

# AVERAGING PRINCIPLE AND NORMAL DEVIATION FOR MULTI-SCALE SDES WITH POLYNOMIAL NONLINEARITY

MENGYU CHENG, ZHENXIN LIU, AND MICHAEL RÖCKNER

ABSTRACT. We investigate three types of averaging principles and the normal deviation for multi-scale stochastic differential equations (in short, SDEs) with polynomial nonlinearity. More specifically, we first demonstrate the strong convergence of the solution of SDEs, which involves highly oscillating components and fast processes, to that of the averaged equation. Then we investigate the small fluctuations of the system around its average, and show that the normalized difference weakly converges to an Ornstein-Uhlenbeck type process, which can be viewed as a functional central limit theorem. Additionally, we show that the attractor of the original system tends to that of the averaged equation in probability measure space as the time scale  $\varepsilon$  goes to zero. Finally, we establish the second Bogolyubov theorem; that is to say, we prove that there exists a quasi-periodic solution in a neighborhood of the stationary solution of the averaged equation when the  $\varepsilon$  is small.

## CONTENTS

|  |    |
|--|----|
| 1. Introduction  | 2  |
| 2. Statement of the main results                                 | 6  |
| 2.1. The first Bogolyubov theorem                                | 6  |
| 2.2. Normal deviation  | 7  |
| 2.3. Global averaging principle                                  | 8  |
| 2.4. The second Bogolyubov theorem                               | 10 |
| 2.5. Examples  | 10 |
| 3. Frozen equation and Poisson equation                          | 12 |
| 4. The first Bogolyubov theorem                                  | 14 |
| 4.1. Auxiliary lemmas  | 15 |
| 4.2. Proof of Theorem 2.3  | 18 |
| 5. Normal deviation  | 21 |
| 5.1. The optimal strong convergence rate                         | 21 |
| 5.2. Proof of Theorem 2.6  | 23 |
| 6. The global averaging principle and second averaging principle | 30 |
| 6.1. Preliminaries for dynamical systems                         | 30 |
| 6.2. Proof of Theorem 2.8  | 31 |
| 6.3. Proof of Theorem 2.12                                       | 35 |
| Appendix A.  | 36 |
| Acknowledgements   | 37 |
| References   | 38 |

---

2010 *Mathematics Subject Classification.* 34C29, 34F05, 60F05, 37C55.

*Key words and phrases.* Multi-scale dynamical system; Averaging principle; Normal deviation; Quasi-periodic solutions; Pullback attractors.

## 1. INTRODUCTION

Consider the following periodically forced Van der Pol's equation:

$$y'' + \mu(y^2 - 1)y' + y = a \sin(2\pi\nu t), \quad (1.1)$$

where  $\mu \gg 1$ ,  $\nu$  represents the frequency and  $a$  the amplitude of the forcing. Let  $t = \mu\tau$  and

$$x = \frac{1}{\mu^2} \frac{dy}{d\tau} + \frac{y^3}{3} - 3,$$

which is also called the Liénard transformation. Define  $\varepsilon = \frac{1}{\mu^2}$ . Hence we can transform (1.1) into the following system

$$\begin{cases} \frac{dx}{d\tau} = a \sin\left(2\pi\nu \frac{1}{\sqrt{\varepsilon}}\tau\right) - y \\ \frac{dy}{d\tau} = \frac{1}{\varepsilon} \left(x - \frac{y^3}{3} + y\right). \end{cases} \quad (1.2)$$

Let  $\theta = \frac{1}{\sqrt{\varepsilon}}\tau$ . Then note that (1.2) is a multi-scale system, including the slow variable  $x$ , the fast variable  $y$  and the highly oscillating time component  $\theta$ . Van der Pol's equation is one of the most important examples of multi-scale systems. It exhibits a wide variety of interesting dynamical phenomena and appears frequently in applications in various fields, including, but not limited to, neuroscience, seismology, electrical circuits, networks, and systems biology. Multi-scale models appear frequently in many real-world dynamical systems, such as climate weather interactions (see e.g. [29, 37]), macro-molecules (see e.g. [3, 25]), stochastic volatility in finance (see e.g. [15]), etc.

Usually, studying multi-scale models is relatively difficult because of the presence of widely separated times scales and the interactions between them. To understand the dynamics of multi-scale models, it is desirable to seek a simplified system, which can simulate and predict the evolution of the original system over a long time scale. This is the basic idea of the averaging principle.

The averaging principle was first developed for deterministic systems by Krylov, Bogolyubov and Miltropolsky [4, 34], and extended to SDEs by Khasminskii [27]. After that, numerous studies have been carried out on the averaging principle for SDEs, see e.g. [2, 16, 19, 28, 36, 38, 40, 44, 46] and the references therein. Furthermore, similar results concerning stochastic partial differential equations can be found in references like [5, 7, 9, 10, 12, 13, 18, 21, 45]. Despite considerable advances in the averaging principle, it seems that there is no work on multi-scale SDEs which includes the slow variable, the fast variable and the highly oscillating time component.

Building upon the motivations mentioned above, in this paper we investigate the averaging principle of the following multi-scale SDEs with polynomial nonlinearity:

$$\begin{cases} dX_t^\varepsilon = f(\varepsilon^{-\gamma}t, X_t^\varepsilon, Y_t^\varepsilon)dt + \sigma(\varepsilon^{-\gamma}t, X_t^\varepsilon)dW_t^1 \\ dY_t^\varepsilon = (\varepsilon^{-2\alpha}B(X_t^\varepsilon, Y_t^\varepsilon) + \varepsilon^{-\beta}b(X_t^\varepsilon, Y_t^\varepsilon))dt + \varepsilon^{-\alpha}g(X_t^\varepsilon, Y_t^\varepsilon)dW_t^2, \end{cases} \quad (1.3)$$

where  $0 \leq \beta < 2\alpha$ ,  $0 < \gamma < 2\alpha$ ,  $f: \mathbb{R}^{1+d_1+d_2} \rightarrow \mathbb{R}^{d_1}$ ,  $\sigma: \mathbb{R}^{1+d_1} \rightarrow \mathbb{R}^{d_1} \otimes \mathbb{R}^{d_1}$ ,  $B: \mathbb{R}^{d_1+d_2} \rightarrow \mathbb{R}^{d_2}$ ,  $b: \mathbb{R}^{d_1+d_2} \rightarrow \mathbb{R}^{d_2}$ ,  $g: \mathbb{R}^{d_1+d_2} \rightarrow \mathbb{R}^{d_2} \otimes \mathbb{R}^{d_2}$ , and  $0 < \varepsilon \ll 1$  is a small parameter. Here  $W^1$  and  $W^2$  are independent standard Brownian motions. See Section 2 for detailed conditions for coefficients  $f$ ,  $\sigma$ ,  $B$ ,  $b$  and  $g$ . If  $f$  and  $\sigma$  are time-independent, and  $b \equiv 0$  then (1.3) reduces to the classical slow-fast SDEs.

More precisely, as the time scale  $\varepsilon$  goes to zero we first consider the so-called *first Bogolyubov theorem*, which focuses on the strong convergence of the solution of (1.3) to that of the following averaged equation on finite time intervals:

$$d\bar{X}_t = \bar{f}(\bar{X}_t)dt + \bar{\sigma}(\bar{X}_t)dW_t^1, \quad (1.4)$$

where

$$\bar{f}(x) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_t^{t+T} \int_{\mathbb{R}^{d_2}} f(s, x, y) \mu^x(dy) ds, \quad \lim_{T \rightarrow \infty} \frac{1}{T} \int_t^{t+T} |\sigma(s, x) - \bar{\sigma}(x)|_{HS}^2 ds = 0$$

for all  $(t, x) \in \mathbb{R}^{1+d_1}$  (see more details about the assumptions in Section 2.1), where  $|\cdot|_{HS}$  is the Hilbert-Schmidt norm; see Theorem 2.3. Here  $\mu^x$  is the invariant measure of

$$dY_t^x = B(x, Y_t^x)dt + g(x, Y_t^x)dW_t. \quad (1.5)$$

Secondly, we consider the case where the coefficients  $f$  and  $\sigma$  are time-independent. In this case, we assume  $2\alpha = 1$  for simplicity. Then by using regularity estimates for the solutions to Poisson equations, we obtain the optimal strong convergence rate, i.e.

$$\mathbb{E} \left( \sup_{t \in [0, T]} |X_t^\varepsilon - \bar{X}_t|^2 \right) \leq C_T \varepsilon, \quad (1.6)$$

where  $C_T$  is a constant (see Theorem 2.5). Furthermore, if  $\sigma$  is constant, we study the *normal deviation*. In other words, we prove that the normalized difference

$$Z_t^\varepsilon := \frac{X_t^\varepsilon - \bar{X}_t}{\sqrt{\varepsilon}}$$

weakly converges to  $\bar{Z}_t$  as  $\varepsilon$  goes to zero. Here  $\bar{Z}_t$  is the solution to

$$d\bar{Z}_t = \nabla \bar{f}(\bar{X}_t) \bar{Z}_t dt + G(\bar{X}_t) d\widetilde{W}_t^1, \quad \bar{Z}_0 = 0 \in \mathbb{R}^{d_1}, \quad (1.7)$$

where

$$G(x) = \sqrt{\int_0^\infty \int_{\mathbb{R}^{d_2}} \mathbb{E} [f(x, Y_t^x(y)) - \bar{f}(x)] [f(x, y) - \bar{f}(x)]^T \mu^x(dy) dt},$$

and  $\widetilde{W}_t^1$  is another standard Brownian motion that is independent of  $W_t^1$  (see Theorem 2.6). Here  $Y_t^x(y), t \geq 0$  is the solution to (1.5) with  $Y_0^x = y$ . Such a result is also known as the Gaussian approximation. In addition, our investigation includes the study of the *global averaging principle* in the weak sense, i.e. we prove that the measure attractor of (1.3) converges, as  $\varepsilon$  goes to zero, to that of (1.4) (see Theorem 2.8). Finally, we establish the *second Bogolyubov theorem*, which states that the stationary solution of (1.4) approximates the recurrent solution of (1.3) in the sense of (2.6) in Theorem 2.12.

Compared with [36], where they proved the first Bogolyubov theorem for two time scale SDEs with locally Lipschitz coefficients, we study a broader class of SDEs (1.3). The slow process  $X_t^\varepsilon$  here interacts not only with the fast process  $Y_t^\varepsilon$  but also with the highly oscillating time component  $\varepsilon^{-\gamma}t$ . To overcome the difficulty, we employ the Poisson equation, the technique of time discretization and the technique of truncation. And we also obtain the optimal rate of strong convergence when the coefficients of the slow equation are time-independent and  $f$  satisfies the following locally monotone condition: for all  $x_1, x_2 \in \mathbb{R}^{d_1}$  and  $y \in \mathbb{R}^{d_2}$

$$\langle f(x_1, y) - f(x_2, y), x_1 - x_2 \rangle \leq M (1 + |y|^{\theta_2}) |x_1 - x_2|^2, \quad (1.8)$$

where  $M \geq 0$  and  $\theta_2 \geq 1$ ; see Section 2.2 for more detailed conditions.

