b}\big(k\delta, X_{k\delta}^{\epsilon}, \mathcal{L}_{X_{k\delta}^{\epsilon}}\big)\big)\big]\big|\big\}\, ds\, dt \\
= {} & C_T \int_{i\delta}^{(i+1)\delta} \int_{j\delta}^{(j+1)\delta} \sum_{k=i+1}^{j-1} \mathbb{E}\big\{\big(1+\big|X_{i\delta}^{\epsilon}\big|+\big|\hat{Y}_s^{\epsilon}\big|\big)\big|\mathbb{E}_{k\delta}\big[\big(b\big((k+1)\delta, X_{(k+1)\delta}^{\epsilon}, \mathcal{L}_{X_{(k+1)\delta}^{\epsilon}}, Z_{k+1,t}^{\epsilon}\big) \\
& - \bar{b}\big((k+1)\delta, X_{(k+1)\delta}^{\epsilon}, \mathcal{L}_{X_{(k+1)\delta}^{\epsilon}}\big)\big) - \big(b\big(k\delta, X_{k\delta}^{\epsilon}, \mathcal{L}_{X_{k\delta}^{\epsilon}}, Z_{k,t}^{\epsilon}\big) - \bar{b}\big(k\delta, X_{k\delta}^{\epsilon}, \mathcal{L}_{X_{k\delta}^{\epsilon}}\big)\big)\big]\big|\big\}\, ds\, dt.
\end{aligned}
$$

Thanks to the Markov property, we get

$$
\begin{aligned}
& \mathbb{E}_{k\delta}\big[b\big((k+1)\delta, X_{(k+1)\delta}^{\epsilon}, \mathcal{L}_{X_{(k+1)\delta}^{\epsilon}}, Z_{k+1,t}^{\epsilon}\big) - \bar{b}\big((k+1)\delta, X_{(k+1)\delta}^{\epsilon}, \mathcal{L}_{X_{(k+1)\delta}^{\epsilon}}\big)\big] \\
& \quad = \mathbb{E}_{k\delta}\big[\tilde{b}\big((k+1)\delta, X_{(k+1)\delta}^{\epsilon}, \mathcal{L}_{X_{(k+1)\delta}^{\epsilon}}, \hat{Y}_{(k+1)\delta}^{\epsilon}, \big[t-(k+1)\delta\big]/\epsilon\big)\big]
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathbb{E}_{k\delta}\big[b\big(k\delta, X_{k\delta}^{\epsilon}, \mathcal{L}_{X_{k\delta}^{\epsilon}}, Z_{k,t}^{\epsilon}\big) - \bar{b}\big(k\delta, X_{k\delta}^{\epsilon}, \mathcal{L}_{X_{k\delta}^{\epsilon}}\big)\big] \\
& \quad = \mathbb{E}_{k\delta}\big[\tilde{b}\big(k\delta, X_{k\delta}^{\epsilon}, \mathcal{L}_{X_{k\delta}^{\epsilon}}, \hat{Y}_{(k+1)\delta}^{\epsilon}, \big[t-(k+1)\delta\big]/\epsilon\big)\big],
\end{aligned}
$$

where $\tilde{b}(t,x,\mu,y,s) := \tilde{\mathbb{E}} b(t,x,\mu,Y_s^{t,x,\mu,y}) - \bar{b}(t,x,\mu)$.

Recall the following properties of $\tilde{b}$ (see the detailed proof in Section A.3):

- For any $t_1, t_2 \in [0,T]$, $s \geq 0$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $\mu \in \mathscr{P}_2$,

$$
\begin{aligned}
& \big|\tilde{b}(t_1,x,\mu,y,s) - \tilde{b}(t_2,x,\mu,y,s)\big| \\
& \quad \leq C_T |t_1-t_2| e^{-\eta s}\big\{1+|x|^{\gamma_1}+|y|^{\gamma_1}+\big[\mu\big(|\cdot|^2\big)\big]^{\gamma_1/2}\big\};
\end{aligned}
\tag{3.22}
$$

- For any $t \in [0,T]$, $s \geq 0$, $x_1, x_2 \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $\mu \in \mathscr{P}_2$,

$$
\begin{aligned}
& \big|\tilde{b}(t,x_1,\mu,y,s) - \tilde{b}(t,x_2,\mu,y,s)\big| \\
& \quad \leq C_T |x_1-x_2| e^{-\eta s}\big\{1+|x_1|^{\gamma_1}+|x_2|^{\gamma_1}+|y|^{\gamma_1}+\big[\mu\big(|\cdot|^2\big)\big]^{\gamma_1/2}\big\};
\end{aligned}
\tag{3.23}
$$

- For any $t \in [0,T]$, $s \geq 0$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $\mu_1, \mu_2 \in \mathscr{P}_2$,

$$
\begin{aligned}
& \big|\tilde{b}(t,x,\mu_1,y,s) - \tilde{b}(t,x,\mu_2,y,s)\big| \\
& \quad \leq C_T \mathbb{W}_2(\mu_1,\mu_2) e^{-\eta s}\big\{1+|x|^{\gamma_1}+|y|^{\gamma_1}+\big[\mu_1\big(|\cdot|^2\big)\big]^{\gamma_1/2}+\big[\mu_2\big(|\cdot|^2\big)\big]^{\gamma_1/2}\big\},
\end{aligned}
\tag{3.24}
$$

where η is a positive constant. Then by estimates (3.22)–(3.24) and Lemma 3.1, we have

$$
\begin{aligned}
B_1 \leq {} & C \int_{i\delta}^{(i+1)\delta} \int_{j\delta}^{(j+1)\delta} \sum_{k=i+1}^{j-1} \mathbb{E}\big\{\big(1+\big|X_{i\delta}^{\epsilon}\big|+\big|\hat{Y}_s^{\epsilon}\big|\big) \\
& \times \big|\tilde{b}\big((k+1)\delta, X_{(k+1)\delta}^{\epsilon}, \mathcal{L}_{X_{(k+1)\delta}^{\epsilon}}, \hat{Y}_{(k+1)\delta}^{\epsilon}, \big[t-(k+1)\delta\big]/\epsilon\big) \\
& - \tilde{b}\big(k\delta, X_{k\delta}^{\epsilon}, \mathcal{L}_{X_{k\delta}^{\epsilon}}, \hat{Y}_{(k+1)\delta}^{\epsilon}, \big[t-(k+1)\delta\big]/\epsilon\big)\big|\big\}\, ds\, dt \\
\leq {} & C_T \int_{i\delta}^{(i+1)\delta} \int_{j\delta}^{(j+1)\delta} \sum_{k=i+1}^{j-1} \mathbb{E}\big[\big(1+\big|X_{i\delta}^{\epsilon}\big|+\big|\hat{Y}_s^{\epsilon}\big|\big)\big(1+\big|X_{k\delta}^{\epsilon}\big|^{\gamma_1}+\big|X_{(k+1)\delta}^{\epsilon}\big|^{\gamma_1}+\big|\hat{Y}_{(k+1)\delta}^{\epsilon}\big|^{\gamma_1}\big) \\
& \times \big(\delta+\big|X_{(k+1)\delta}^{\epsilon}-X_{k\delta}^{\epsilon}\big|+\big[\mathbb{E}\big|X_{(k+1)\delta}^{\epsilon}-X_{k\delta}^{\epsilon}\big|^2\big]^{1/2}\big)\big] e^{\frac{-\beta[t-(k+1)\delta]}{4\epsilon}}\, ds\, dt
\end{aligned}
$$

$$\leq C_T\big(1+|x|^3+|y|^3\big)\delta^{1/2}\int_{i\delta}^{(i+1)\delta}\int_{j\delta}^{(j+1)\delta}\sum_{k=i+1}^{j-1}e^{\frac{-\beta[t-(k+1)\delta]}{4\epsilon}}\,ds\,dt$$

$$\leq C_T\big(1+|x|^3+|y|^3\big)\delta^{1/2}\int_{i\delta}^{(i+1)\delta}\int_{j\delta}^{(j+1)\delta}\frac{e^{\frac{-\beta(t-j\delta)}{4\epsilon}}}{1-e^{\frac{-\beta\delta}{4\epsilon}}}\,ds\,dt$$

$$\leq C_T\big(1+|x|^3+|y|^3\big)\delta^{3/2}\epsilon. \tag{3.25}$$

Combining estimates (3.20), (3.21) and (3.25), we obtain

$$\sup_{t\in[0,T]}\mathbb{E}I_{12}(t)\leq C_T\big(1+|x|^3+|y|^3\big)\sum_{0\leq i<j\leq[T/\delta]-1}\Big[\delta^{3/2}\epsilon+\epsilon\delta e^{\frac{-\beta(j-i)\delta}{4\epsilon}}\big(1-e^{\frac{-\beta\delta}{4\epsilon}}\big)\Big]$$

$$\leq C_T\big(1+|x|^3+|y|^3\big)\bigg(\frac{\epsilon}{\delta^{1/2}}+\epsilon\bigg),$$

which is the estimate (3.18). The proof is complete. □

Now we are in a position to complete our first result.

Proof of Theorem 2.3. Taking $\delta=\epsilon^{2/3}$, Lemmas 3.5 and 3.9 imply that for any $T>0$, initial values $x\in\mathbb{R}^n$ and $y\in\mathbb{R}^m$, there exists $C_T>0$ such that

$$\sup_{t\in[0,T]}\mathbb{E}\big|X_t^\epsilon-\bar{X}_t\big|^2\leq C_T\big(1+|x|^3+|y|^3\big)\epsilon^{2/3},$$

which proves the first part of Theorem 2.3, i.e., (2.16) holds.

Furthermore, if there is no noise in the slow equation (i.e., $\sigma=0$), we can improve the Hölder continuity in time in Lemma 3.2, i.e., for any $T>0$, $0\leq t\leq t+h\leq T$, there exists a positive constant C_T such that

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

Then, following almost the same procedure as above, it is easy to see that

$$\sup_{t\in[0,T]}\mathbb{E}\big|X_t^\epsilon-\bar{X}_t\big|^2\leq C_T\big(1+|x|^3+|y|^3\big)\bigg(\epsilon+\frac{\epsilon^2}{\delta}+\delta^2\bigg).$$

Hence, taking $\delta=\epsilon$ yields (2.17). The proof is complete. □

4. Proof of Theorem 2.5

In this section, we will use the technique of Poisson equation to prove the strong convergence order, which is quite different from the method used in Section 3. Because we will study the regularity of second-order derivatives of the solution for the corresponding Poisson equation, more conditions (see assumption A3) are needed. This section is divided into two subsections. In Section 4.1, we study the regularity of the solution for the corresponding Poisson equation. In Section 4.2, we prove Theorem 2.5 by using the technique of Poisson equation. Note that we always assume conditions A1–A3 hold.