The rate of convergence is interesting in its own right since it plays a crucial role in constructing efficient numerical schemes. The main motivation comes from the well-known Heterogeneous Multi-Scale Methods used to approximate the slow component; see e.g. [14, 31]. Recall that the optimal strong convergence order is also obtained in [43] for monotone SDEs. However, it should be noted that our result cannot be covered by those in [43] because there it is assumed that the coefficient  $f$  must be monotone uniformly with respect to (in short, w.r.t.)  $y$ . There are certain classes of systems, such as  $f(x, y) = x - x^3 + y^k \sin x$ ,  $k \in \mathbb{N}$ , that do not satisfy monotonicity uniformly w.r.t.  $y$  but satisfy condition (1.8); see Example 2.14.

In order to obtain the optimal strong convergence order under monotonicity conditions, we need to estimate two crucial terms:

$$\mathcal{I}_1 := \mathbb{E} \left( \sup_{t \in [0, T]} \int_0^t \langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(X_s^\varepsilon), X_s^\varepsilon - \bar{X}_s \rangle ds \right),$$

$$\mathcal{I}_2 := \mathbb{E} \left( \sup_{t \in [0, T]} \int_0^t \langle \bar{f}(X_s^\varepsilon) - \bar{f}(\bar{X}_s), X_s^\varepsilon - \bar{X}_s \rangle ds \right).$$

For  $\mathcal{I}_1$ , regularity estimates for solutions to the Poisson equation can be employed to handle its estimation. The remaining part of the proof is to show that  $\bar{f}$  is monotone when  $f$  satisfies the locally monotone condition (1.8). Actually, thanks to the stability of the stationary solution to (1.5), we can complete the proof (see Lemma 5.1 for more details).

The first Bogolyubov theorem can be viewed as a functional law of large numbers, indicating the convergence of the slow process  $X_t^\varepsilon$  to the averaged process  $\bar{X}_t$ . However, it is crucial to acknowledge that even for small positive values of  $\varepsilon$ , the slow process  $X_t^\varepsilon$  still experiences fluctuations around the averaged process  $\bar{X}_t$ . Consequently, it is natural to go one step further and consider the functional central limit theorem, i.e. the normal deviation. By studying these deviations, we can contribute to the understanding of the behavior of the system and its relationship with the averaged process. The fundamental paper about the normal deviation of multi-scale SDEs is by Khasminskii [26]. Since then, further developments were acquired; see e.g. [6, 22, 30, 31, 32, 39, 40, 45].

To the best of our knowledge, it seems that there is no work on the normal deviation for SDEs with polynomial nonlinearity. Therefore, we focus on investigating the deviations of the solutions  $X_t^\varepsilon$  to monotone SDEs of type (1.3) from  $\bar{X}_t$  in this paper. Specifically, we establish that, under appropriate conditions, the deviation process  $Z_t^\varepsilon$  converges weakly to an Ornstein-Uhlenbeck type process  $\bar{Z}_t$ . Such a result is closely related to the homogenization for solutions of partial differential equations with singularly perturbed terms; see e.g. [17].

More specifically, we prove that for any  $\varphi \in C_b^\infty(\mathbb{R}^{d_1})$

$$\lim_{\varepsilon \rightarrow 0} \sup_{t \in [0, T]} |\mathbb{E}\varphi(Z_t^\varepsilon) - \mathbb{E}\varphi(\bar{Z}_t)| = 0, \quad (1.9)$$

where  $C_b^\infty(\mathbb{R}^{d_1})$  is the space of all smooth functions with bounded  $j$ -th derivatives for all integers  $j \in [0, \infty)$ . To this end, employing Itô's formula, we have

$$\begin{aligned} & |\mathbb{E}\varphi(Z_t^\varepsilon) - \mathbb{E}\varphi(\bar{Z}_t)| \\ & \leq \left| \mathbb{E} \int_0^t \frac{1}{\sqrt{\varepsilon}} \langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(X_s^\varepsilon), \nabla\varphi(Z_s^\varepsilon) \rangle - \frac{1}{2} \text{Tr}[\nabla^2\varphi(Z_s^\varepsilon)GG^T(X_s^\varepsilon)] ds \right| \\ & \quad + \left| \mathbb{E} \int_0^t \langle \nabla\bar{f}(\bar{X}_s + \iota(X_s^\varepsilon - \bar{X}_s))Z_s^\varepsilon, \nabla\varphi(Z_s^\varepsilon) \rangle - \langle \nabla\bar{f}(\bar{X}_s)\bar{Z}_s, \nabla\varphi(\bar{Z}_s) \rangle ds \right| \\ & \quad + \left| \mathbb{E} \int_0^t \frac{1}{2} \text{Tr}[\nabla^2\varphi(Z_s^\varepsilon)GG^T(X_s^\varepsilon)] - \frac{1}{2} \text{Tr}[\nabla^2\varphi(\bar{Z}_s)GG^T(\bar{X}_s)] ds \right| =: \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3, \end{aligned}$$

where  $\iota \in [0, 1]$ . Therefore, we just need to show that  $\mathcal{I}_1$ ,  $\mathcal{I}_2$  and  $\mathcal{I}_3$  go to zero as  $\varepsilon \rightarrow 0$ . Combining the regularity estimates of the solutions to the Poisson equation and the optimal strong convergence (1.6), we prove that  $\lim_{\varepsilon \rightarrow 0} \mathcal{I}_1 = 0$ . For  $\mathcal{I}_2$  and  $\mathcal{I}_3$ , we first prove that the subset

$$\{Z^\varepsilon, \bar{Z} : 0 < \varepsilon \leq 1\} \subset C([0, T]; \mathbb{R}^{d_1})$$

is tight; see Lemma 5.3. By utilizing the tightness of the set and the separation properties of  $C([0, T]; \mathbb{R}^{d_1})$ , we can then conclude that  $\lim_{\varepsilon \rightarrow 0} \mathcal{I}_2 = 0$  and  $\lim_{\varepsilon \rightarrow 0} \mathcal{I}_3 = 0$ , and complete the proof of (1.9); see Section 5.2 for details. This provides a better approximation and is also known as Van Kampen's scheme in physics; see e.g. [1].

Another main ingredient of this paper is to study the long-time asymptotic behavior of solutions to (1.3). So, we aim to establish the global averaging principle in the weak sense. Namely, we prove that the attractor of (1.3) tends to that of (1.4) in the space of probability measures. The global averaging principle of deterministic systems was proved in [20, 23, 24, 47] and the references therein. There are few works on the global averaging principle for stochastic systems. In [9, 10] the global averaging principle was established in the weak sense for stochastic partial differential equations with highly time oscillating components.

As we mentioned before, the drift coefficient  $f$  in (1.3) exhibits a complex and general structure, and it encompasses not only the highly time oscillating component  $\varepsilon^{-\gamma}t$ , but also the fast variable  $Y_t^\varepsilon$ . Due to the coupling between the slow process  $X_t^\varepsilon$  and the fast process  $Y_t^\varepsilon$  in (1.3), it is necessary to consider the attractor of the entire multi-scale system as a unified entity. The dynamics of the slow process and the fast process are interconnected, and their mutual influence plays a role in shaping the behavior of the system as a whole. Therefore, we consider the attractor of the whole multi-scale system (1.3) instead of the single slow equation.

More exactly, fix  $0 < \varepsilon \leq 1$ . Define the transition probability

$$P^\varepsilon(s, (x, y), t, D) := \mathbb{P} \circ [(X_{s,t}^\varepsilon(x), Y_{s,t}^\varepsilon(y))]^{-1}(D)$$

for all  $s \leq t$ ,  $(x, y) \in \mathbb{R}^{d_1+d_2}$  and  $D \in \mathcal{B}(\mathbb{R}^{d_1+d_2})$ , where  $\mathcal{B}(\mathbb{R}^{d_1+d_2})$  is the Borel  $\sigma$ -algebra of  $\mathbb{R}^{d_1+d_2}$ . Then for any  $0 < \varepsilon \leq 1$  it associates the Markov operator  $P_\varepsilon^*$  acting on the probability measure space  $\mathcal{P}(\mathbb{R}^{d_1+d_2})$ :

$$P_\varepsilon^*(s, t, m)(D) := \int_{\mathbb{R}^{d_1+d_2}} P^\varepsilon(s, (x, y), t, D) m(d(x, y)) \quad (1.10)$$

for any  $m \in \mathcal{P}(\mathbb{R}^{d_1+d_2})$  and  $D \in \mathcal{B}(\mathbb{R}^{d_1+d_2})$ .

Note that  $P_\varepsilon^*$  is time inhomogeneous, so we employ the method of skew product to consider its pullback attractors. For detailed definitions of skew product flows, attractors, and pullback attractors, please see Section 6.1. Finally, we investigate the convergence of pullback measure attractors for (1.3).

Finally, we establish the second Bogolyubov theorem for (1.3). Since numerous physical models have periodic forces, such as (1.1), we approximate the periodic solution of the original system by utilizing the stationary solution to the averaged equation. It is worth noting that we obtain the convergence of a broader class of recurrent solutions, including periodic, quasi-periodic, almost periodic solutions among others; see Remark 2.13. For brevity, we focus on the analysis of quasi-periodic solutions in this paper. To be specific, under some suitable conditions we show that there exists a unique solution to (1.3), which is quasi-periodic in distribution, if  $f$  and  $\sigma$  are quasi-periodic. Then the law of the slow component of the quasi-periodic solution converges to the law of the stationary solution for (1.4) uniformly w.r.t.  $t \in \mathbb{R}$  as the time scale  $\varepsilon$  goes to zero.

Now we summarize the structure of the paper. In Section 2, we state our main results. In Section 3, we study the frozen equation and the Poisson equation. In Section 4, we investigate the first Bogolyubov theorem. In Section 5, we first prove the optimal strong convergence order. Then we establish the normal deviation. In Section 6, we prove the global averaging principle in the weak sense and the second Bogolyubov theorem. In the appendix at the end, we show the existence and uniqueness of solutions to (1.3).

**Notations.** Throughout this paper, let  $|\cdot|$  be the Euclidean norm and  $\langle \cdot, \cdot \rangle$  be the Euclidean inner product on  $\mathbb{R}^d$ ,  $d \in \mathbb{N}$ . For a vector-valued or matrix-valued function  $x \mapsto \varphi(x)$  defined on  $\mathbb{R}^d$  or  $(x, y) \mapsto \varphi(x, y)$  defined on  $\mathbb{R}^{d_1+d_2}$ ,  $d_1, d_2 \in \mathbb{N}$ , we denote the  $i$ -th order derivative of  $\varphi$  by  $\nabla^i \varphi(x)$ , and the  $i$ -th and  $j$ -th order partial derivative of  $\varphi(x, y)$  w.r.t.  $x$  and  $y$  by  $\partial_x^i \partial_y^j \varphi(x, y)$ , respectively, where  $i, j \in \mathbb{N}$ . For all  $i \in \mathbb{N}$ , let  $C^{i,2,4}(\mathbb{R}^{d+d_1+d_2})$  be the space of all continuous mappings  $f : \mathbb{R}^{d+d_1+d_2} \rightarrow \mathbb{R}^{d_1+d_2}$  such that  $\partial_h^{i'} f$  and  $\partial_x^{j'} \partial_y^{k'} f$  are continuous for any  $0 \leq i' \leq i$  and  $0 \leq 2j' + k' \leq 4$ . Let  $[C]$  denote the integer part of  $C$  for any  $C \geq 0$ . We use  $C_b^k(\mathbb{R}^{d_1})$  to denote the space of all functions  $f : \mathbb{R}^{d_1} \rightarrow \mathbb{R}$  whose  $j$ -th derivative is continuous and bounded for all  $j \in [0, k]$ . For any complete metric space  $(\mathcal{X}, d)$ , let  $C(\mathbb{R}, \mathcal{X})$  be the space of all continuous mappings  $\varphi : \mathbb{R} \rightarrow \mathcal{X}$  with the compact-open topology. Define the Hausdorff semi-metric  $\text{dist}_{\mathcal{X}}(A, B) := \sup_{x \in A} \inf_{y \in B} d(x, y)$  for any  $A, B \subset \mathcal{X}$ . Let  $A^T$  denote the transpose of a matrix  $A$ . Let  $\mathcal{L}(X)$  denote the distribution or law of random variable  $X$ , and  $(\Omega, \mathcal{F}, \mathbb{P})$  be a complete probability space. We use  $C$  with or without subscripts to denote some constant, which

may change from line to line. In this paper, solutions to SDEs are always meant to be strong solutions.

## 2. STATEMENT OF THE MAIN RESULTS

In this section, we formulate our main results.

**2.1. The first Bogolyubov theorem.** First of all, we introduce the following conditions about the coefficients  $B$ ,  $b$ , and  $g$ :

( $\mathbf{H}_y^1$ ) There exist constants  $\eta > 0$ ,  $\eta' \geq 0$ ,  $\theta \geq 2$ ,  $\tilde{\eta} \in \mathbb{R}$  and  $K_1 \in \mathbb{R}$  such that for all  $(x, y) \in \mathbb{R}^{d_1+d_2}$

$$2\langle B(x, y), y \rangle + |g(x, y)|_{HS}^2 \leq -\eta|y|^2 - \eta'|y|^\theta + K_1, \quad 2\langle b(x, y), y \rangle \leq \tilde{\eta}|y|^2 + K_1.$$

( $\mathbf{H}_y^2$ ) There exists a constant  $K_2 > 0$  such that for all  $(x, y) \in \mathbb{R}^{d_1+d_2}$

$$K_2^{-1}I \leq a(x, y) \leq K_2I,$$

where  $a(x, y) = \frac{1}{2}gg^T(x, y)$ .

( $\mathbf{H}_y^3$ ) (i) There exist constants  $\eta > 0$  and  $\theta \geq 2$  such that for all  $x \in \mathbb{R}^{d_1}, y_1, y_2 \in \mathbb{R}^{d_2}$

$$2\langle B(x, y_1) - B(x, y_2), y_1 - y_2 \rangle + |g(x, y_1) - g(x, y_2)|_{HS}^2 \leq -\eta|y_1 - y_2|^2.$$

(ii) There exists a constant  $L_g > 0$  such that for all  $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^{d_1+d_2}$

$$|g(x_1, y_1) - g(x_2, y_2)|_{HS} \leq L_g(|x_1 - x_2| + |y_1 - y_2|).$$

(iii) There exist constants  $\kappa_1, \kappa_2 \geq 1$  and  $K_3 > 0$  such that for all  $x_1, x_2 \in \mathbb{R}^{d_1}$  and  $y \in \mathbb{R}^{d_2}$

$$|B(x_1, y) - B(x_2, y)| \leq K_3(1 + |y|^{\kappa_2})|x_1 - x_2|,$$

$$\begin{aligned} & |b(x_1, y_1) - b(x_2, y_2)| \\ & \leq K_3(1 + |x_1|^{\kappa_1} + |x_2|^{\kappa_1} + |y_1|^{\kappa_2} + |y_2|^{\kappa_2})(|x_1 - x_2| + |y_1 - y_2|). \end{aligned}$$

( $\mathbf{H}_y^4$ ) There exist constants  $\varsigma > 0$  and  $\varsigma_1 > 8$  such that for all  $(x, y) \in \mathbb{R}^{d_1+d_2}$  and  $\xi \in \mathbb{R}^{d_2}$

$$2\langle \partial_y B(x, y)\xi, \xi \rangle + (\varsigma_1 - 1)|\partial_y g(x, y)\xi|^2 \leq -\varsigma|\xi|^2.$$

( $\mathbf{H}_y^5$ ) There exist  $K_3 > 0$  and  $\kappa_1 \geq 1$  such that  $B \in C^{3,3}(\mathbb{R}^{d_1+d_2})$  and  $g \in C^{3,3}(\mathbb{R}^{d_1+d_2})$  satisfy

$$\sum_{1 \leq i+j \leq 3} (|\partial_y^j \partial_x^i B| + |\partial_y^j \partial_x^i g|) \leq K_3(1 + |y|^\kappa).$$

**Remark 2.1.** Note that ( $\mathbf{H}_y^4$ ) and ( $\mathbf{H}_y^5$ ) are not necessarily required if we study the averaging principle by the technique of time discretization. However, in this context, we aim to investigate the optimal strong convergence rate for the averaging principle based on the Poisson equation, and we need ( $\mathbf{H}_y^4$ )–( $\mathbf{H}_y^5$ ) to obtain the well-posedness of the Poisson equation (see e.g. [11, 43]).

Note that if ( $\mathbf{H}_y^1$ ) and ( $\mathbf{H}_y^2$ ) hold, then for any  $x \in \mathbb{R}^{d_1}$

$$dY_t^x = B(x, Y_t^x)dt + g(x, Y_t^x)dW_t^2 \quad (2.1)$$

admits a unique invariant measure  $\mu^x$ ; see e.g. [42]. Set

$$\hat{f}(t, x) := \int_{\mathbb{R}^{d_2}} f(t, x, y)\mu^x(dy), \quad \forall (t, x, y) \in \mathbb{R}^{1+d_1+d_2}.$$

Next, we introduce the conditions concerning the coefficients  $f$  and  $\sigma$ :

( $\mathbf{A}_f$ ) Let  $R \in \mathbb{R}_+$ . There exist  $\omega_R^f : \mathbb{R} \rightarrow \mathbb{R}_+$  satisfying  $\omega_R^f(T) \rightarrow 0$  as  $T \rightarrow \infty$  and  $\bar{f} : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_1}$  such that for all  $t \in \mathbb{R}$  and  $|x| \leq R$

$$\frac{1}{T} \left| \int_t^{t+T} (\hat{f}(s, x) - \bar{f}(x))ds \right| \leq \omega_R^f(T).$$

( $\mathbf{A}_\sigma$ ) There exists  $\omega^\sigma : \mathbb{R} \rightarrow \mathbb{R}_+$  satisfying  $\omega^\sigma(T) \rightarrow 0$  as  $T \rightarrow \infty$  and  $\bar{\sigma} : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_1} \otimes \mathbb{R}^{d_1}$  such that for all  $(t, x) \in \mathbb{R}^{1+d_1}$

$$\frac{1}{T} \int_t^{t+T} |\sigma(s, x) - \bar{\sigma}(x)|_{HS}^2 ds \leq \omega^\sigma(T)(1 + |x|^2).$$

( $\mathbf{H}_x^1$ ) There exist constants  $K_4, K_5 \in \mathbb{R}$  such that for all  $(t, x, y) \in \mathbb{R}^{1+d_1+d_2}$

$$2\langle f(t, x, y), x \rangle + |\sigma(t, x)|^2 \leq K_4(1 + |x|^2) + K_5|y|^\theta,$$

where  $\theta$  is as in ( $\mathbf{H}_y^1$ ).

( $\mathbf{H}_x^2$ ) There exist constants  $K_6 > 0, \theta_1, \theta_2 > 1$  such that for all  $(t, x, y) \in \mathbb{R}^{1+d_1+d_2}$

$$|\partial_t f(t, x, y)| + \sum_{0 \leq 2i+j \leq 4} |\partial_y^j \partial_x^i f(t, x, y)| \leq K_6(1 + |x|^{\theta_1} + |y|^{\theta_2}).$$

( $\mathbf{H}_x^3$ ) There exist constants  $K_7, L_\sigma > 0$  such that for all  $t \in \mathbb{R}$  and  $x_1, x_2 \in \mathbb{R}^{d_1}$

$$|\sigma(t, x_1) - \sigma(t, x_2)|_{HS} \leq L_\sigma |x_1 - x_2|, \quad |\sigma(t, 0)| \leq K_7.$$

**Remark 2.2.** (i) If  $\theta = 2$  or  $K_5 = 0$  in ( $\mathbf{H}_x^1$ ), then we can assume that  $\eta' = 0$  in ( $\mathbf{H}_y^1$ ).

(ii) Note that ( $\mathbf{H}_x^2$ ) implies that there exists a constant  $C > 0$ , depending only on  $K_6, \theta_1, \theta_2$ , such that for all  $t \in \mathbb{R}, x_1, x_2 \in \mathbb{R}^{d_1}$  and  $y_1, y_2 \in \mathbb{R}^{d_2}$

$$\begin{aligned} & |f(t, x_1, y_1) - f(t, x_2, y_2)| \\ & \leq C(1 + |x_1|^{\theta_1} + |x_2|^{\theta_1} + |y_1|^{\theta_2} + |y_2|^{\theta_2})(|x_1 - x_2| + |y_1 - y_2|). \end{aligned}$$

(iii) If  $f$  and  $\sigma$  satisfy ( $\mathbf{A}_f$ ) and ( $\mathbf{A}_\sigma$ ) respectively then for all  $t \in \mathbb{R}$  and  $x \in \mathbb{R}^{d_1}$

$$\bar{f}(x) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_t^{t+T} \hat{f}(s, x) ds, \quad \bar{\sigma}(x) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_t^{t+T} \sigma(s, x) ds.$$

(iv) Assume that ( $\mathbf{H}_y^1$ ), ( $\mathbf{H}_y^3$ ) and ( $\mathbf{H}_x^1$ )-( $\mathbf{H}_x^3$ ) hold. Let  $0 < \varepsilon \leq \sqrt[2\alpha]{\eta'/K_5} \wedge 1$ . Then for any  $(x, y) \in \mathbb{R}^{d_1+d_2}$  there exists a unique solution  $(X_t^\varepsilon(x), Y_t^\varepsilon(y))$  to (1.3) satisfying  $(X_0^\varepsilon(x), Y_0^\varepsilon(y)) = (x, y)$ ; see Lemma A.1 for details. Moreover, if  $g \in C_b(\mathbb{R}^{d_1+d_2})$  then for any  $x \in \mathbb{R}^{d_1}$  there exists a unique solution  $\bar{X}_t(x)$  to (1.4) with  $\bar{X}_0(x) = x$ ; see Remark 4.2.

(v) In this paper, we focus on the asymptotic dynamics of the multi-scale system (1.3) when  $\varepsilon$  goes to zero. There exists  $\varepsilon_0 > 0$  such that our results in this paper hold for any  $0 < \varepsilon \leq \varepsilon_0$ . Therefore, we state our results for all  $0 < \varepsilon \leq 1$  in this section for brevity.

Now we establish the first Bogolyubov theorem for (1.3).