4.1. *Poisson equation*

Consider the following Poisson equation:

$$-\mathscr{L}_2(t,x,\mu)\Phi(t,x,\mu,y)=b(t,x,\mu,y)-\bar{b}(t,x,\mu), \tag{4.1}$$

where $\Phi(t,x,\mu,y)=(\Phi_1(t,x,\mu,y),\dots,\Phi_n(t,x,\mu,y))$,

$$\mathscr{L}_2(t,x,\mu)\Phi(t,x,\mu,y):=\big(\mathscr{L}_2(t,x,\mu)\Phi_1(t,x,\mu,y),\dots,\mathscr{L}_2(t,x,\mu)\Phi_n(t,x,\mu,y)\big)$$

and for any $k=1,\dots,n$,

$$\mathcal{L}_2(t,x,\mu)\Phi_k(t,x,\mu,y) := \langle f(t,x,\mu,y), \partial_y\Phi_k(t,x,\mu,y)\rangle + \frac{1}{2}\mathrm{Tr}\big[gg^*(t,x,\mu,y)\partial^2_{yy}\Phi_k(t,x,\mu,y)\big].$$

The smoothness of the solution of the Poisson equation with respect to parameters have been studied in many references, see [31,32,34] for example. Note that here the solution for the Poisson equation (4.1) depends on the parameter μ, so here we have to check the regularity *w.r.t.* μ. The main result of this subsection is the following:

Proposition 4.1. *Assume the assumptions* A1–A3 *hold. Define*

$$\Phi(t,x,\mu,y) := \int_0^\infty \tilde{\mathbb{E}}\big[b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)\big] - \bar{b}(t,x,\mu)\,ds. \tag{4.2}$$

Then $\Phi(t,x,\mu,y)$ *is the unique solution of Eq.* (4.1) *and it satisfies that* $\Phi(\cdot,\cdot,\mu,\cdot)\in C^{1,2,2}([0,\infty)\times\mathbb{R}^n\times\mathbb{R}^m,\mathbb{R}^n)$, $\Phi(t,x,\cdot,y)\in C^{1,1}(\mathscr{P}_2,\mathbb{R}^n)$. *Moreover, for any* $t\in[0,T]$,

$$\max\big\{\big|\Phi(t,x,\mu,y)\big|, \big\|\partial_y\Phi(t,x,\mu,y)\big\|, \big|\partial_t\Phi(t,x,\mu,y)\big|, \big\|\partial_x\Phi(t,x,\mu,y)\big\|, \big\|\partial_\mu\Phi(t,x,\mu,y)\big\|_{L^2(\mu)}\big\} \le C_T\big\{1+|x|+|y|+\big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\} \tag{4.3}$$

and

$$\max\big\{\big\|\partial^2_{xx}\Phi(t,x,\mu,y)\big\|, \big\|\partial_z\partial_\mu\Phi(t,x,\mu,y)(\cdot)\big\|_{L^2(\mu)}\big\} \le C_T\big\{1+|x|+|y|+\big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\}. \tag{4.4}$$

Proof. We will divide the proof into three steps.

Step 1. Noting that $\mathcal{L}_2(t,x,\mu)$ is the infinitesimal generator of the frozen process $\{Y_s^{t,x,\mu}\}$, we easily check that (4.2) is the unique solution of the Poisson equation (4.1) under the assumptions A1–A3. Moreover, by a straightforward computation, we also have that $\Phi(\cdot,\cdot,\mu,\cdot)\in C^{1,2,2}([0,\infty)\times\mathbb{R}^n\times\mathbb{R}^m,\mathbb{R}^n)$, $\Phi(t,x,\cdot,y)\in C^{1,1}(\mathscr{P}_2,\mathbb{R}^n)$.

By Proposition 3.7, we get

$$\begin{aligned}\big|\Phi(t,x,\mu,y)\big| &\le \int_0^\infty \big|\tilde{\mathbb{E}}\big[b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)\big] - \bar{b}(t,x,\mu)\big|\,ds\\ &\le C_T\big\{1+|x|+|y|+\big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\}\int_0^\infty e^{-\frac{\beta s}{2}}\,ds\\ &\le C_T\big\{1+|x|+|y|+\big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\}.\end{aligned}$$

By Lemma 3.6, we have $\tilde{\mathbb{E}}\|\partial_y Y_s^{t,x,\mu,y}\|^2 \le C_T e^{-\beta s}$, which implies

$$\big\|\partial_y\Phi(t,x,\mu,y)\big\| \le C_T.$$

Furthermore, the remaining estimates in (4.3) can be obtained easily by (3.22)–(3.24). Therefore, it is sufficient to estimate (4.4) below.

We first recall that (see Section A.3 in the Appendix)

$$\tilde{b}_{s_0}(t,x,\mu,y,s) = \hat{b}(t,x,\mu,y,s) - \hat{b}(t,x,\mu,y,s+s_0),$$

where $\hat{b}(t,x,\mu,y,s) = \tilde{\mathbb{E}}b(t,x,\mu,Y_s^{t,x,\mu,y})$. Note that

$$\lim_{s_0\to\infty}\tilde{b}_{s_0}(t,x,\mu,y,s) = \tilde{\mathbb{E}}\big[b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)\big] - \bar{b}(t,x,\mu).$$

So, in order to prove (4.4), it suffices to show there exists $\eta>0$ such that for any $s_0>0$, $t\in[0,T]$, $s\ge 0$, $x\in\mathbb{R}^n$, $y\in\mathbb{R}^m$ and $\mu\in\mathscr{P}_2$,

$$\big\|\partial^2_{xx}\tilde{b}_{s_0}(t,x,\mu,y,s)\big\| \le C_T e^{-\eta s}\big\{1+|x|+|y|+\big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\} \tag{4.5}$$

and

$$\left\|\partial_z\partial_\mu\Phi(t,x,\mu,y)(\cdot)\right\|_{L^2(\mu)}\le C_Te^{-\eta s}\left\{1+|x|+|y|+\left[\mu\left(|\cdot|^2\right)\right]^{1/2}\right\},\tag{4.6}$$

which will be proved in the following two steps.

Step 2. In this step, we intend to prove estimate (4.5). We recall that in (A.5) below

$$\tilde{b}_{s_0}(t,x,\mu,y,s)=\hat{b}(t,x,\mu,y,s)-\tilde{\mathbb{E}}\hat{b}\left(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\right).$$

Then the chain rule yields

$$\begin{aligned}\partial_x\tilde{b}_{s_0}(t,x,\mu,y,s)=\partial_x\hat{b}(t,x,\mu,y,s)&-\tilde{\mathbb{E}}\partial_x\hat{b}\left(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\right)\\&-\tilde{\mathbb{E}}\left[\partial_y\hat{b}\left(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\right)\cdot\partial_xY_{s_0}^{t,x,\mu,y}\right],\end{aligned}$$

and furthermore,

$$\begin{aligned}&\partial^2_{xx}\tilde{b}_{s_0}(t,x,\mu,y,s)\\&\quad=\partial^2_{xx}\hat{b}(t,x,\mu,y,s)-\tilde{\mathbb{E}}\partial^2_{xx}\hat{b}\left(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\right)\\&\qquad-\tilde{\mathbb{E}}\left[\partial^2_{xy}\hat{b}\left(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\right)\cdot\partial_xY_{s_0}^{t,x,\mu,y}\right]\\&\qquad-\tilde{\mathbb{E}}\left\{\left[\partial^2_{yx}\hat{b}\left(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\right)+\partial^2_{yy}\hat{b}\left(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\right)\cdot\partial_xY_{s_0}^{t,x,\mu,y}\right]\cdot\partial_xY_{s_0}^{t,x,\mu,y}\right\}\\&\qquad-\tilde{\mathbb{E}}\left[\partial_y\hat{b}\left(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\right)\cdot\partial^2_{xx}Y_{s_0}^{t,x,\mu,y}\right]\\&\quad:=\sum_{i=1}^4J_i.\end{aligned}$$

(i) For the term J_1, note that

$$\partial_x\hat{b}(t,x,\mu,y,s)=\tilde{\mathbb{E}}\left[\partial_xb\left(t,x,\mu,Y_s^{t,x,\mu,y}\right)\right]+\tilde{\mathbb{E}}\left[\partial_yb\left(t,x,\mu,Y_s^{t,x,\mu,y}\right)\cdot\partial_xY_s^{t,x,\mu,y}\right],$$

which implies

$$\begin{aligned}\partial^2_{xx}\hat{b}(t,x,\mu,y,s)=\tilde{\mathbb{E}}\left[\partial^2_{xx}b\left(t,x,\mu,Y_s^{t,x,\mu,y}\right)\right]&+\tilde{\mathbb{E}}\left[\partial^2_{xy}b\left(t,x,\mu,Y_s^{t,x,\mu,y}\right)\cdot\partial_xY_s^{t,x,\mu,y}\right]\\&+\tilde{\mathbb{E}}\left[\partial^2_{yx}b\left(t,x,\mu,Y_s^{t,x,\mu,y}\right)\cdot\partial_xY_s^{t,x,\mu,y}\right]\\&+\tilde{\mathbb{E}}\left[\partial^2_{yy}b\left(t,x,\mu,Y_s^{t,x,\mu,y}\right)\cdot\left(\partial_xY_s^{t,x,\mu,y},\partial_xY_s^{t,x,\mu,y}\right)\right]\\&+\tilde{\mathbb{E}}\left[\partial_yb\left(t,x,\mu,Y_s^{t,x,\mu,y}\right)\cdot\partial^2_{xx}Y_s^{t,x,\mu,y}\right].\end{aligned}$$