**Theorem 2.3.** *Assume that ( $\mathbf{H}_x^1$ )-( $\mathbf{H}_x^3$ ), ( $\mathbf{H}_y^1$ )-( $\mathbf{H}_y^5$ ), ( $\mathbf{A}_f$ ) and ( $\mathbf{A}_\sigma$ ) hold. Then we have*

$$\lim_{\varepsilon \rightarrow 0} \sup_{t \in [0, T]} \mathbb{E} |X_t^\varepsilon(x) - \bar{X}_t(x)|^2 = 0.$$

**2.2. Normal deviation.** Let  $f$  and  $\sigma$  be independent of time  $t$ . Without loss of generality, assume that  $\alpha = \frac{1}{2}$ . Then we can consider the following system

$$\begin{cases} dX_t^\varepsilon = f(X_t^\varepsilon, Y_t^\varepsilon) dt + \sigma(X_t^\varepsilon) dW_t^1 \\ dY_t^\varepsilon = \left( \frac{1}{\varepsilon} B(X_t^\varepsilon, Y_t^\varepsilon) + \frac{1}{\varepsilon^\beta} b(X_t^\varepsilon, Y_t^\varepsilon) \right) dt + \varepsilon^{-\frac{1}{2}} g(X_t^\varepsilon, Y_t^\varepsilon) dW_t^2, \end{cases} \quad (2.2)$$

where  $\beta < 1$ . To obtain the normal deviation, we need the following condition:

( $\mathbf{H}_x^4$ ) (i) There exist  $M > 0$  and  $\theta_2 > 1$  such that for all  $x_1, x_2 \in \mathbb{R}^{d_1}$  and  $y \in \mathbb{R}^{d_2}$

$$\langle f(x_1, y) - f(x_2, y), x_1 - x_2 \rangle \leq M(1 + |y|^{\theta_2}) |x_1 - x_2|^2.$$

(ii) There exists  $C > 0$  and  $\theta_2 > 1$  such that for all  $x \in \mathbb{R}^{d_1}$  and  $y_1, y_2 \in \mathbb{R}^{d_2}$

$$|f(x, y_1) - f(x, y_2)| \leq C(1 + |y_1|^{\theta_2} + |y_2|^{\theta_2}) |y_1 - y_2|.$$

**Remark 2.4.** There is a work on the optimal strong convergence rate for monotone SDEs in [43]. Our condition  $(\mathbf{H}_x^4)$ –(i), however, is more general than the condition [43, (2.1)].

The optimal strong convergence rate of the first Bogolyubov theorem for (2.2) we prove in this paper is contained in the following theorem.

**Theorem 2.5.** *Assume that  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^4)$  and  $(\mathbf{H}_y^1)$ – $(\mathbf{H}_y^5)$  hold. Then there exists a constant  $C_T > 0$  such that*

$$\mathbb{E} \left( \sup_{t \in [0, T]} |X_t^\varepsilon(x) - \bar{X}_t(x)|^2 \right) \leq C_T \varepsilon. \quad (2.3)$$

For simplicity, set  $X_t^\varepsilon := X_t^\varepsilon(x)$ ,  $\bar{X}_t := \bar{X}_t(x)$  for all  $x \in \mathbb{R}^{d_1}$  in the following. Define  $Z_t^\varepsilon := \frac{X_t^\varepsilon - \bar{X}_t}{\sqrt{\varepsilon}}$ . It is clear that  $Z_t^\varepsilon$  solves

$$dZ_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}} (f(X_t^\varepsilon, Y_t^\varepsilon) - \bar{f}(\bar{X}_t)) dt + \frac{1}{\sqrt{\varepsilon}} (\sigma(X_t^\varepsilon) - \sigma(\bar{X}_t)) dW_t^1, \quad Z_0^\varepsilon = 0 \in \mathbb{R}^{d_1}.$$

If  $\sigma(x) \equiv \sigma$  is a constant, then we can show that, as  $\varepsilon \rightarrow 0$ ,  $Z_t^\varepsilon$  converges weakly to  $\bar{Z}_t$ , which is the solution of (1.7) with  $\bar{Z}_0 = 0$ .

**Theorem 2.6.** *Assume that  $\sigma$  is a constant. Furthermore, suppose that  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^4)$  and  $(\mathbf{H}_y^1)$ – $(\mathbf{H}_y^5)$  hold. Then for any  $\varphi \in C_b^\infty(\mathbb{R}^{d_1})$  we have*

$$\lim_{\varepsilon \rightarrow 0} \sup_{t \in [0, T]} |\mathbb{E} \varphi(Z_t^\varepsilon) - \mathbb{E} \varphi(\bar{Z}_t)| = 0.$$

**2.3. Global averaging principle.** Now we investigate the convergence of measure attractors for multi-scale SDEs (1.3). As mentioned in the Introduction,  $P_\varepsilon^*$  defined by (1.10) is time inhomogeneous for any fixed  $0 < \varepsilon \leq 1$ . Therefore, we employ the classical method called the method of skew product, which has been widely used in studying non-autonomous problems arising from deterministic differential equations and dynamical systems, to analyze its pullback attractor; see Section 6.1 for detailed definitions of cocycle, skew product flow, attractors, and pullback attractors.

More precisely, for any  $0 < \varepsilon \leq 1$  we characterize  $P_\varepsilon^*$  as a cocycle over some base space. Indeed, let

$$v := (x, y)^T, \quad F_\varepsilon(t, v) := (f(\varepsilon^{-\gamma}t, x, y), \varepsilon^{-2\alpha}B(x, y) + \varepsilon^{-\beta}b(x, y))^T$$

and

$$G_\varepsilon(t, v) := (\sigma(\varepsilon^{-\gamma}t, x), \varepsilon^{-\alpha}g(x, y))^T, \quad W := (W^1, W^2)^T$$

for all  $0 < \varepsilon \leq 1$  and  $(t, x, y) \in \mathbb{R}^{1+d_1+d_2}$ . Then (1.3) can be written as

$$dV_t^\varepsilon = F_\varepsilon(t, V_t^\varepsilon)dt + G_\varepsilon(t, V_t^\varepsilon)dW_t. \quad (2.4)$$

Fix  $0 < \varepsilon \leq 1$ . Let  $\mathbb{F}_\varepsilon := (F_\varepsilon, G_\varepsilon)$ , and

$$\mathcal{H}(\mathbb{F}_\varepsilon) := \overline{\{\mathbb{F}_\varepsilon^\tau : \tau \in \mathbb{R}\}}$$

with the closure being taken under the metric  $d$  given by (2.5) below, where  $\mathbb{F}^\tau$  is the  $\tau$ -translation of  $\mathbb{F}$  for each  $\mathbb{F} : \mathbb{R} \times \mathbb{R}^{d_1+d_2} \rightarrow \mathbb{R}^{d_1+d_2}$ , i.e.  $\mathbb{F}^\tau(t, v) := \mathbb{F}(t + \tau, v)$ , for all  $(t, v) \in \mathbb{R}^{1+d_1+d_2}$ . For any  $\mathbb{F}_1, \mathbb{F}_2 : \mathbb{R} \times \mathbb{R}^{d_1+d_2} \rightarrow \mathbb{R}^{d_1+d_2}$ , let

$$d(\mathbb{F}_1, \mathbb{F}_2) := \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{d_n(\mathbb{F}_1, \mathbb{F}_2)}{1 + d_n(\mathbb{F}_1, \mathbb{F}_2)}, \quad (2.5)$$

where  $d_n(\mathbb{F}_1, \mathbb{F}_2) := \sup_{|t| \leq n, |v| \leq n} |\mathbb{F}_1(t, v) - \mathbb{F}_2(t, v)|$ .

**Remark 2.7.** Fix  $0 < \varepsilon \leq 1$ . We note that for any  $\tilde{\mathbb{F}}_\varepsilon := (\tilde{F}_\varepsilon, \tilde{G}_\varepsilon) \in \mathcal{H}(\mathbb{F}_\varepsilon)$  there exists  $\{t_n\} \subset \mathbb{R}$  such that for all  $l, r > 0$

$$\lim_{n \rightarrow \infty} \sup_{|t| \leq l, |v| \leq r} |F_\varepsilon(t + t_n, v) - \tilde{F}_\varepsilon(t, v)| = 0, \quad \lim_{n \rightarrow \infty} \sup_{|t| \leq l, |v| \leq r} |G_\varepsilon(t + t_n, v) - \tilde{G}_\varepsilon(t, v)|_{HS} = 0.$$

Therefore, if  $(F_\varepsilon, G_\varepsilon)$  satisfies  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$  and  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^3)$  then  $\tilde{\mathbb{F}}_\varepsilon := (\tilde{F}_\varepsilon, \tilde{G}_\varepsilon)$  does so for any  $\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)$ . Furthermore, let  $\sigma_\tau \mathbb{F} := \mathbb{F}^\tau$ ,  $\tau \in \mathbb{R}$ . Then  $(\mathcal{H}(\mathbb{F}_\varepsilon), \mathbb{R}, \sigma)$  is a shift dynamical system (also called Bebutov shift flow); see Definition 6.3.

Assume that  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$  and  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^3)$  hold. Fix  $0 < \varepsilon \leq 1$ . We aim to show that  $P_\varepsilon^*$  is a cocycle over the base space  $(\mathcal{H}(\mathbb{F}_\varepsilon), \mathbb{R}, \sigma)$ . It follows from Remark 2.7 that for each  $\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)$  and  $v \in \mathbb{R}^{d_1+d_2}$ , there exists a unique solution  $V_{\tilde{\mathbb{F}}_\varepsilon}(t, s, v)$  of (2.4) by replacing  $\mathbb{F}_\varepsilon$  with  $\tilde{\mathbb{F}}_\varepsilon$ , i.e.

$$dV_{\tilde{\mathbb{F}}_\varepsilon}(t, s, v) = \tilde{F}_\varepsilon(t, V_{\tilde{\mathbb{F}}_\varepsilon}(t, s, v))dt + \tilde{G}_\varepsilon(t, V_{\tilde{\mathbb{F}}_\varepsilon}(t, s, v))dW_t, \quad V_{\tilde{\mathbb{F}}_\varepsilon}(s, s, v) = v.$$

Here we explicitly denote the solution  $V_{\tilde{\mathbb{F}}_\varepsilon}(t, s, v)$  with a subscript  $\tilde{\mathbb{F}}_\varepsilon$  to indicate its dependence on  $\tilde{\mathbb{F}}_\varepsilon$ . Similarly, we express the dependence of the associated Markov operators on  $\tilde{\mathbb{F}}_\varepsilon$  by writing

$$P_\varepsilon^*(t, s, \tilde{\mathbb{F}}_\varepsilon, m)(D) := \int_{\mathbb{R}^{d_1+d_2}} \mathbb{P} \circ \left( V_{\tilde{\mathbb{F}}_\varepsilon}(t, s, v) \right)^{-1}(D) m(dv)$$

for all  $D \in \mathcal{B}(\mathbb{R}^{d_1+d_2})$  and  $m \in \mathcal{P}(\mathbb{R}^{d_1+d_2})$ . Set

$$P_\varepsilon^*(t, \tilde{\mathbb{F}}_\varepsilon, m)(D) := P_\varepsilon^*(0, t, \tilde{\mathbb{F}}_\varepsilon, m)(D).$$

Then we show that for any  $0 < \varepsilon \leq 1$   $P_\varepsilon^*$  is a cocycle over the base space  $(\mathcal{H}(\mathbb{F}_\varepsilon), \mathbb{R}, \sigma)$ , and

$$\Phi(t, \tilde{\mathbb{F}}_\varepsilon, m) := \left( \sigma_t \tilde{\mathbb{F}}_\varepsilon, P_\varepsilon^*(t, \sigma_t \tilde{\mathbb{F}}_\varepsilon, m) \right)$$

is the homogeneous Markov semi-flow in the extended phase space; see Lemma 6.17. We call  $(P_\varepsilon^*, \sigma)$  a skew product flow. Furthermore, we consider the existence of pullback attractors for  $(P_\varepsilon^*, \sigma)$  and the convergence of the pullback attractors when the time scale goes to zero; see Theorem 2.8 below for more details. To this end, we need the following dissipativity condition:

$(\mathbf{H}_x^5)$  There exist  $\lambda_1, K_4, K_5 > 0$  such that for any  $(t, x, y) \in \mathbb{R}^{1+d_1+d_2}$

$$2\langle f(t, x, y), x \rangle + |\sigma(t, x)|_{HS}^2 \leq -\lambda_1|x|^2 + K_5|y|^\theta + K_4,$$

where  $\theta$  is as in  $(\mathbf{H}_y^1)$  and  $(\mathbf{H}_x^1)$ .

For any  $v := (x, y)^T \in \mathbb{R}^{d_1+d_2}$ , let  $\pi_1(v) := x$  and  $\pi_2(v) := y$ . Define

$$\mathcal{P}_{2,\theta}(\mathbb{R}^{d_1+d_2}) := \left\{ m \in \mathcal{P}_2(\mathbb{R}^{d_1+d_2}) : \int_{\mathbb{R}^{d_2}} |y|^\theta m \circ \pi_2^{-1}(dy) < \infty \right\},$$

equipped with the following bounded Lipschitz distance (also called Fortet-Mourier distance)

$$d_{BL}(m_1, m_2) := \sup \left\{ \left| \int_{\mathbb{R}^{d_1+d_2}} f dm_1 - \int_{\mathbb{R}^{d_1+d_2}} f dm_2 \right| : \|f\|_{BL} \leq 1 \right\}$$

for all  $m_1, m_2 \in \mathcal{P}_{2,\theta}(\mathbb{R}^{d_1+d_2})$ , where  $\|f\|_{BL} := Lip(f) + \|f\|_\infty$  for all Lipschitz continuous  $f \in C_b(\mathbb{R}^{d_1+d_2})$ . It can be verified that  $(\mathcal{P}_{2,\theta}(\mathbb{R}^{d_1+d_2}), d_{BL})$  is a Polish space. We say that  $D \subset \mathcal{P}_{2,\theta}(\mathbb{R}^{d_1+d_2})$  is *bounded* if there exist  $r_1, r_2 > 0$  such that for all  $m \in D$

$$\int_{\mathbb{R}^{d_1+d_2}} |v|^2 m(dv) \leq r_1, \quad \int_{\mathbb{R}^{d_2}} |y|^\theta m \circ \pi_2^{-1}(dy) \leq r_2.$$

**Theorem 2.8.** *Consider equation (1.3). Assume that  $(\mathbf{H}_y^1)$ – $(\mathbf{H}_y^5)$ ,  $(\mathbf{H}_x^2)$ – $(\mathbf{H}_x^3)$ ,  $(\mathbf{H}_x^5)$ ,  $(\mathbf{A}_f)$  and  $(\mathbf{A}_\sigma)$  hold. If  $\mathcal{H}(\mathbb{F}_\varepsilon)$  is compact for any  $0 < \varepsilon \leq 1$ , then the following conclusions hold:*

- (i) *For any  $0 < \varepsilon \leq 1$  (1.3) is associated with a skew product flow  $(\sigma, P_\varepsilon^*)$  on  $(\mathcal{P}_{2,\theta}(\mathbb{R}^{d_1+d_2}), d_{BL})$ , and  $(\sigma, P_\varepsilon^*)$  admits a pullback attractor  $\mathcal{A}^\varepsilon$  with component subsets  $\mathcal{A}_{\tilde{\mathbb{F}}_\varepsilon}^\varepsilon$ ,  $\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)$ ;*
- (ii) *The averaged equation has a global attractor  $\bar{\mathcal{A}}$ ;*
- (iii) *Furthermore, for all  $\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)$*

$$\lim_{\varepsilon \rightarrow 0} \text{dist}_{\mathcal{P}(\mathbb{R}^{d_1})} \left( \Pi_1(\mathcal{A}_{\tilde{\mathbb{F}}_\varepsilon}^\varepsilon), \bar{\mathcal{A}} \right) = 0,$$

where  $\bar{\mathcal{A}}$  is the global attractor of  $\bar{P}^*$  and  $\Pi_1 m := m \circ \pi_1^{-1}$  for all  $m \in \mathcal{P}(\mathbb{R}^{d_1+d_2})$ .

**2.4. The second Bogolyubov theorem.** In this subsection, we will consider the convergence of the recurrent solutions for (1.3). More precisely, we show that, as the time scale  $\varepsilon$  goes to 0, the slow components of the quasi-periodic solutions weakly converge to the stationary solution of the averaged equation. To this end, we need the following conditions:

( $\mathbf{H}_x^6$ ) There exist constants  $\lambda_1, \lambda_2 > 0$  such that for all  $t \in \mathbb{R}$  and  $x_1, x_2 \in \mathbb{R}^{d_1}$ ,  $y_1, y_2 \in \mathbb{R}^{d_2}$

$$\begin{aligned} & 2\langle f(t, x_1, y_1) - f(t, x_2, y_2), x_1 - x_2 \rangle + |\sigma(t, x_1) - \sigma(t, x_2)|_{HS}^2 \\ & \leq -\lambda_1 |x_1 - x_2|^2 + \lambda_2 |y_1 - y_2|^2. \end{aligned}$$

( $\mathbf{H}_y^6$ ) There exists  $L_b > 0$  such that for all  $t \in \mathbb{R}$  and  $y_1, y_2 \in \mathbb{R}^{d_2}$

$$|b(x_1, y_1) - b(x_2, y_2)| \leq L_b (|x_1 - x_2| + |y_1 - y_2|).$$

**Remark 2.9.** It can be verified that ( $\mathbf{H}_x^3$ ) and ( $\mathbf{H}_x^6$ ) imply ( $\mathbf{H}_x^5$ ).

First, we recall the definition of quasi-periodic functions. Let  $\mathcal{X}$  be a Polish space.

**Definition 2.10.** A function  $\varphi \in C(\mathbb{R}, \mathcal{X})$  is called *quasi-periodic with the spectrum of frequencies*  $\nu_1, \dots, \nu_k$  if it satisfies the following conditions:

- (i) the numbers  $\nu_1, \dots, \nu_k$  are rationally independent;
- (ii) there exists a continuous function  $\phi : \mathbb{R}^k \rightarrow \mathcal{X}$  such that for all  $(t_1, \dots, t_k) \in \mathbb{R}^k$

$$\phi(t_1 + 2\pi, \dots, t_k + 2\pi) = \phi(t_1, \dots, t_k);$$

- (iii)  $\varphi(t) = \phi(\nu_1 t, \dots, \nu_k t)$  for  $t \in \mathbb{R}$ .

**Definition 2.11.** We say a  $\mathcal{X}$ -valued continuous stochastic process  $X_t, t \in \mathbb{R}$  is *quasi-periodic in distribution*, if the mapping  $\mathcal{L}(X_\cdot) : \mathbb{R} \rightarrow \mathcal{P}(\mathcal{X})$  is quasi-periodic.

Now we can formulate our result, which is called the second Bogolyubov theorem.

**Theorem 2.12.** *Let  $B(x, y) \equiv B(y)$  and  $g(x, y) \equiv g(y)$  for all  $(x, y) \in \mathbb{R}^{d_1+d_2}$ . Assume that  $\beta < \alpha$  or  $\beta = \alpha$  and  $\lambda_1 > \frac{L_b^2}{\eta}$ . Furthermore, suppose that ( $\mathbf{H}_y^1$ )-( $\mathbf{H}_y^6$ ), ( $\mathbf{H}_x^2$ )-( $\mathbf{H}_x^3$ ) and ( $\mathbf{H}_x^6$ ) hold. If  $f$  and  $g$  are quasi-periodic, then for any  $0 < \varepsilon \leq 1$  there exists a unique solution  $V_t^\varepsilon := (X_t^\varepsilon, Y_t^\varepsilon), t \in \mathbb{R}$ , of (1.3), which is quasi-periodic in distribution, and*

$$\limsup_{\varepsilon \rightarrow 0} \sup_{t \in \mathbb{R}} d_{BL}(\mathcal{L}(X_t^\varepsilon), \mathcal{L}(\bar{X}_t)) = 0, \quad (2.6)$$

where  $\bar{X}$  is the unique stationary solution of the averaged equation (1.4).

**Remark 2.13.** (i) For brevity, we just illustrate the case of quasi-periodic solutions in this paper. Indeed, our method applies to more general compact recurrent solutions.

(ii) Although there is a more general result on the second Bogolyubov theorem in [9], which can cover unbounded recurrent solutions such as Levitan almost periodic solutions, the proof presented here is more concise than [9]. Furthermore, the system (1.3) is more general, and the result (2.6) is stronger than [9, Theorem 4.7]. Since we employ the global averaging principle to establish the second Bogolyubov theorem, it is required that the hull  $\mathcal{H}(\mathbb{F}_\varepsilon)$  is compact. It is worth noting that  $\mathcal{H}(\mathbb{F}_\varepsilon)$  is compact provided  $\mathbb{F}_\varepsilon$  is Birkhoff recurrent, which includes periodic, quasi-periodic, almost periodic, almost automorphic, and Birkhoff recurrent functions.

(iii) It is well-known that the uniform attractor (see [9, Definition 5.6]) is a pullback attractor, but not vice versa. Compared to [9] and [10], we consider the more general pullback attractor instead of the uniform attractor.

**2.5. Examples.** To illustrate our results, we will present two examples in this subsection. For simplicity, we just consider the one-dimensional case, but one can easily extend this to the multi-dimensional case. Let  $W_t^1, t \in \mathbb{R}$  and  $W_t^2, t \in \mathbb{R}$  be independent two-sided standard Brownian motions.

**Example 2.14.** Consider the following slow-fast SDEs:

$$\begin{cases} dX_t^\varepsilon = (X_t^\varepsilon - (X_t^\varepsilon)^3 + (Y_t^\varepsilon)^2 \sin X_t^\varepsilon + Y_t^\varepsilon) dt + dW_t^1 \\ dY_t^\varepsilon = \frac{1}{\varepsilon} (-(\sin X_t^\varepsilon)^2 (Y_t^\varepsilon)^5 - (Y_t^\varepsilon)^3 - Y_t^\varepsilon) dt + \frac{1}{\sqrt{\varepsilon}} dW_t^2. \end{cases}$$

We define  $f(x, y) := x - x^3 + y^2 \sin x + y$ ,  $B(x, y) := -(\sin x)^2 y^5 - y^3 - y$ ,  $\forall (x, y) \in \mathbb{R}^2$ . It can be verified that  $f$  and  $B$  satisfy  $(\mathbf{H}_y^1)$ – $(\mathbf{H}_y^5)$  and  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^4)$ . Then by Theorems 2.5 and 2.6, one sees that there exists a constant  $C > 0$  such that

$$\mathbb{E} \left( \sup_{t \in [0, T]} |X_t^\varepsilon - \bar{X}_t|^2 \right) \leq C\varepsilon,$$

where  $\bar{X}_t$  is the solution to the corresponding equation, and that

$$Z_t^\varepsilon := \frac{X_t^\varepsilon - \bar{X}_t}{\sqrt{\varepsilon}}$$

weakly converges to an Ornstein-Uhlenbeck type process, as  $\varepsilon$  goes to zero.