Then for any $y_1,y_2\in\mathbb{R}^m$,

$$\begin{aligned}&\left\|\partial^2_{xx}\hat{b}(t,x,\mu,y_1,s)-\partial^2_{xx}\hat{b}(t,x,\mu,y_2,s)\right\|\\&\quad\le\left\|\tilde{\mathbb{E}}\left[\partial^2_{xx}b\left(t,x,\mu,Y_s^{t,x,\mu,y_1}\right)-\partial^2_{xx}b\left(t,x,\mu,Y_s^{t,x,\mu,y_2}\right)\right]\right\|\\&\qquad+\left\|\tilde{\mathbb{E}}\left[\partial^2_{xy}b\left(t,x,\mu,Y_s^{t,x,\mu,y_1}\right)\cdot\partial_xY_s^{t,x,\mu,y_1}-\partial^2_{xy}b\left(t,x,\mu,Y_s^{t,x,\mu,y_2}\right)\cdot\partial_xY_s^{t,x,\mu,y_2}\right]\right\|\\&\qquad+\left\|\tilde{\mathbb{E}}\left[\partial^2_{yx}b\left(t,x,\mu,Y_s^{t,x,\mu,y_1}\right)\cdot\partial_xY_s^{t,x,\mu,y_1}-\partial^2_{yx}b\left(t,x,\mu,Y_s^{t,x,\mu,y_2}\right)\cdot\partial_xY_s^{t,x,\mu,y_2}\right]\right\|\\&\qquad+\left\|\tilde{\mathbb{E}}\left[\partial^2_{yy}b\left(t,x,\mu,Y_s^{t,x,\mu,y_1}\right)\cdot\left(\partial_xY_s^{t,x,\mu,y_1},\partial_xY_s^{t,x,\mu,y_1}\right)\right.\right.\\&\qquad\left.\left.-\partial^2_{yy}b\left(t,x,\mu,Y_s^{t,x,\mu,y_2}\right)\cdot\left(\partial_xY_s^{t,x,\mu,y_2},\partial_xY_s^{t,x,\mu,y_2}\right)\right]\right\|\\&\qquad+\left\|\tilde{\mathbb{E}}\left[\partial_yb\left(t,x,\mu,Y_s^{t,x,\mu,y_1}\right)\cdot\partial^2_{xx}Y_s^{t,x,\mu,y_1}-\partial_yb\left(t,x,\mu,Y_s^{t,x,\mu,y_2}\right)\cdot\partial^2_{xx}Y_s^{t,x,\mu,y_2}\right]\right\|\\&\quad:=\sum_{i=1}^5J_{1i}.\end{aligned}\tag{4.7}$$

By condition (2.8) and Lemma 3.6, there exists $\eta > 0$ such that

$$J_{11} \le C_T \tilde{\mathbb{E}} \big| Y_s^{t,x,\mu,y_1} - Y_s^{t,x,\mu,y_2} \big|^{\gamma_2} \le C_T e^{-\eta s} |y_1 - y_2|^{\gamma_2}. \tag{4.8}$$

By the boundedness of $\|\partial^2_{xy} b\|$ and condition (2.9), we have

$$\begin{aligned} J_{12} \le\, & \tilde{\mathbb{E}} \big\| \partial^2_{xy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big) \cdot \partial_x Y_s^{t,x,\mu,y_1} - \partial^2_{xy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \cdot \partial_x Y_s^{t,x,\mu,y_1} \big\| \\ & + \tilde{\mathbb{E}} \big\| \partial^2_{xy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \cdot \partial_x Y_s^{t,x,\mu,y_1} - \partial^2_{xy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \cdot \partial_x Y_s^{t,x,\mu,y_2} \big\| \\ \le\, & \tilde{\mathbb{E}} \big[\big\| \partial^2_{xy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big) - \partial^2_{xy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \big\| \big\| \partial_x Y_s^{t,x,\mu,y_1} \big\| \big] \\ & + \tilde{\mathbb{E}} \big\| \partial^2_{xy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \big\| \big\| \partial_x Y_s^{t,x,\mu,y_1} - \partial_x Y_s^{t,x,\mu,y_2} \big\| \\ \le\, & C_T \big[\tilde{\mathbb{E}} \big| Y_s^{t,x,\mu,y_1} - Y_s^{t,x,\mu,y_2} \big|^{2\gamma_2} \big]^{1/2} \big[\tilde{\mathbb{E}} \big\| \partial_x Y_s^{t,x,\mu,y_1} \big\|^2 \big]^{1/2} \\ & + C_T \tilde{\mathbb{E}} \big\| \partial_x Y_s^{t,x,\mu,y_1} - \partial_x Y_s^{t,x,\mu,y_2} \big\|, \end{aligned}$$

where $\partial_x Y_s^{t,x,\mu,y_2}$ satisfies

$$\begin{cases} d\partial_x Y_s^{t,x,\mu,y} = [\partial_x f(t,x,\mu,Y_s^{t,x,\mu,y}) + \partial_y f(t,x,\mu,Y_s^{t,x,\mu,y}) \partial_x Y_s^{t,x,\mu,y}]\, ds \\ \qquad\qquad\qquad + [\partial_x g(t,x,\mu,Y_s^{t,x,\mu,y}) + \partial_y g(t,x,\mu,Y_s^{t,x,\mu,y}) \partial_x Y_s^{t,x,\mu,y}]\, d\tilde{W}_s^2, \\ \partial_x Y_0^{t,x,\mu,y} = 0. \end{cases} \tag{4.9}$$

Note that condition (2.3) in (A1) implies that for any $(t,x,\mu,y) \in [0,t] \times \mathbb{R}^n \times \mathscr{P}_2 \times \mathbb{R}^m$,

$$2\big\langle \partial_y f(t,x,\mu,y) \cdot y, y \big\rangle + 3 \big\| \partial_y g(t,x,\mu,y) \cdot y \big\|^2 \le -\beta |y|^2.$$

Then by Itô's formula, the boundedness of $\partial_x f$, $\partial_x g$, and a straightforward computation, we obtain that

$$\sup_{t\in[0,T], s\ge 0, x\in\mathbb{R}^n, y\in\mathbb{R}^m, \mu\in\mathscr{P}_2} \tilde{\mathbb{E}} \big\| \partial_x Y_s^{t,x,\mu,y} \big\|^4 \le C_T, \tag{4.10}$$

and by Lemma 3.6 and the boundedness of $\partial_{xy} f$, $\partial_{yy} f$, $\partial_{xy} g$ and $\partial_{yy} g$, we have

$$\sup_{t\in[0,T], x\in\mathbb{R}^n, \mu\in\mathscr{P}_2} \tilde{\mathbb{E}} \big\| \partial_x Y_s^{t,x,\mu,y_1} - \partial_x Y_s^{t,x,\mu,y_2} \big\|^2 \le C_T e^{-\frac{\beta s}{2}} |y_1 - y_2|^2. \tag{4.11}$$

Then Lemma 3.6, (4.10) and (4.11) imply that there exists $\eta > 0$ such that

$$J_{12} \le C_T e^{-\eta s} \big(|y_1 - y_2| + 1 \big). \tag{4.12}$$

By condition (2.9) and a similar arguments as in estimating J_{12}, we also have

$$J_{13} \le C e^{-\eta s} \big(|y_1 - y_2| + 1 \big). \tag{4.13}$$

By condition (2.10) and a straightforward computation,

$$\begin{aligned} J_{14} \le\, & \tilde{\mathbb{E}} \big\| \partial^2_{yy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big) \cdot \big(\partial_x Y_s^{t,x,\mu,y_1}, \partial_x Y_s^{t,x,\mu,y_1} \big) \\ & - \partial^2_{yy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \cdot \big(\partial_x Y_s^{t,x,\mu,y_1} \partial_x Y_s^{t,x,\mu,y_1} \big) \big\| \\ & + \tilde{\mathbb{E}} \big\| \partial^2_{yy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \cdot \big(\partial_x Y_s^{t,x,\mu,y_1}, \partial_x Y_s^{t,x,\mu,y_1} \big) \\ & - \partial^2_{yy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \cdot \big(\partial_x Y_s^{t,x,\mu,y_2}, \partial_x Y_s^{t,x,\mu,y_2} \big) \big\| \\ \le\, & \tilde{\mathbb{E}} \big[\big\| \partial^2_{yy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big) - \partial^2_{yy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \big\| \big\| \partial_x Y_s^{t,x,\mu,y_1} \big\|^2 \big] \\ & + \tilde{\mathbb{E}} \big[\big\| \partial^2_{yy} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \big\| \big\| \partial_x Y_s^{t,x,\mu,y_1} - \partial_x Y_s^{t,x,\mu,y_2} \big\| \big(\big\| \partial_x Y_s^{t,x,\mu,y_1} \big\| + \big\| \partial_x Y_s^{t,x,\mu,y_2} \big\| \big) \big] \\ \le\, & C_T \big[\tilde{\mathbb{E}} \big| Y_s^{t,x,\mu,y_1} - Y_s^{t,x,\mu,y_2} \big|^{2\gamma_2} \big]^{1/2} \big[\tilde{\mathbb{E}} \big\| \partial_x Y_s^{t,x,\mu,y_1} \big\|^4 \big]^{1/2} \\ & + C_T \big[\tilde{\mathbb{E}} \big\| \partial_x Y_s^{t,x,\mu,y_1} - \partial_x Y_s^{t,x,\mu,y_2} \big\|^2 \big]^{1/2} \big[\tilde{\mathbb{E}} \big(\big\| \partial_x Y_s^{t,x,\mu,y_1} \big\|^2 + \big\| \partial_x Y_s^{t,x,\mu,y_2} \big\|^2 \big) \big]^{1/2}. \end{aligned}$$

Then by Lemma 3.6, (4.10) and (4.11), we get

$$J_{14} \le C_T e^{-\eta s}\big(|y_1 - y_2| + 1\big). \tag{4.14}$$

Under the assumptions A1–A3, it is easy to prove that

$$\sup_{t\in[0,T], s\ge 0, x\in\mathbb{R}^n, y\in\mathbb{R}^m, \mu\in\mathcal{P}_2} \tilde{\mathbb{E}}\big\|\partial^2_{xx} Y_s^{t,x,\mu,y}\big\|^2 \le C_T \tag{4.15}$$

and

$$\sup_{t\in[0,T], x\in\mathbb{R}^n, \mu\in\mathcal{P}_2} \tilde{\mathbb{E}}\big\|\partial^2_{xx} Y_s^{t,x,\mu,y_1} - \partial^2_{xx} Y_s^{t,x,\mu,y_2}\big\|^2 \le C_T e^{-2\eta s}|y_1 - y_2|^{2\gamma_2}.$$

Then, we get

$$J_{15} \le C_T e^{-\eta s}|y_1 - y_2|^{\gamma_2}. \tag{4.16}$$

Hence, by (4.7), (4.8), (4.12), (4.13), (4.14) and (4.16) we obtain

$$J_1 \le C e^{-\eta s}\big(\tilde{\mathbb{E}}\big|y - Y_{s_0}^{t,x,\mu,y}\big| + 1\big) \le C_T e^{-\eta s}\big\{1 + |x| + |y| + \big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\}.$$

(ii) For the term J_2, note that

$$\begin{aligned}
&\partial^2_{xy}\hat{b}(t,x,\mu,y,s)\\
&\quad = \partial_y \tilde{\mathbb{E}}\big[\partial_x b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)\big] + \partial_y \tilde{\mathbb{E}}\big[\partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)\cdot \partial_x Y_s^{t,x,\mu,y}\big]\\
&\quad = \tilde{\mathbb{E}}\big[\partial^2_{xy} b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)\partial_y Y_s^{t,x,\mu,y}\big] + \tilde{\mathbb{E}}\big[\partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)\partial^2_{xy} Y_s^{t,x,\mu,y}\big]\\
&\qquad + \tilde{\mathbb{E}}\big[\partial^2_{yy} b\big(t,x,\mu,Y_s^{t,x,\mu,y}\big)\cdot\big(\partial_x Y_s^{t,x,\mu,y}, \partial_y Y_s^{t,x,\mu,y}\big)\big].
\end{aligned}$$

Lemma 3.6 and (4.11) imply

$$\sup_{t\in[0,T], x\in\mathbb{R}^n, \mu\in\mathcal{P}_2, y\in\mathbb{R}^m} \big(\tilde{\mathbb{E}}\big\|\partial_y Y_s^{t,x,\mu,y}\big\|^2 + \tilde{\mathbb{E}}\big\|\partial^2_{xy} Y_s^{t,x,\mu,y}\big\|^2\big) \le C_T e^{-\frac{\beta s}{2}}.$$

Hence we have

$$\sup_{t\in[0,T], x\in\mathbb{R}^n, \mu\in\mathcal{P}_2, y\in\mathbb{R}^m} \big\|\partial^2_{xy}\hat{b}(t,x,\mu,y,s)\big\| \le C_T e^{-\frac{\beta s}{4}}.$$

Hence, it is easy to see that

$$J_2 \le C_T e^{-\frac{\beta s}{4}}\tilde{\mathbb{E}}\big\|\partial_x Y_{s_0}^{t,x,\mu,y}\big\| \le C_T e^{\frac{-\beta s}{4}}.$$

(iii) For the term J_3, by a similar argument as in (ii), we have

$$\sup_{t\in[0,T], x\in\mathbb{R}^n, \mu\in\mathcal{P}_2, y\in\mathbb{R}^m} \big\|\partial^2_{yx}\hat{b}(t,x,\mu,y,s)\big\| \le C_T e^{-\frac{\beta s}{4}}$$

and

$$\sup_{t\in[0,T], x\in\mathbb{R}^n, \mu\in\mathcal{P}_2, y\in\mathbb{R}^m} \big\|\partial^2_{yy}\hat{b}(t,x,\mu,y,s)\big\| \le C_T e^{-\frac{\beta s}{4}}.$$

Hence, it is easy to see that

$$J_3 \le C_T e^{\frac{-\beta s}{4}}.$$

(iv) For the term J_4, by estimates (4.15) and (A.7), we easily get

$$J_4 \le C_T e^{\frac{-\beta s}{4}}.$$

Hence, combining (i)–(iv), we prove estimate (4.5).

Step 3. In this step, we intend to prove estimate (4.6). Recall that

$$\partial_\mu \tilde{b}_{s_0}(t,x,\mu,y,s)(z) = \partial_\mu \hat{b}(t,x,\mu,y,s)(z) - \tilde{\mathbb{E}}\partial_\mu \hat{b}\big(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\big)(z) - \tilde{\mathbb{E}}\big[\big\langle \partial_y \hat{b}\big(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\big), \partial_\mu Y_{s_0}^{t,x,\mu,y}(z)\big\rangle\big].$$

So we have

$$\partial_z\partial_\mu \tilde{b}_{s_0}(t,x,\mu,y,s)(z) = \partial_z\partial_\mu \hat{b}(t,x,\mu,y,s)(z) - \tilde{\mathbb{E}}\partial_z\partial_\mu \hat{b}\big(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\big)(z) - \tilde{\mathbb{E}}\big[\big\langle \partial_y \hat{b}\big(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\big), \partial_z\partial_\mu Y_{s_0}^{t,x,\mu,y}(z)\big\rangle\big],$$

where $\partial_z\partial_\mu Y_s^{t,x,\mu,y}(z)$ satisfies

$$\begin{cases} d\partial_z\partial_\mu Y_s^{t,x,\mu,y}(z) = \partial_z\partial_\mu f(t,x,\mu,Y_s^{t,x,\mu,y})(z)\,ds + \partial_y f(t,x,\mu,Y_s^{t,x,\mu,y})\partial_z\partial_\mu Y_s^{t,x,\mu,y}(z)\,ds \\ \qquad + [\partial_z\partial_\mu g(t,x,\mu,Y_s^{t,x,\mu,y})(z) + \partial_y g(t,x,\mu,Y_s^{t,x,\mu,y})\partial_z\partial_\mu Y_s^{t,x,\mu,y}(z)]\,d\tilde{W}_s^2, \\ \partial_z\partial_\mu Y_s^{t,x,\mu,y}(z) = 0. \end{cases} \tag{4.17}$$

Under the assumptions A1–A3, it is easy to prove that for any $T>0$, we have

$$\sup_{t\in[0,T],s\geq 0,x\in\mathbb{R}^n,y\in\mathbb{R}^m,\mu\in\mathscr{P}_2} \tilde{\mathbb{E}}\big\|\partial_z\partial_\mu Y_s^{t,x,\mu,y}\big\|_{L^2(\mu)}^2 \leq C_T \tag{4.18}$$

and there exists $\eta>0$ such that

$$\sup_{t\in[0,T],x\in\mathbb{R}^n,\mu\in\mathscr{P}_2} \tilde{\mathbb{E}}\big\|\partial_z\partial_\mu Y_s^{t,x,\mu,y_1} - \partial_z\partial_\mu Y_s^{t,x,\mu,y_2}\big\|_{L^2(\mu)}^2 \leq C_T e^{-2\eta s}|y_1-y_2|^{2\gamma_2}. \tag{4.19}$$

Then we have

$$\begin{aligned} &\big\|\partial_z\partial_\mu \hat{b}(t,x,\mu,y_1,s) - \partial_z\partial_\mu \hat{b}(t,x,\mu,y_2,s)\big\|_{L^2(\mu)} \\ &\quad = \big\|\partial_z\partial_\mu \tilde{\mathbb{E}}b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big) - \partial_z\partial_\mu \tilde{\mathbb{E}}b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\big\|_{L^2(\mu)} \\ &\quad \leq \tilde{\mathbb{E}}\big\|\partial_z\partial_\mu b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big) - \partial_z\partial_\mu b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\big\|_{L^2(\mu)} \\ &\qquad + \mathbb{E}\big\|\partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big)\partial_z\partial_\mu Y_s^{t,x,\mu,y_1}(z) - \partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\partial_z\partial_\mu Y_s^{t,x,\mu,y_2}\big\|_{L^2(\mu)} \\ &\quad \leq \tilde{\mathbb{E}}\big\|\partial_z\partial_\mu b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big) - \partial_z\partial_\mu b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\big\|_{L^2(\mu)} \\ &\qquad + \tilde{\mathbb{E}}\big\|\partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big)\partial_z\partial_\mu Y_s^{t,x,\mu,y_1} - \partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\partial_z\partial_\mu Y_s^{t,x,\mu,y_1}\big\|_{L^2(\mu)} \\ &\qquad + \tilde{\mathbb{E}}\big\|\partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\partial_z\partial_\mu Y_s^{t,x,\mu,y_1} - \partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\partial_z\partial_\mu Y_s^{t,x,\mu,y_2}\big\|_{L^2(\mu)} \\ &\quad := \sum_{i=1}^3 K_i. \end{aligned}$$

For the terms K_1 and K_2, it follows from condition (2.11) that

$$K_1 \leq C_T\tilde{\mathbb{E}}\big|Y_s^{t,x,\mu,y_1} - Y_s^{t,x,\mu,y_2}\big|^{\gamma_2} \leq C_T e^{-\eta s}|y_1-y_2|^{\gamma_2} \tag{4.20}$$

and by (4.18)

$$\begin{aligned} K_2 &\leq C_T\tilde{\mathbb{E}}\big\|\big(Y_s^{t,x,\mu,y_1} - Y_s^{t,x,\mu,y_2}\big)\partial_z\partial_\mu Y_s^{t,x,\mu,y_1}\big\|_{L^2(\mu)} \\ &\leq C_T\big[\tilde{\mathbb{E}}\big|Y_s^{t,x,\mu,y_1} - Y_s^{t,x,\mu,y_2}\big|^2\big]^{1/2}\big[\tilde{\mathbb{E}}\big\|\partial_z\partial_\mu Y_s^{t,x,\mu,y_1}\big\|_{L^2(\mu)}^2\big]^{1/2} \\ &\leq C_T e^{\frac{-\beta s}{2}}|y_1-y_2|. \end{aligned} \tag{4.21}$$

For the term K_3, by (4.19), it is easy to see that

$$K_3 \leq C_T\tilde{\mathbb{E}}\big\|\partial_z\partial_\mu Y_s^{t,x,\mu,y_1} - \partial_z\partial_\mu Y_s^{t,x,\mu,y_2}\big\|_{L^2(\mu)} \leq C_T e^{-\eta s}|y_1-y_2|^{\gamma_2}. \tag{4.22}$$

Therefore, estimates (4.20) to (4.22) imply

$$\left\|\partial_z\partial_\mu\hat{b}(t,x,\mu,y_1,s)-\partial_z\partial_\mu\hat{b}(t,x,\mu,y_2,s)\right\|_{L^2(\mu)}\le C_T e^{-\eta s}\big(|y_1-y_2|+1\big).$$