**Example 2.15.** Consider the following multi-scale SDEs:

$$\begin{cases} dX_t^\varepsilon = \left[ -a_1 X_t^\varepsilon - (X_t^\varepsilon)^3 + (a_2 Y_t^\varepsilon + a_3 (Y_t^\varepsilon)^3) \left( \cos(t/\sqrt{\varepsilon}) + \sin(\sqrt{2}t/\sqrt{\varepsilon}) \right) \right. \\ \quad \left. - a_4 X_t^\varepsilon (Y_t^\varepsilon)^4 (\sin(t/\sqrt{\varepsilon}))^2 \right] dt + dW_t^1 \\ dY_t^\varepsilon = \left[ \frac{1}{\varepsilon} (-(Y_t^\varepsilon)^3 - Y_t^\varepsilon) + \frac{1}{\sqrt[3]{\varepsilon}} (X_t^\varepsilon + Y_t^\varepsilon) \right] dt + \frac{1}{\sqrt{\varepsilon}} dW_t^2, \end{cases} \quad (2.7)$$

where  $a_1 > 0$ ,  $a_2, a_3 \in \mathbb{R}$  and  $a_4 \geq 0$ . For all  $(t, x, y) \in \mathbb{R}^3$ , define

$$f(t, x, y) := -a_1 x - x^3 + (a_2 y + a_3 y^3)(\cos t + \sin \sqrt{2}t) - a_4 x y^4 (\sin t)^2,$$

$B(y) := -y^3 - y$  and  $b(x, y) := x + y$ . We note that  $f$ ,  $B$  and  $b$  satisfy conditions  $(\mathbf{H}_y^1)$ – $(\mathbf{H}_y^5)$ ,  $(\mathbf{H}_x^2)$ – $(\mathbf{H}_x^3)$  and  $(\mathbf{H}_x^5)$ . Set

$$\mathbb{F}_\varepsilon(t, x, y) := \left( f(\varepsilon^{-1/2}t, x, y), \varepsilon^{-1}B(y) + \varepsilon^{-1/3}b(x, y) \right)^T, \quad \forall (t, x, y) \in \mathbb{R}^3.$$

Then by Theorem 2.8, we have the following conclusions:

- (i) For any  $0 < \varepsilon \leq 1$  (2.7) is associated to a skew product flow  $(\sigma, P_\varepsilon^*)$  on  $(\mathcal{P}_{2,6}(\mathbb{R}^2), d_{BL})$ , and  $(\sigma, P_\varepsilon^*)$  admits a pullback attractor  $\mathcal{A}^\varepsilon$  with component subsets  $\mathcal{A}_{\mathbb{F}_\varepsilon}$ ,  $\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)$ ;
- (ii) The corresponding averaged equation has a global attractor  $\bar{\mathcal{A}}$ ;
- (iii) Furthermore, for all  $\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)$

$$\lim_{\varepsilon \rightarrow 0} \text{dist}_{\mathcal{P}(\mathbb{R}^{d_1})} \left( \Pi_1(\mathcal{A}_{\tilde{\mathbb{F}}_\varepsilon}), \bar{\mathcal{A}} \right) = 0.$$

Furthermore, assume that  $a_4 = 0$ . Then it can be verified that  $(\mathbf{H}_x^6)$  also holds. Recall that  $f$  is quasi-periodic. Then in view of Theorem 2.12, for all  $0 < \varepsilon \leq 1$  there is a unique solution  $(X_t^\varepsilon, Y_t^\varepsilon)$ ,  $t \in \mathbb{R}$ , which is quasi-periodic in distribution, and

$$\limsup_{\varepsilon \rightarrow 0} \sup_{t \in \mathbb{R}} d_{BL}(\mathcal{L}(X_t^\varepsilon), \mathcal{L}(\bar{X}_t)) = 0,$$

where  $\bar{X}$  is the stationary solution to the following averaged equation:

$$d\bar{X}_t = (-a_1 \bar{X}_t - (\bar{X}_t)^3) dt + dW_t^1.$$

## 3. FROZEN EQUATION AND POISSON EQUATION

Consider the following so-called frozen equation

$$dY_s^{x,y} = B(x, Y_s^{x,y})ds + g(x, Y_s^{x,y})dW_s^2, \quad Y_0^{x,y} = y \in \mathbb{R}^{d_2},$$

where  $x \in \mathbb{R}^{d_1}$  is a frozen parameter.

**Lemma 3.1.** *Assume that  $(\mathbf{H}_y^1)$  and  $(\mathbf{H}_y^3)$  hold. For any  $\xi \in \mathcal{L}^2(\Omega, \mathcal{F}_s, \mathbb{P}; \mathbb{R}^{d_2})$ ,  $t \geq s$  and  $x \in \mathbb{R}^{d_1}$ , let  $Y_{s,t}^x(\xi), t \geq s$  be the unique solution to (2.1). Then for any  $\xi_1, \xi_2 \in \mathcal{L}^2(\Omega, \mathcal{F}_s, \mathbb{P}; \mathbb{R}^{d_2})$  we have*

$$\mathbb{E} |Y_{s,t}^x(\xi_1) - Y_{s,t}^x(\xi_2)|^2 \leq \mathbb{E} |\xi_1 - \xi_2|^2 e^{-\eta(t-s)}. \quad (3.1)$$

Moreover, if  $\xi \in \mathcal{L}^{2p}(\Omega, \mathcal{F}_s, \mathbb{P}; \mathbb{R}^{d_2})$  and  $g \in C_b(\mathbb{R}^{d_1+d_2})$  for any  $p \geq 1$ , then we have

$$\mathbb{E} |Y_{s,t}^x(\xi)|^{2p} \leq \mathbb{E} |\xi|^{2p} e^{-\frac{\eta p}{2}(t-s)} + M, \quad (3.2)$$

and there exists a constant  $C > 0$  such that for all  $x_1, x_2 \in \mathbb{R}^{d_1}$

$$\mathbb{E} |Y_{s,t}^{x_1}(y) - Y_{s,t}^{x_2}(y)|^2 \leq C(1 + |y|^{2\kappa_2}) |x_1 - x_2|^2.$$

Here  $M$  depends on  $p, K_1, \|g\|_\infty$ .

*Proof.* (i) Employing Itô's formula and  $(\mathbf{H}_y^3)$ -(i), we have

$$\begin{aligned} & \mathbb{E} |Y_{s,t}^x(\xi_1) - Y_{s,t}^x(\xi_2)|^2 \\ &= \mathbb{E} |\xi_1 - \xi_2|^2 + \mathbb{E} \int_0^t 2 \langle B(x, Y_{s,r}^x(\xi_1)) - B(x, Y_{s,r}^x(\xi_2)), Y_{s,r}^x(\xi_1) - Y_{s,r}^x(\xi_2) \rangle dr \\ & \quad + \mathbb{E} \int_0^t |g(x, Y_{s,r}^x(\xi_1)) - g(x, Y_{s,r}^x(\xi_2))|_{HS}^2 dr \\ & \leq \mathbb{E} |\xi_1 - \xi_2|^2 + \mathbb{E} \int_0^t -\eta |Y_{s,r}^x(\xi_1) - Y_{s,r}^x(\xi_2)|^2 dr, \end{aligned}$$

which by Gronwall's inequality implies that

$$\mathbb{E} |Y_{s,t}^x(\xi_1) - Y_{s,t}^x(\xi_2)|^2 \leq \mathbb{E} |\xi_1 - \xi_2|^2 e^{-\eta(t-s)}.$$

(ii) By Itô's formula,  $(\mathbf{H}_y^1)$  and Young's inequality, one sees that for any  $p \geq 1$

$$\begin{aligned} & \mathbb{E} |Y_{s,t}^x(\xi)|^{2p} \\ &= \mathbb{E} |\xi|^{2p} + p \mathbb{E} \int_0^t |Y_{s,r}^x(\xi)|^{2p-2} \left( 2 \langle B(x, Y_{s,r}^x(\xi)), Y_{s,r}^x(\xi) \rangle + |g(x, Y_{s,r}^x(\xi))|_{HS}^2 \right) dr \\ & \quad + 2p(p-1) \mathbb{E} \int_0^t |Y_{s,r}^x(\xi)|^{2p-4} |g^T(x, Y_{s,r}^x(\xi)) Y_{s,r}^x(\xi)|^2 dr \\ & \leq \mathbb{E} |\xi|^{2p} + \mathbb{E} \int_0^t p [(-\eta |Y_{s,r}^x(\xi)|^2 + K_1) |Y_{s,r}^x(\xi)|^{2p-2} + 2(p-1) \|g\|_\infty |Y_{s,r}^x(\xi)|^{2p-2}] dr \\ & \leq \mathbb{E} |\xi|^{2p} + \mathbb{E} \int_0^t -\frac{\eta p}{2} |Y_{s,r}^x(\xi)|^{2p} + p^{-p}(p-1)^{p-1} (\eta/2)^{1-p} (K_1 p + 2(p-1) \|g\|_\infty)^p dr. \end{aligned}$$

Then we have

$$\mathbb{E} |Y_{s,t}^x(\xi)|^{2p} \leq \mathbb{E} |\xi|^{2p} e^{-\frac{\eta p}{2}(t-s)} + 2p^{-p}(p-1)^{p-1} (\eta/2)^{1-p} (K_1 p + 2(p-1) \|g\|_\infty)^p (\eta p)^{-1}.$$

(iii) In view of Itô's formula, we get

$$\begin{aligned} & \mathbb{E} |Y_{s,t}^{x_1}(y) - Y_{s,t}^{x_2}(y)|^2 \\ &= \mathbb{E} \int_s^t 2 \langle B(x_1, Y_{s,r}^{x_1}(y)) - B(x_2, Y_{s,r}^{x_2}(y)), Y_{s,r}^{x_1}(y) - Y_{s,r}^{x_2}(y) \rangle dr \end{aligned}$$

$$\begin{aligned}
& + \mathbb{E} \int_s^t |g(x_1, Y_{s,r}^{x_1}(y)) - g(x_2, Y_{s,r}^{x_2}(y))|_{HS}^2 dr \\
= & \mathbb{E} \int_s^t 2 \langle B(x_1, Y_{s,r}^{x_1}(y)) - B(x_1, Y_{s,r}^{x_2}(y)), Y_{s,r}^{x_1}(y) - Y_{s,r}^{x_2}(y) \rangle dr \\
& + \mathbb{E} \int_s^t 2 \langle B(x_1, Y_{s,r}^{x_2}(y)) - B(x_2, Y_{s,r}^{x_2}(y)), Y_{s,r}^{x_1}(y) - Y_{s,r}^{x_2}(y) \rangle dr \\
& + \mathbb{E} \int_s^t \left( |g(x_1, Y_{s,r}^{x_1}(y)) - g(x_1, Y_{s,r}^{x_2}(y))|_{HS}^2 + |g(x_1, Y_{s,r}^{x_2}(y)) - g(x_2, Y_{s,r}^{x_2}(y))|_{HS}^2 \right) dr \\
& + \mathbb{E} \int_s^t 2 \langle g(x_1, Y_{s,r}^{x_1}(y)) - g(x_1, Y_{s,r}^{x_2}(y)), g(x_1, Y_{s,r}^{x_2}(y)) - g(x_2, Y_{s,r}^{x_2}(y)) \rangle_{HS} dr.
\end{aligned}$$