Hence, we finally have

$$\begin{aligned}
&\left\|\partial_z\partial_\mu\tilde{b}_{s_0}(t,x,\mu,y,s)\right\|_{L^2(\mu)}\\
&\quad\le \left\|\partial_z\partial_\mu\hat{b}(t,x,\mu,y,s)(z)-\tilde{\mathbb{E}}\partial_z\partial_\mu\hat{b}\big(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\big)(z)\right\|_{L^2(\mu)}\\
&\qquad+\tilde{\mathbb{E}}\big[\left\|\partial_y\hat{b}\big(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\big)\right\|\left\|\partial_z\partial_\mu Y_{s_0}^{t,x,\mu,y}\right\|_{L^2(\mu)}\big]\\
&\quad\le C_T e^{-\eta s}\big(\tilde{\mathbb{E}}\big|y-Y_{s_0}^{t,x,\mu,y}\big|+1\big)+C_T e^{-\eta s}\big[\tilde{\mathbb{E}}\left\|\partial_z\partial_\mu Y_{s_0}^{t,x,\mu,y}\right\|^2_{L^2(\mu)}\big]^{1/2}\\
&\quad\le C_T e^{-\eta s}\big\{1+|x|+|y|+\big[\mu\big(|\cdot|^2\big)\big]^{1/2}\big\},
\end{aligned}$$

which completes the proof of estimate (4.6). □

4.2. *Proof of Theorem* 2.5

Proof. Note that

$$\begin{aligned}
X_t^\epsilon-\bar{X}_t&=\int_0^t\big[b\big(s,X_s^\epsilon,\mathcal{L}_{X_s^\epsilon},Y_s^\epsilon\big)-\bar{b}(s,\bar{X}_s,\mathcal{L}_{\bar{X}_s})\big]\,ds\\
&\quad+\int_0^t\big[\sigma\big(s,X_s^\epsilon,\mathcal{L}_{X_s^\epsilon}\big)-\sigma(s,\bar{X}_s,\mathcal{L}_{\bar{X}_s})\big]\,dW_s^1\\
&=\int_0^t\big[b\big(s,X_s^\epsilon,\mathcal{L}_{X_s^\epsilon},Y_s^\epsilon\big)-\bar{b}\big(s,X_s^\epsilon,\mathcal{L}_{X_s^\epsilon}\big)\big]\,ds\\
&\quad+\int_0^t\big[\bar{b}\big(s,X_s^\epsilon,\mathcal{L}_{X_s^\epsilon}\big)-\bar{b}(s,\bar{X}_s,\mathcal{L}_{\bar{X}_s})\big]\,ds\\
&\quad+\int_0^t\big[\sigma\big(s,X_s^\epsilon,\mathcal{L}_{X_s^\epsilon}\big)-\sigma(s,\bar{X}_s,\mathcal{L}_{\bar{X}_s})\big]\,dW_s^1.
\end{aligned}$$

Then it is easy to see that for any $t\in[0,T]$, we have

$$\begin{aligned}
\sup_{t\in[0,T]}\mathbb{E}\big|X_t^\epsilon-\bar{X}_t\big|^2&\le C\sup_{t\in[0,T]}\mathbb{E}\left|\int_0^t b\big(s,X_s^\epsilon,\mathcal{L}_{X_s^\epsilon},Y_s^\epsilon\big)-\bar{b}\big(s,X_s^\epsilon,\mathcal{L}_{X_s^\epsilon}\big)\,ds\right|^2\\
&\quad+C_T\mathbb{E}\int_0^T\big|X_t^\epsilon-\bar{X}_t\big|^2\,dt.
\end{aligned}$$

Then Grownall's inequality implies that

$$\sup_{t\in[0,T]}\mathbb{E}\big|X_t^\epsilon-\bar{X}_t\big|^2\le C_T\sup_{t\in[0,T]}\mathbb{E}\left|\int_0^t b\big(s,X_s^\epsilon,\mathcal{L}_{X_s^\epsilon},Y_s^\epsilon\big)-\bar{b}\big(s,X_s^\epsilon,\mathcal{L}_{X_s^\epsilon}\big)\,ds\right|^2. \tag{4.23}$$

By Proposition 4.1, there exists $\Phi(t,x,\mu,y)$ such that

$$-\mathcal{L}_2(t,x,\mu)\Phi(t,x,\mu,y)=b(t,x,\mu,y)-\bar{b}(t,x,\mu).$$

Then by Itô's formula for a function which depends on measures (see [5, Theorem 7.1]), we have

$$\begin{aligned}
&\Phi\big(t,X_t^\epsilon,\mathcal{L}_{X_t^\epsilon},Y_t^\epsilon\big)\\
&\quad=\Phi(0,x,\delta_x,y)+\int_0^t\partial_t\Phi\big(s,X_s^\epsilon,\mathcal{L}_{X_s^\epsilon},Y_s^\epsilon\big)\,ds
\end{aligned}$$

$$
\begin{aligned}
&+\int_0^t \mathbb{E}\big[b\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\partial_\mu\Phi(s,x,\mu,y)\big(X_s^\epsilon\big)\big]\big|_{x=X_s^\epsilon,\mu=\mathcal{L}_{X_s^\epsilon},y=Y_s^\epsilon}\,ds \\
&+\int_0^t \frac{1}{2}\mathbb{E}\,\mathrm{Tr}\big[\sigma\sigma^*\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\partial_z\partial_\mu\Phi(s,x,\mu,y)\big(X_s^\epsilon\big)\big]\big|_{x=X_s^\epsilon,\mu=\mathcal{L}_{X_s^\epsilon},y=Y_s^\epsilon}\,ds \\
&+\int_0^t \mathcal{L}_1\big(s, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\Phi\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\,ds \\
&+\frac{1}{\epsilon}\int_0^t \mathcal{L}_2\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\Phi\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\,ds + M_t^{\epsilon,1} + \frac{1}{\sqrt{\epsilon}}M_t^{\epsilon,2},
\end{aligned}
$$

where $\mathcal{L}_1(t,\mu,y)\Phi(t,x,\mu,y) := (\mathcal{L}_1(t,\mu,y)\Phi_1(t,x,\mu,y),\dots,\mathcal{L}_2(t,\mu,y)\Phi_n(t,x,\mu,y))$ with

$$
\begin{aligned}
\mathcal{L}_1(t,\mu,y)\Phi_k(t,x,\mu,y) :=\; &\big\langle b(t,x,\mu,y), \partial_x\Phi(t,x,\mu,y)\big\rangle \\
&+\frac{1}{2}\mathrm{Tr}\big[\sigma\sigma^*(t,x,\mu)\partial_{xx}^2\Phi_k(t,x,\mu,y)\big], \quad k=1,\dots,n,
\end{aligned}
$$

and $M_t^{\epsilon,1}$, $M_t^{\epsilon,2}$ are two martingales, which are defined by

$$
\begin{aligned}
M_t^{\epsilon,1} &:= \int_0^t \partial_x\Phi\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\cdot\sigma\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\,dW_s^1; \\
M_t^{\epsilon,2} &:= \int_0^t \partial_y\Phi\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\cdot g\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\,dW_s^2.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&\sup_{t\in[0,T]}\mathbb{E}\left|\int_0^t b\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big) - \bar{b}\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\,ds\right|^2 \\
&\quad= \sup_{t\in[0,T]}\mathbb{E}\left|\int_0^t \mathcal{L}_2\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\Phi\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\,ds\right|^2 \\
&\quad\le \epsilon^2\sup_{t\in[0,T]}\mathbb{E}\bigg|\Phi\big(t, X_t^\epsilon, \mathcal{L}_{X_t^\epsilon}, Y_t^\epsilon\big) - \Phi(0,x,\delta_x,y) - \int_0^t \partial_t\Phi\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\,ds \\
&\qquad-\int_0^t \mathbb{E}\big[b\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\partial_\mu\Phi(s,x,\mu,y)\big(X_s^\epsilon\big)\big]\big|_{x=X_s^\epsilon,\mu=\mathcal{L}_{X_s^\epsilon},y=Y_s^\epsilon}\,ds \\
&\qquad-\int_0^t \mathbb{E}\,\mathrm{Tr}\big[\sigma\sigma^*\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\partial_z\partial_\mu\Phi(s,x,\mu,y)\big(X_s^\epsilon\big)\big]\big|_{x=X_s^\epsilon,\mu=\mathcal{L}_{X_s^\epsilon},y=Y_s^\epsilon}\,ds \\
&\qquad-\int_0^t \mathcal{L}_1\big(s, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\Phi\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big)\,ds\bigg|^2 \\
&\qquad+\epsilon^2\sup_{t\in[0,T]}\mathbb{E}\big|M_t^{\epsilon,1}\big|^2 + \epsilon\sup_{t\in[0,T]}\mathbb{E}\big|M_t^{\epsilon,2}\big|^2.
\end{aligned}
$$

By Itô's isometry and estimates (4.3) and (4.4), we finally get

$$
\begin{aligned}
&\sup_{t\in[0,T]}\mathbb{E}\left|\int_0^t b\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}, Y_s^\epsilon\big) - \bar{b}\big(s, X_s^\epsilon, \mathcal{L}_{X_s^\epsilon}\big)\,ds\right|^2 \\
&\quad\le C_T\epsilon\Big[\sup_{t\in[0,T]}\mathbb{E}\big|X_t^\epsilon\big|^4 + \sup_{t\in[0,T]}\mathbb{E}\big|Y_t^\epsilon\big|^4 + 1\Big] \\
&\quad\le C_T\big(1+|x|^4+|y|^4\big)\epsilon.
\end{aligned}
$$

This and (4.23) imply the assertion. □

5. Example

Here we give a simple example as an application of our results.

Example 5.1. Let $b_0 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $f_0 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ and satisfying the following conditions:

(1) The first-order partial derivatives $\partial_x b_0(x, y)$, $\partial_y b_0(x, y)$, $\partial_x f_0(x, y)$, $\partial_y f_0(x, y)$ exist for any $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Moreover, all these first-order partial derivatives are bounded uniformly in (x, y) and Lipschitz continuous *w.r.t.* y uniformly in x.

(2) There exists $\beta > 0$ such that for any $x \in \mathbb{R}^n$ and $y_1, y_2 \in \mathbb{R}^m$,

$$\langle f_0(x, y_1) - f_0(x, y_2), y_1 - y_2 \rangle \le -\beta |y_1 - y_2|^2;$$

(3) The second-order partial derivatives $\partial^2_{xx} b_0(x, y)$, $\partial^2_{xy} b_0(x, y)$, $\partial^2_{xx} f_0(x, y)$ and $\partial^2_{xy} f_0(x, y)$ exist for any $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$. Moreover, all these second-order partial derivatives are bounded uniformly in (x, y) and Lipschitz continuous *w.r.t.* y uniformly in x.

Now, let us consider the following slow–fast distribution dependent stochastic differential equations,

$$\begin{cases} dX^\epsilon_t = b(X^\epsilon_t, \mathcal{L}_{X^\epsilon_t}, Y^\epsilon_t)\,dt + dW^1_t, & X^\epsilon_0 = x \in \mathbb{R}^n, \\ dY^\epsilon_t = \frac{1}{\epsilon} f(X^\epsilon_t, \mathcal{L}_{X^\epsilon_t}, Y^\epsilon_t)\,dt + \frac{1}{\sqrt{\epsilon}}\,dW^2_t, & Y^\epsilon_0 = y \in \mathbb{R}^m, \end{cases} \tag{5.1}$$

where $\{W^1_t\}_{t\ge 0}$ and $\{W^2_t\}_{t\ge 0}$ are mutually independent n- and m-dimensional standard Brownian motions and

$$b(x, \mu, y) := \int_{\mathbb{R}^n} b_0(x + z, y)\mu(dz), \qquad f(x, \mu, y) := \int_{\mathbb{R}^n} f_0(x + z, y)\mu(dz).$$

Then we have

$$\partial_\mu b(x, \mu, y)(\cdot) = \partial_x b_0(x + \cdot, y), \qquad \partial_z \partial_\mu b(x, \mu, y)(z) = \partial^2_{xx} b_0(x + z, y)$$

and

$$\partial_\mu f(x, \mu, y)(\cdot) = \partial_x f_0(x + \cdot, y), \qquad \partial_z \partial_\mu f(x, \mu, y)(z) = \partial^2_{xx} f_0(x + z, y).$$

If the conditions (1) and (2) hold, it is easy to check that the coefficients above satisfy assumptions A1–A2. Hence, by Theorem 2.3, we have

$$\sup_{t\in[0,T]} \mathbb{E}\left|X^\epsilon_t - \bar{X}_t\right|^2 \le C\epsilon^{2/3},$$

where $\bar{X}$ solves the corresponding averaged equation.

If the conditions (1)–(3) hold, it is easy to check that the coefficients above satisfy assumptions A1–A3. Hence, by Theorem 2.5, we have

$$\sup_{t\in[0,T]} \mathbb{E}\left|X^\epsilon_t - \bar{X}_t\right|^2 \le C\epsilon,$$

where $\bar{X}$ solves the corresponding averaged equation.

Appendix

In this section, by using the result due to Wang in [38], we prove the existence and uniqueness of solutions to system (1.1) and the corresponding averaged equation.

A.1. Proof of Theorem 2.2

Proof. We set

$$Z_t^\epsilon := \begin{pmatrix} X_t^\epsilon \\ Y_t^\epsilon \end{pmatrix}, \qquad \tilde{b}^\epsilon(t,x,y,\tilde{\mu}) := \begin{pmatrix} b(t,x,\mu,y) \\ \frac{1}{\epsilon} f(t,x,\mu,y) \end{pmatrix}$$

and

$$\tilde{\sigma}^\epsilon(t,x,y,\tilde{\mu}) := \begin{pmatrix} \sigma(t,x,\mu) & 0 \\ 0 & \frac{1}{\sqrt{\epsilon}} g(t,x,\mu,y) \end{pmatrix}, \qquad W_t := \begin{pmatrix} W_t^1 \\ W_t^2 \end{pmatrix},$$

where $t \geq 0$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, $\tilde{\mu} \in \mathscr{P}_2(\mathbb{R}^{n+m})$ with its marginal distribution μ on $\mathbb{R}^n$. Then system (1.1) can be rewritten as the following equation:

$$dZ_t^\epsilon = \tilde{b}^\epsilon\big(t, Z_t^\epsilon, \mathscr{L}_{Z_t^\epsilon}\big)\,dt + \tilde{\sigma}^\epsilon\big(t, Z_t^\epsilon, \mathscr{L}_{Z_t^\epsilon}\big)\,dW_t, \qquad Z_0^\epsilon = \begin{pmatrix} x \\ y \end{pmatrix}. \tag{A.1}$$

Under the assumption A1, we intend to prove that the coefficients in equation (A.1) satisfy Lipschitz and linear growth conditions, uniformly *w.r.t.* $t \in [0,T]$.

In fact, for $T > 0$, and any $z_i = (x_i, y_i) \in \mathbb{R}^{n+m}$, $\tilde{\mu}_i \in \mathscr{P}_2(\mathbb{R}^{n+m})$ with its marginal distributions μ_i on $\mathbb{R}^n$, $i = 1,2$, $t \in [0,T]$

$$\begin{aligned}
&\big|\tilde{b}^\epsilon(t,z_1,\tilde{\mu}_1) - \tilde{b}^\epsilon(t,z_2,\tilde{\mu}_2)\big| + \big\|\tilde{\sigma}^\epsilon(t,z_1,\tilde{\mu}_1) - \tilde{\sigma}^\epsilon(t,z_2,\tilde{\mu}_2)\big\| \\
&\quad \leq \big|b(t,x_1,\mu_1,y_1) - b(t,x_2,\mu_2,y_2)\big| + \big\|\sigma(t,x_1,\mu_1) - \sigma(t,x_2,\mu_2)\big\| \\
&\qquad + \frac{1}{\epsilon}\big|f(t,x_1,\mu_1,y_1) - f(t,x_2,\mu_2,y_2)\big| + \frac{1}{\epsilon}\big\|g(t,x_1,\mu_1,y_1) - g(t,x_2,\mu_2,y_2)\big\| \\
&\quad \leq C_T\left(1 + \frac{1}{\epsilon}\right)\big[|x_1 - x_2| + |y_1 - y_2| + \mathbb{W}_2(\mu_1,\mu_2)\big] \\
&\quad \leq C_T\left(1 + \frac{1}{\epsilon}\right)\big[|z_1 - z_2| + \mathbb{W}_2(\tilde{\mu}_1,\tilde{\mu}_2)\big].
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\big|\tilde{b}^\epsilon(t,z_1,\tilde{\mu}_1)\big| + \big\|\tilde{\sigma}^\epsilon(t,z_1,\tilde{\mu}_1)\big\| \\
&\quad \leq \big|b(t,x_1,\mu_1,y_1)\big| + \big\|\sigma(t,x_1,\mu_1)\big\| + \frac{1}{\epsilon}\big|f(t,x_1,\mu_1,y_1)\big| + \frac{1}{\epsilon}\big\|g(t,x_1,\mu_1,y_1)\big\| \\
&\quad \leq C_T\left(1 + \frac{1}{\epsilon}\right)\big[1 + |x_1| + |y_1| + \mu_1\big(|\cdot|^2\big)\big] \\
&\quad \leq C_T\left(1 + \frac{1}{\epsilon}\right)\big[1 + |z_1| + \tilde{\mu}_1\big(|\cdot|^2\big)\big].
\end{aligned}$$

Hence by [38, Theorem 4.1], there exists a unique solution $\{(X_t^\epsilon, Y_t^\epsilon), t \geq 0\}$ to system (1.1). The proof is complete. □

A.2. Proof of Lemma 3.8

Proof. We first check that the coefficients of Eq. (3.5) satisfy the following condition:

For any $T > 0$, there exists $C_T > 0$ such that for any $t_i \in [0,T]$, $x_i \in \mathbb{R}^n$, $\mu_i \in \mathscr{P}_2$, $i = 1,2$,

$$\begin{aligned}
&\big|\bar{b}(t_1,x_1,\mu_1) - \bar{b}(t_2,x_2,\mu_2)\big| + \big\|\sigma(t_1,x_1,\mu_1) - \sigma(t_2,x_2,\mu_2)\big\| \\
&\quad \leq C_T\big[|t_1 - t_2| + |x_1 - x_2| + \mathbb{W}_2(\mu_1,\mu_2)\big].
\end{aligned} \tag{A.2}$$

Indeed, by Proposition 3.7 and Lemma 3.6, for any $s>0$, we have

$$\begin{aligned}
&\big|\bar b(t_1,x_1,\mu_1)-\bar b(t_2,x_2,\mu_2)\big|+\big\|\sigma(t_1,x_1,\mu_1)-\sigma(t_2,x_2,\mu_2)\big\|\\
&\quad\le\big|\bar b(t_1,x_1,\mu_1)-\tilde{\mathbb{E}}b\big(t_1,x_1,\mu_1,Y_s^{t_1,x_1,\mu_1,0}\big)\big|+\big|\tilde{\mathbb{E}}b\big(t_2,x_2,\mu_2,Y_s^{t_2,x_2,\mu_2,0}\big)-\bar b(t_2,x_2,\mu_2,)\big|\\
&\qquad+\tilde{\mathbb{E}}\big|b\big(t_1,x_1,\mu_1,Y_s^{t_1,x_1,\mu_1,0}\big)-b\big(t_2,x_2,\mu_2,Y_s^{t_2,x_2,\mu_2,0}\big)\big|+\big\|\sigma(t_1,x_1,\mu_1)-\sigma(t_2,x_2,\mu_2)\big\|\\
&\quad\le C_Te^{-\frac{\beta s}{2}}\big(1+|x_1|+|x_2|+\big[\mu_1\big(|\cdot|^2\big)\big]^{1/2}+\big[\mu_2\big(|\cdot|^2\big)\big]^{1/2}\big)\\
&\qquad+C_T\big\{|t_1-t_2|+|x_1-x_2|+\tilde{\mathbb{E}}\big|Y_s^{t_1,x_1,\mu_1,0}-Y_s^{t_2,x_2,\mu_2,0}\big|+\mathbb{W}_2(\mu_1,\mu_2)\big\}\\
&\quad\le C_Te^{-\frac{\beta s}{2}}\big\{1+|x_1|+|x_2|+\big[\mu_1\big(|\cdot|^2\big)\big]^{1/2}+\big[\mu_2\big(|\cdot|^2\big)\big]^{1/2}\big\}\\
&\qquad+C_T\big[|t_1-t_2|+|x_1-x_2|+\mathbb{W}_2(\mu_1,\mu_2)\big].
\end{aligned}$$