Then thanks to  $(\mathbf{H}_y^3)$ , Young's inequality and (3.2), we obtain

$$\begin{aligned}
& \mathbb{E} |Y_{s,t}^{x_1}(y) - Y_{s,t}^{x_2}(y)|^2 \\
& \leq \mathbb{E} \int_s^t \left( -\eta |Y_{s,r}^{x_1}(y) - Y_{s,r}^{x_2}(y)|^2 + L_g^2 |x_1 - x_2|^2 + 2L_g^2 |Y_{s,r}^{x_1}(y) - Y_{s,r}^{x_2}(y)| |x_1 - x_2| \right. \\
& \quad \left. + 2K_3 (1 + |Y_{s,r}^{x_2}(y)|^{\kappa_2}) |x_1 - x_2| |Y_{s,r}^{x_1}(y) - Y_{s,r}^{x_2}(y)| \right) dr \\
& \leq \mathbb{E} \int_s^t \left( -\frac{\eta}{2} |Y_{s,r}^{x_1}(y) - Y_{s,r}^{x_2}(y)|^2 + \left( \frac{4}{\eta} L_g^4 + L_g^2 \right) |x_1 - x_2|^2 \right) dr \\
& \quad + \int_s^t \frac{C}{\eta} (1 + \mathbb{E} |Y_{s,r}^{x_2}(y)|^{2\kappa_2}) |x_1 - x_2|^2 dr \\
& \leq \mathbb{E} \int_s^t \left( -\frac{\eta}{2} |Y_{s,r}^{x_1}(y) - Y_{s,r}^{x_2}(y)|^2 + \left( \frac{4}{\eta} L_g^4 + L_g^2 \right) |x_1 - x_2|^2 \right) dr \\
& \quad + \int_s^t C \left( 1 + |y|^{2\kappa_2} e^{-\frac{\eta\kappa_2}{2}(r-s)} \right) |x_1 - x_2|^2 dr.
\end{aligned}$$

Therefore, by Gronwall's inequality, we obtain

$$\mathbb{E} |Y_{s,t}^{x_1}(y) - Y_{s,t}^{x_2}(y)|^2 \leq C (1 + |y|^{2\kappa_2}) |x_1 - x_2|^2.$$

□

Combining (3.1) and (3.2), we have the following result.

**Corollary 3.2.** *Assume that  $(\mathbf{H}_y^1)$  and  $(\mathbf{H}_y^3)$  hold. If  $g \in C_b(\mathbb{R}^{d_1+d_2})$ , then for any  $m > 0$*

$$\sup_{x \in \mathbb{R}^{d_1}} \int_{\mathbb{R}^{d_2}} |y|^m \mu^x(dy) < M.$$

Now we show the continuous dependence on the parameter  $x \in \mathbb{R}^{d_1}$  for stationary solutions to equation (2.1).

**Lemma 3.3.** *Assume that  $(\mathbf{H}_y^1)$  and  $(\mathbf{H}_y^3)$  hold. If  $g \in C_b(\mathbb{R}^{d_1+d_2})$ , then for any  $x \in \mathbb{R}^{d_1}$  there exist a unique stationary solution  $Y_t^x$ ,  $t \in \mathbb{R}$ , to (2.1), and a constant  $C > 0$  such that for any  $x_1, x_2 \in \mathbb{R}^{d_1}$*

$$\sup_{t \in \mathbb{R}} \mathbb{E} |Y_t^{x_1} - Y_t^{x_2}|^2 \leq C |x_1 - x_2|^2.$$

*Proof.* It follows from (3.1) and (3.2) that for any  $x \in \mathbb{R}^{d_1}$  there exists a unique stationary solution  $Y_t^x$ ,  $t \in \mathbb{R}$ , to (2.1). By Corollary 3.2, we have for all  $-n \leq t$  and  $x_1, x_2 \in \mathbb{R}^{d_1}$

$$\begin{aligned}
\mathbb{E} |Y_t^{x_1} - Y_t^{x_2}|^2 & \leq 3\mathbb{E} |Y_t^{x_1} - Y_{-n,t}^{x_1}(0)|^2 + 3\mathbb{E} |Y_{-n,t}^{x_1}(0) - Y_{-n,t}^{x_2}(0)|^2 + 3\mathbb{E} |Y_{-n,t}^{x_2}(0) - Y_t^{x_2}|^2 \\
& \leq 6Me^{-\eta(t+n)} + 3C|x_1 - x_2|^2,
\end{aligned}$$

which implies that

$$\mathbb{E}|Y_t^{x_1} - Y_t^{x_2}|^2 \leq C|x_1 - x_2|^2$$

by letting  $n \rightarrow \infty$ , where  $C$  does not depend on  $t$ .  $\square$

Finally, we investigate the well-posedness of the Poisson equation. Consider the following equation

$$\mathfrak{L}_2(x, y)u(h, x, y) = -(\phi(h, x, y) - \bar{\phi}(h, x)), \quad y \in \mathbb{R}^{d_2}, \quad (3.3)$$

where  $(h, x) \in \mathbb{R}^{d+d_1}$  is a parameter,  $\bar{\phi}(h, x) := \int_{\mathbb{R}^{d_2}} \phi(h, x, y) \mu^x(dy)$  and

$$\mathfrak{L}_2 u(h, x, y) := \mathfrak{L}_2(x, y)u(h, x, y) := \langle B(x, y), \partial_y u(h, x, y) \rangle + \sum_{i,j=1}^{d_2} a_{ij} \partial_{y_j y_i}^2 u(h, x, y).$$

Here  $(a_{ij}) = gg^T/2$ . Similarly, for all  $(h, x, y) \in \mathbb{R}^{d+d_1+d_2}$  we define

$$\mathfrak{L}_1^\varepsilon \psi(h, x, y) := \mathfrak{L}_1^\varepsilon(x, y)\psi(h, x, y) := \langle f(\varepsilon^{-\gamma}t, x, y), \partial_x \psi(h, x, y) \rangle + \sum_{i,j=1}^{d_1} A_{ij}^\varepsilon \partial_{x_j x_i}^2 \psi(h, x, y),$$

$$\mathfrak{L}_3 \psi(h, x, y) := \mathfrak{L}_3(x, y)\psi(h, x, y) := \langle b(x, y), \partial_y \psi(h, x, y) \rangle,$$

where  $(A_{ij}^\varepsilon) = \sigma_\varepsilon \sigma_\varepsilon^T / 2$ . When  $f$  and  $\sigma$  are time independent, we let  $(A_{ij}) := \sigma \sigma^T / 2$ , and

$$\mathfrak{L}_1 \psi(h, x, y) := \mathfrak{L}_1(x, y)\psi(h, x, y) := \langle f(x, y), \partial_x \psi(h, x, y) \rangle + \sum_{i,j=1}^{d_1} A_{ij} \partial_{x_j x_i}^2 \psi(h, x, y),$$

$$\mathfrak{L}_x \psi(x, y) := \mathfrak{L}_x \psi(x, y) := \langle \bar{f}(x), \partial_x \psi(x, y) \rangle + \sum_{i,j=1}^{d_1} A_{ij} \partial_{x_j x_i}^2 \psi(x, y).$$

Let us first introduce the condition  $(\mathbf{H}_\phi^i)$ , where  $i = 1, 2$ .

$(\mathbf{H}_\phi^i)$  There exist constants  $C_1 > 0$  and  $m_1, m_2, m_3 \geq 0$  such that for all  $(h, x, y) \in \mathbb{R}^{d+d_1+d_2}$

$$\sum_{0 \leq 2k+j \leq 4} |\partial_x^k \partial_y^j \phi(h, x, y)| + |\partial_h^i \phi(h, x, y)| \leq C_1(1 + |h|^{m_1} + |x|^{m_2} + |y|^{m_3}).$$

Similar to [43, Proposition 4.1] and [11, Theorem 3.1], we have the following lemma about the existence and uniqueness of solutions to (3.3).

**Lemma 3.4.** *Assume that  $(\mathbf{H}_y^1)$ – $(\mathbf{H}_y^5)$  hold. Furthermore, suppose that  $\phi \in C^{i,2,4}(\mathbb{R}^{d+d_1+d_2})$  satisfies  $(\mathbf{H}_\phi^i)$ . Then there exist a unique solution  $u(t, x, \cdot) \in C^2(\mathbb{R}^{d_2})$  to (3.3) and constants  $m'_1, m'_2, m'_3, C > 0$  such that*

$$\begin{aligned} & |u(h, x, y)| + |\partial_y u(h, x, y)| + \sum_{j=1}^i |\partial_h^j u(h, x, y)| \\ & + |\partial_x u(h, x, y)| + |\partial_x^2 u(h, x, y)| \leq C(1 + |h|^{m'_1} + |x|^{m'_2} + |y|^{m'_3}), \end{aligned}$$

and

$$\sum_{j=1}^2 |\partial_x^j \bar{\phi}(h, x)| \leq C \left( 1 + |h|^{m'_1} + |x|^{m'_2} \right).$$

#### 4. THE FIRST BOGOLYUBOV THEOREM

In Section 4.1, we will prove some lemmas, which give the properties for  $\hat{f}, \bar{f}$  and  $\bar{\sigma}$ , moment estimates of solutions to (1.3) and (1.4), and Hölder continuity of the slow variable in (1.3). In Section 4.2, we prove Theorem 2.3.

**4.1. Auxiliary lemmas.** In the following lemma, we show that  $\hat{f}$ ,  $\bar{f}$  and  $\bar{\sigma}$  inherit some properties from  $f$  and  $\sigma$ .

**Lemma 4.1.** *Assume that  $f \in C^{1,2,4}(\mathbb{R}^{1+d_1+d_2})$  and  $\sigma$  satisfy  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^3)$ ,  $(\mathbf{A}_f)$  and  $(\mathbf{A}_\sigma)$ . Furthermore, suppose that  $(\mathbf{H}_y^1)$  and  $(\mathbf{H}_y^3)$  hold. Then the following conclusions hold.*

(i) *If  $g \in C_b(\mathbb{R}^{d_1+d_2})$  then  $\hat{f}$  and  $\bar{f}$  satisfy  $(\mathbf{H}_x^1)$ ,  $\bar{\sigma}$  satisfies  $(\mathbf{H}_x^3)$  and there exists  $C > 0$  such that for all  $t \in \mathbb{R}$  and  $x_1, x_2 \in \mathbb{R}^{d_1}$*

$$\left| \hat{f}(t, x_1) - \hat{f}(t, x_2) \right| + \left| \bar{f}(x_1) - \bar{f}(x_2) \right| \leq C (1 + |x_1|^{\theta_1} + |x_2|^{\theta_1}) |x_1 - x_2|,$$

where  $\theta_1$  is as in  $(\mathbf{H}_x^2)$ .