Then (A.2) follows by letting $s\to\infty$. Moreover, the estimate (A.2) implies

$$\big|\bar b(t_1,x_1,\mu_1)\big|+\big\|\sigma(t_1,x_1,\mu_1)\big\|\le C_T\big\{1+|x_1|+\big[\mu_1\big(|\cdot|^2\big)\big]^{1/2}\big\}. \tag{A.3}$$

Hence by [38, Theorem 4.1], there exists a unique solution $\{\bar X_t, t\ge 0\}$ to Eq. (3.5) and (3.6) can be easily obtained by following the same arguments as in the proof of Lemma 3.1. The proof is complete. □

A.3. *Proof of* (3.22)–(3.24)

Proof. We here only prove (3.24). (3.22) and (3.23) can be proved by the same procedure. For any $s_0>0$, we define

$$\tilde b_{s_0}(t,x,\mu,y,s):=\hat b(t,x,\mu,y,s)-\hat b(t,x,\mu,y,s+s_0),$$

where $\hat b(t,x,\mu,y,s):=\tilde{\mathbb{E}}b(t,x,\mu,Y_s^{t,x,\mu,y})$. Proposition 3.7 implies that

$$\lim_{s_0\to\infty}\hat b(t,x,\mu,y,s+s_0)=\bar b(t,x,\mu).$$

As a consequence, it is easy to see that

$$\lim_{s_0\to\infty}\tilde b_{s_0}(t,x,\mu,y,s)=\hat b(t,x,\mu,y,s)-\bar b(t,x,\mu)=\tilde b(t,x,\mu,y,s).$$

Hence, in order to prove (3.24), it suffices to show there exists $\eta>0$ such that for any $s_0>0$, $t\in[0,T]$, $s\ge 0$, $x\in\mathbb{R}^n$, $y\in\mathbb{R}^m$ and $\mu_1,\mu_2\in\mathscr{P}_2$,

$$\begin{aligned}
&\big|\tilde b_{s_0}(t,x,\mu_1,y,s)-\tilde b_{s_0}(t,x,\mu_2,y,s)\big|\\
&\quad\le C_T\mathbb{W}_2(\mu_1,\mu_2)e^{-\eta s}\big\{1+|x|^{\gamma_1}+|y|^{\gamma_1}+\big[\mu_1\big(|\cdot|^2\big)\big]^{\gamma_1/2}+\big[\mu_2\big(|\cdot|^2\big)\big]^{\gamma_1/2}\big\},
\end{aligned}$$

which can be obtained by

$$\sup_{t\in[0,T]}\big\|\partial_\mu\tilde b_{s_0}(t,x,\mu,y,s)\big\|_{L^2(\mu)}\le C_Te^{-\eta s}\big\{1+|x|^{\gamma_1}+|y|^{\gamma_1}+\big[\mu\big(|\cdot|^2\big)\big]^{\gamma_1/2}\big\}. \tag{A.4}$$

Indeed, by the Markov property,

$$\begin{aligned}
\tilde b_{s_0}(t,x,\mu,y,s)&=\hat b(t,x,\mu,y,s)-\tilde{\mathbb{E}}b\big(t,x,\mu,Y_{s+s_0}^{t,x,\mu,y}\big)\\
&=\hat b(t,x,\mu,y,s)-\tilde{\mathbb{E}}\big\{\tilde{\mathbb{E}}\big[b\big(t,x,\mu,Y_{s+s_0}^{t,x,\mu,y}\big)\mid\tilde{\mathscr{F}}_{s_0}\big]\big\}\\
&=\hat b(t,x,\mu,y,s)-\tilde{\mathbb{E}}\hat b\big(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\big).
\end{aligned}\tag{A.5}$$

Then we obtain

$$\begin{aligned}
\partial_\mu\tilde b_{s_0}(t,x,\mu,y,s)&=\partial_\mu\hat b(t,x,\mu,y,s)-\tilde{\mathbb{E}}\partial_\mu\hat b\big(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\big)\\
&\quad-\tilde{\mathbb{E}}\big[\big\langle\partial_y\hat b\big(t,x,\mu,Y_{s_0}^{t,x,\mu,y},s\big),\partial_\mu Y_{s_0}^{t,x,\mu,y}\big\rangle\big].
\end{aligned}\tag{A.6}$$

Next, we intend to prove the following two statements.

- For any $t \in [0,T]$, $s \geq 0$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $\mu \in \mathscr{P}_2$,

$$\left\| \partial_y \hat{b}(t,x,\mu,y,s) \right\| \leq C_T e^{-\frac{\beta s}{2}}. \tag{A.7}$$

- For any $t \in [0,T]$, $s \geq 0$, $x \in \mathbb{R}^n$, $y_1, y_2 \in \mathbb{R}^m$ and $\mu \in \mathscr{P}_2$,

$$\left\| \partial_\mu \hat{b}(t,x,\mu,y_1,s) - \partial_\mu \hat{b}(t,x,\mu,y_2,s) \right\|_{L^2(\mu)} \leq C_T e^{-\eta s} |y_1 - y_2|. \tag{A.8}$$

For the first statement, by Lemma 3.6,

$$\begin{aligned} \left|\hat{b}(t,x,\mu,y_1,s) - \hat{b}(t,x,\mu,y_2,s)\right| &= \left|\tilde{\mathbb{E}} b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big) - \tilde{\mathbb{E}} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\right| \\ &\leq C_T \tilde{\mathbb{E}}\left|Y_s^{t,x,\mu,y_1} - Y_s^{t,x,\mu,y_2}\right| \\ &\leq C_T e^{-\frac{\beta s}{2}} |y_1 - y_2|, \end{aligned}$$

which implies (A.7).

For the second statement, the assumptions A1 and A2 imply $Y_s^{t,x,\mu,y}$ that is differentiable *w.r.t.* μ and its derivative $\partial_\mu Y_s^{t,x,\mu,y}(z)$ satisfies

$$\begin{cases} d\partial_\mu Y_s^{t,x,\mu,y}(z) = \partial_\mu f(t,x,\mu,Y_s^{t,x,\mu,y})(z)\,ds + \partial_y f(t,x,\mu,Y_s^{t,x,\mu,y})\partial_\mu Y_s^{t,x,\mu,y}(z)\,ds \\ \qquad\qquad + [\partial_\mu g(t,x,\mu,Y_s^{t,x,\mu,y})(z) + \partial_y g(t,x,\mu,Y_s^{t,x,\mu,y})\partial_\mu Y_s^{t,x,\mu,y}(z)]\,d\tilde{W}_s^2, \\ \partial_\mu Y_s^{t,x,\mu,y}(z) = 0. \end{cases} \tag{A.9}$$

Moreover, it is easy to see that for any $T > 0$, there exists C_T such that

$$\sup_{t\in[0,T],s\geq 0,x\in\mathbb{R}^n,y\in\mathbb{R}^m,\mu\in\mathscr{P}_2} \tilde{\mathbb{E}}\left\| \partial_\mu Y_s^{t,x,\mu,y} \right\|_{L^2(\mu)}^2 \leq C_T.$$

Then we have

$$\begin{aligned} &\left\| \partial_\mu \hat{b}(t,x,\mu,y_1,s) - \partial_\mu \hat{b}(t,x,\mu,y_2,s) \right\|_{L^2(\mu)} \\ &\quad= \left\| \partial_\mu \tilde{\mathbb{E}} b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big) - \partial_\mu \tilde{\mathbb{E}} b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \right\|_{L^2(\mu)} \\ &\quad\leq \tilde{\mathbb{E}}\left\| \partial_\mu b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big) - \partial_\mu b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \right\|_{L^2(\mu)} \\ &\qquad+ \mathbb{E}\left\| \partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big)\partial_\mu Y_s^{t,x,\mu,y_1} - \partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\partial_\mu Y_s^{t,x,\mu,y_2} \right\|_{L^2(\mu)} \\ &\quad\leq \tilde{\mathbb{E}}\left\| \partial_\mu b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big) - \partial_\mu b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big) \right\|_{L^2(\mu)} \\ &\qquad+ \tilde{\mathbb{E}}\left\| \partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_1}\big)\partial_\mu Y_s^{t,x,\mu,y_1} - \partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\partial_\mu Y_s^{t,x,\mu,y_1} \right\|_{L^2(\mu)} \\ &\qquad+ \tilde{\mathbb{E}}\left\| \partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\partial_\mu Y_s^{t,x,\mu,y_1} - \partial_y b\big(t,x,\mu,Y_s^{t,x,\mu,y_2}\big)\partial_\mu Y_s^{t,x,\mu,y_2} \right\|_{L^2(\mu)} \\ &\quad:= \sum_{i=1}^{3} S_i. \end{aligned}$$

For the terms S_1 and S_2, it follows from conditions (2.6), (2.7) and Lemma 3.6 that there exists $\eta > 0$ such that

$$S_1 \leq C_T \tilde{\mathbb{E}}\left|Y_s^{t,x,\mu,y_1} - Y_s^{t,x,\mu,y_2}\right|^{\gamma_1} \leq C_T e^{-\eta s}|y_1 - y_2|^{\gamma_1} \tag{A.10}$$

and

$$\begin{aligned} S_2 &\leq C_T \big[\tilde{\mathbb{E}}\left|Y_s^{t,x,\mu,y_1} - Y_s^{t,x,\mu,y_2}\right|^{2\gamma_1}\big]^{1/2} \big[\tilde{\mathbb{E}}\left\|\partial_\mu Y_s^{t,x,\mu,y_1}\right\|_{L^2(\mu)}^2\big]^{1/2} \\ &\leq C_T e^{-\eta s}|y_1 - y_2|^{\gamma_1}. \end{aligned} \tag{A.11}$$

For the term S_3, by a straightforward computer, we obtain that

$$\tilde{\mathbb{E}}\big\|\partial_\mu Y_s^{t,x,\mu,y_1}-\partial_\mu Y_s^{t,x,\mu,y_2}\big\|_{L^2(\mu)}^2\leq C_T e^{-\frac{\beta s}{2}}|y_1-y_2|^{2\gamma_1},$$

which implies

$$S_3\leq C_T\tilde{\mathbb{E}}\big\|\partial_\mu Y_s^{t,x,\mu,y_1}-\partial_\mu Y_s^{t,x,\mu,y_2}\big\|_{L^2(\mu)}\leq C_T e^{\frac{-\beta s}{4}}|y_1-y_2|^{\gamma_1}. \tag{A.12}$$

Therefore, estimates (A.10) to (A.12) imply (A.8).

Finally, by estimates (A.6), (A.7) and (A.8), there exists $\eta>0$ such that

$$\begin{aligned}\big\|\partial_\mu\tilde{b}_{s_0}(t,x,\mu,y,s)\big\|_{L^2(\mu)}&\leq Ce^{-\eta s}\tilde{\mathbb{E}}\big|y-Y_{s_0}^{t,x,\mu,y}\big|^{\gamma_1}+Ce^{-\eta t}\\&\leq C_T e^{-\eta s}\big\{1+|x|^{\gamma_1}+|y|^{\gamma_1}+\big[\mu\big(|\cdot|^2\big)\big]^{\gamma_1/2}\big\},\end{aligned}$$

which proves (A.4). The proof is complete. □

Acknowledgements

This work is supported by NSFC (No. 11771187, 11931004) and the Priority Academic Program Development of Jiangsu Higher Education Institutions. Financial support by the DFG through CRC 1283 is also gratefully acknowledged.

References

[1] R. Bertram and J. E. Rubin. Multi-timescale systems and fast–slow analysis. *Math. Biosci.* **287** (2017) 105–121. MR3634156 https://doi.org/10.1016/j.mbs.2016.07.003
[2] N. N. Bogoliubov and Y. A. Mitropolsky. *Asymptotic Methods in the Theory of Non-linear Oscillations*. Gordon and Breach Science Publishers, New York, 1961. MR0141845
[3] C. E. Bréhier. Strong and weak orders in averaging for SPDEs. *Stochastic Process. Appl.* **122** (2012) 2553–2593. MR2926167 https://doi.org/10.1016/j.spa.2012.04.007
[4] C. E. Bréhier. Orders of convergence in the averaging principle for SPDEs: The case of a stochastically forced slow component. *Stochastic Process. Appl.* **130** (2020) 3325–3368. MR4092407 https://doi.org/10.1016/j.spa.2019.09.015
[5] R. Buckdahn, J. Li, S. Peng and C. Rainer. Mean-field stochastic differential equations and associated PDEs. *Ann. Probab.* **45** (2) (2017) 824–878. MR3630288 https://doi.org/10.1214/15-AOP1076
[6] P. Cardaliaguet. Notes on mean field games (from P.L. Lions' lectures at Collége de France), 2012. Available at https://www.ceremade.dauphine.fr/cardalia/MFG100629.pdf.
[7] S. Cerrai. A Khasminskii type averaging principle for stochastic reaction–diffusion equations. *Ann. Appl. Probab.* **19** (2009) 899–948. MR2537194 https://doi.org/10.1214/08-AAP560
[8] S. Cerrai. Averaging principle for systems of reaction–diffusion equations with polynomial nonlinearities perturbed by multiplicative noise. *SIAM J. Math. Anal.* **43** (2011) 2482–2518. MR2854919 https://doi.org/10.1137/100806710
[9] S. Cerrai and M. Freidlin. Averaging principle for stochastic reaction–diffusion equations. *Probab. Theory Related Fields* **144** (2009) 137–177. MR2480788 https://doi.org/10.1007/s00440-008-0144-z
[10] S. Cerrai and A. Lunardi. Averaging principle for nonautonomous slow–fast systems of stochastic reaction–diffusion equations: The almost periodic case. *SIAM J. Math. Anal.* **49** (2017) 2843–2884. MR3679916 https://doi.org/10.1137/16M1063307
[11] Y. Chen, Y. Shi and X. Sun. Averaging principle for slow–fast stochastic Burgers equation driven by α-stable process. *Appl. Math. Lett.* **103** (2020) 106199. MR4053803 https://doi.org/10.1016/j.aml.2019.106199
[12] Z. Dong, X. Sun, H. Xiao and J. Zhai. Averaging principle for one dimensional stochastic Burgers equation. *J. Differential Equations* **265** (2018) 4749–4797. MR3848236 https://doi.org/10.1016/j.jde.2018.06.020
[13] W. E and B. Engquist. Multiscale modeling and computations. *Not. Amer. Math. Soc.* **50** (2003) 1062–1070. MR2002752
[14] W. E, D. Liu and E. Vanden-Eijnden. Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.* **58** (2005) 1544–1585. MR2165382 https://doi.org/10.1002/cpa.20088
[15] H. Fu, L. Wan and J. Liu. Strong convergence in averaging principle for stochastic hyperbolic–parabolic equations with two time-scales. *Stochastic Process. Appl.* **125** (2015) 3255–3279. MR3343294 https://doi.org/10.1016/j.spa.2015.03.004
[16] H. Fu, L. Wan, J. Liu and X. Liu. Weak order in averaging principle for stochastic wave equation with a fast oscillation. *Stochastic Process. Appl.* **128** (2018) 2557–2580. MR3811697 https://doi.org/10.1016/j.spa.2017.09.021
[17] P. Gao. Averaging principle for stochastic Kuramoto–Sivashinsky equation with a fast oscillation. *Discrete Contin. Dyn. Syst. Ser. A* **38** (2018) 5649–5684. MR3917783 https://doi.org/10.3934/dcds.2018247
[18] P. Gao. Averaging principle for the higher order nonlinear Schrödinger equation with a random fast oscillation. *J. Stat. Phys.* **171** (2018) 897–926. MR3800899 https://doi.org/10.1007/s10955-018-2048-3
[19] P. Gao. Averaging principle for multiscale stochastic Klein–Gordon-heat system. *J. Nonlinear Sci.* **29** (4) (2019) 1701–1759. MR3993181 https://doi.org/10.1007/s00332-019-09529-4

[20] D. Givon, I. G. Kevrekidis and R. Kupferman. Strong convergence of projective integeration schemes for singularly perturbed stochastic differential systems. *Commun. Math. Sci.* **4** (2006) 707–729. MR2264816

[21] J. Golec. Stochastic averaging principle for systems with pathwise uniqueness. *Stoch. Anal. Appl.* **13** (1995) 307–322. MR1328391 https://doi.org/10.1080/07362999508809400

[22] J. Golec and G. Ladde. Averaging principle and systems of singularly perturbed stochastic differential equations. *J. Math. Phys.* **31** (1990) 1116–1123. MR1050462 https://doi.org/10.1063/1.528792

[23] E. Harvey, V. Kirk, M. Wechselberger and J. Sneyd. Multiple timescales, mixed mode oscillations and canards in models of intracellular calcium dynamics. *J. Nonlinear Sci.* **21** (2011) 639–683. MR2768076

[24] R. Z. Khasminskii. On an averging principle for Itô stochastic differential equations. *Kibernetica* **4** (1968) 260–279. MR0260052

[25] D. Liu. Strong convergence of principle of averaging for multiscale stochastic dynamical systems. *Commun. Math. Sci.* **8** (2010) 999–1020. MR2744917

[26] W. Liu and M. Röckner. *Stochastic Partial Differential Equations: An Introduction*. *Universitext*. Springer, Berlin, 2015. MR3410409 https://doi.org/10.1007/978-3-319-22354-4

[27] W. Liu, M. Röckner, X. Sun and Y. Xie. Strong averaging principle for slow–fast stochastic partial differential equations with locally monotone coefficients, 2019. Available at arXiv:1907.03260v2.

[28] W. Liu, M. Röckner, X. Sun and Y. Xie. Averaging principle for slow–fast stochastic differential equations with time dependent locally Lipschitz coefficients. *J. Differential Equations* **268** (6) (2020) 2910–2948. MR4047972 https://doi.org/10.1016/j.jde.2019.09.047

[29] E. A. Mastny, E. L. Haseltine and J. B. Rawlings. Two classes of quasi-steady-state model reductions for stochastic kinetics. *J. Chem. Phys.* **127** (2007) 094106.

[30] Y. S. Mishura and A. Y. Veretennikov. Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations, 2018. Available at arXiv:1603.02212v8.

[31] E. Pardoux and A. Y. Veretennikov. On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** (3) (2001) 1061–1085. MR1872736 https://doi.org/10.1214/aop/1015345596

[32] E. Pardoux and A. Y. Veretennikov. On the Poisson equation and diffusion approximation. 2. *Ann. Probab.* **31** (3) (2003) 1166–1192. MR1988467 https://doi.org/10.1214/aop/1055425774

[33] P. Ren and F.-Y. Wang. Bismut formula for Lions derivative of distribution dependent SDEs and applications. *J. Differential Equations* **267** (8) (2019) 4745–4777. MR3983053 https://doi.org/10.1016/j.jde.2019.05.016

[34] M. Röckner, X. Sun and L. Xie. Strong and weak convergence in the averaging principle for SDEs with Hölder coefficients, 2019. Available at arXiv:1907.09256v1.

[35] M. Röckner and X. Zhang. Well-posedness of distribution dependent SDEs with singular drifts, 2019. Available at arXiv:1809.02216v3.

[36] X. Sun and J. Zhai. Averaging principle for stochastic real Ginzburg–Landau equation driven by α-stable process. *Commun. Pure Appl. Anal.* **19** (3) (2020) 1291–1319.

[37] A. Y. Veretennikov. On the averaging principle for systems of stochastic differential equations. *Math. USSR, Sb.* **69** (1991) 271–284. MR1046602 https://doi.org/10.1070/SM1991v069n01ABEH001237

[38] F.-Y. Wang. Distribution dependent SDEs for Landau type equations. *Stochastic Process. Appl.* **128** (2018) 595–621. MR3739509 https://doi.org/10.1016/j.spa.2017.05.006

[39] W. Wang and A. J. Roberts. Average and deviation for slow–fast stochastic partial differential equations. *J. Differential Equations* **253** (2012) 1265–1286. MR2927381 https://doi.org/10.1016/j.jde.2012.05.011

[40] W. Wang, A. J. Roberts and J. Duan. Large deviations and approximations for slow–fast stochastic reaction–diffusion equations. *J. Differential Equations* **253** (2012) 3501–3522. MR2981263 https://doi.org/10.1016/j.jde.2012.08.041

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