(ii) *If  $(\mathbf{H}_y^2)$  and  $(\mathbf{H}_y^4)$ – $(\mathbf{H}_y^5)$  hold then  $\hat{f} \in C^2(\mathbb{R}^{d_1})$  and  $\hat{f}$  satisfies  $(\mathbf{H}_x^2)$ .*

*Proof.* (i) We note that

$$\hat{f}(t, x) = \int_{\mathbb{R}^{d_2}} f(t, x, y) \mu^x(dy), \quad \forall (t, x) \in \mathbb{R}^{1+d_1}.$$

In view of  $(\mathbf{H}_x^1)$  and Corollary 3.2, we have for any  $(t, x) \in \mathbb{R}^{1+d_1}$

$$\begin{aligned} \langle \hat{f}(t, x), x \rangle &= \left\langle \int_{\mathbb{R}^{d_2}} f(t, x, y) \mu^x(dy), x \right\rangle \\ &\leq \int_{\mathbb{R}^{d_2}} (K_4(1 + |x|^2) + K_5|y|^\theta) \mu^x(dy) \leq K_4(1 + |x|^2) + C. \end{aligned}$$

Therefore, for any  $t \in \mathbb{R}$  and  $x \in \mathbb{R}^{d_1}$ , one sees that

$$\begin{aligned} \langle \bar{f}(x), x \rangle &= \left\langle \bar{f}(x) - \frac{1}{T} \int_0^T \hat{f}(s, x) ds, x \right\rangle + \left\langle \frac{1}{T} \int_0^T \hat{f}(s, x) ds, x \right\rangle \\ &\leq \left| \bar{f}(x) - \frac{1}{T} \int_0^T \hat{f}(s, x) ds \right| |x| + K_4(1 + |x|^2) + C, \end{aligned}$$

which implies that

$$\langle \bar{f}(x), x \rangle \leq K_4(1 + |x|^2) + C, \quad \forall x \in \mathbb{R}^{d_1}$$

by letting  $T \rightarrow \infty$  and because of Remark 2.2 (iii).

Combing  $(\mathbf{H}_x^2)$ , Corollary 3.2, Hölder's inequality and Lemma 3.3, we have for any  $t \in \mathbb{R}$  and  $x_1, x_2 \in \mathbb{R}^{d_1}$

$$\begin{aligned} &\left| \hat{f}(t, x_1) - \hat{f}(t, x_2) \right| \\ &= \left| \int_{\mathbb{R}^{d_2}} f(t, x_1, y) \mu^{x_1}(dy) - \int_{\mathbb{R}^{d_2}} f(t, x_2, y) \mu^{x_2}(dy) \right| \\ &\leq \int_{\mathbb{R}^{d_2}} |f(t, x_1, y) - f(t, x_2, y)| \mu^{x_1}(dy) + \left| \int_{\mathbb{R}^{d_2}} f(t, x_2, y) (\mu^{x_1}(dy) - \mu^{x_2}(dy)) \right| \\ &\leq C (1 + |x_1|^{\theta_1} + |x_2|^{\theta_1}) |x_1 - x_2| \\ &\quad + C \mathbb{E} \left( (1 + |x_2|^{\theta_1} + |Y_s^{x_1}|^{\theta_2} + |Y_s^{x_2}|^{\theta_2}) |Y_s^{x_1} - Y_s^{x_2}| \right) \\ &\leq C (1 + |x_1|^{\theta_1} + |x_2|^{\theta_1}) |x_1 - x_2|, \end{aligned} \tag{4.1}$$

where  $s \in \mathbb{R}$ ,  $Y_s^{x_1}$  and  $Y_s^{x_2}$  are stationary solutions to (2.1) with  $x_1$  and  $x_2$  respectively replacing  $x$ . Then (4.1) and  $(\mathbf{A}_f)$  imply that for all  $x_1, x_2 \in \mathbb{R}^{d_1}$

$$\left| \bar{f}(x_1) - \bar{f}(x_2) \right| \leq C (1 + |x_1|^{\theta_1} + |x_2|^{\theta_1}) |x_1 - x_2|.$$

It follows from  $(\mathbf{H}_x^3)$  and  $(\mathbf{A}_\sigma)$  that  $\bar{\sigma}$  satisfies  $(\mathbf{H}_x^3)$ .

(ii) By  $(\mathbf{H}_x^2)$  and Lemma 3.4, one sees that there exists  $m_1 > 0$  such that

$$\left| \partial_x^i \hat{f}(t, x) \right| \leq K_6 (1 + |x|^{m_1} + C), \quad i = 0, 1, 2.$$

It follows from  $(\mathbf{H}_x^2)$  that for all  $(t, x) \in \mathbb{R}^{1+d_1}$

$$\left| \partial_t \hat{f}(t, x) \right| \leq \int_{\mathbb{R}^{d_2}} |\partial_t f(t, x, y)| \mu^x(dy) \leq K_6 (1 + |x|^{\theta_1}) + C.$$

□

**Remark 4.2.** Assume that  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$ ,  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^3)$ ,  $(\mathbf{A}_f)$  and  $(\mathbf{A}_\sigma)$  hold. If  $g \in C_b(\mathbb{R}^{d_1+d_2})$ , then it follows from the above lemma and Theorem 3.1.1 in [35] that for any  $\zeta \in \mathcal{L}^2(\Omega, \mathcal{F}_0, \mathbb{P}; \mathbb{R}^{d_1})$  there exists a unique solution  $\bar{X}_t(\zeta)$  to (1.4) satisfying  $\bar{X}_0(\zeta) = \zeta$ .

Now we prove moment estimates for solutions to the slow-fast SDEs (1.3) and the averaged equation (1.4).

**Lemma 4.3.** Assume that  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$  and  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^3)$  hold. Let  $(X_t^\varepsilon(\zeta^\varepsilon), Y_t^\varepsilon(\xi^\varepsilon)), t \geq 0$ , be the solution to (1.3), and  $\bar{X}_t(\zeta), t \geq 0$ , the solution to (1.4). If  $g \in C_b(\mathbb{R}^{d_1+d_2})$ , then there exists a constant  $0 < \varepsilon_0 \leq 1$  such that for all  $p \geq 1$ ,  $T > 0$  and  $0 < \varepsilon \leq \varepsilon_0$

$$\mathbb{E} \left( \sup_{t \in [0, T]} |Y_t^\varepsilon(\xi^\varepsilon)|^{2p} \right) \leq C_{p, T} (1 + \mathbb{E}|\xi^\varepsilon|^{2p}), \quad (4.2)$$

$$\mathbb{E} \left( \sup_{t \in [0, T]} |X_t^\varepsilon(\zeta^\varepsilon)|^{2p} \right) \leq C_{p, T} (1 + \mathbb{E}|\zeta^\varepsilon|^{2p} + \mathbb{E}|\zeta^\varepsilon|^{\theta p}) \quad (4.3)$$

and

$$\mathbb{E} \left( \sup_{t \in [0, T]} |\bar{X}_t(\zeta)|^{2p} \right) \leq C_{p, T} (1 + \mathbb{E}|\zeta|^{2p}), \quad (4.4)$$

where  $\theta$  is as in  $(\mathbf{H}_x^1)$ , and  $C_{p, T}$  is independent of  $\varepsilon$ .

*Proof.* Employing Itô's formula and  $(\mathbf{H}_y^1)$ , we have

$$\begin{aligned} |Y_t^\varepsilon(\xi^\varepsilon)|^{2p} &= |\xi^\varepsilon|^{2p} + p \int_0^t |Y_s^\varepsilon(\xi^\varepsilon)|^{2p-2} \left( 2\langle \varepsilon^{-2\alpha} B(X_s^\varepsilon(\zeta^\varepsilon), Y_s^\varepsilon(\xi^\varepsilon)), Y_s^\varepsilon(\xi^\varepsilon) \rangle \right. \\ &\quad \left. + 2\langle \varepsilon^{-\beta} b(X_s^\varepsilon(\zeta^\varepsilon), Y_s^\varepsilon(\xi^\varepsilon)), Y_s^\varepsilon(\xi^\varepsilon) \rangle + \varepsilon^{-2\alpha} |g(X_s^\varepsilon(\zeta^\varepsilon), Y_s^\varepsilon(\xi^\varepsilon))|_{HS}^2 \right) ds \\ &\quad + 2p(p-1)\varepsilon^{-2\alpha} \int_0^t |Y_s^\varepsilon(\xi^\varepsilon)|^{2p-4} |g^T(X_s^\varepsilon(\zeta^\varepsilon), Y_s^\varepsilon(\xi^\varepsilon)) Y_s^\varepsilon(\xi^\varepsilon)|^2 ds \\ &\quad + 2p\varepsilon^{-\alpha} \int_0^t |Y_s^\varepsilon(\xi^\varepsilon)|^{2p-2} \langle Y_s^\varepsilon(\xi^\varepsilon), g(X_s^\varepsilon(\zeta^\varepsilon), Y_s^\varepsilon(\xi^\varepsilon)) dW_s^2 \rangle \\ &\leq |\xi^\varepsilon|^{2p} + p \int_0^t -\varepsilon^{-2\alpha} (\eta - \varepsilon^{2\alpha-\beta} \tilde{\eta}) |Y_s^\varepsilon(\xi^\varepsilon)|^{2p} ds \\ &\quad + p \int_0^t (\varepsilon^{-2\alpha} K_1 + \varepsilon^{-\beta} K_1 + 2(p-1)\varepsilon^{-2\alpha} \|g\|_\infty) |Y_s^\varepsilon(\xi^\varepsilon)|^{2p-2} ds \\ &\quad + 2p\varepsilon^{-\alpha} \int_0^t |Y_s^\varepsilon(\xi^\varepsilon)|^{2p-2} \langle Y_s^\varepsilon(\xi^\varepsilon), g(X_s^\varepsilon(\zeta^\varepsilon), Y_s^\varepsilon(\xi^\varepsilon)) dW_s^2 \rangle. \end{aligned}$$

Note that there exists a constant  $0 < \varepsilon_0 \leq 1$  such that  $\eta - \varepsilon^{2\alpha-\beta} \tilde{\eta} > \frac{\eta}{2}$  for all  $0 < \varepsilon \leq \varepsilon_0$ . Then by Young's inequality, one sees that for any  $0 < \varepsilon \leq \varepsilon_0$

$$\begin{aligned} |Y_t^\varepsilon(\xi^\varepsilon)|^{2p} &\leq |\xi^\varepsilon|^{2p} + p \int_0^t -\frac{1}{4}\varepsilon^{-2\alpha}\eta |Y_s^\varepsilon(\xi^\varepsilon)|^{2p} ds \\ &\quad + 2p\varepsilon^{-\alpha} \int_0^t |Y_s^\varepsilon(\xi^\varepsilon)|^{2p-2} \langle Y_s^\varepsilon(\xi^\varepsilon), g(X_s^\varepsilon(\zeta^\varepsilon), Y_s^\varepsilon(\xi^\varepsilon)) dW_s^2 \rangle \end{aligned} \quad (4.5)$$

$$+ \int_0^t \left( \frac{p}{4(p-1)} \varepsilon^{-2\alpha} \eta \right)^{1-p} (\varepsilon^{-2\alpha} K_1 + \varepsilon^{-\beta} K_1 + 2(p-1) \varepsilon^{-2\alpha} \|g\|_\infty)^p ds,$$

which implies that for any stopping time  $\tau \leq T$

$$\mathbb{E}|Y_\tau^\varepsilon(\xi^\varepsilon)|^{2p} \leq \mathbb{E} \left( |\xi^\varepsilon|^{2p} e^{-\frac{1}{4}\varepsilon^{-2\alpha} p \eta \tau} \right) + C_{p,T,\eta,\|g\|_\infty} \leq C_{p,T} (1 + \mathbb{E}|\xi^\varepsilon|^{2p}). \quad (4.6)$$

Define  $\tau_r^\varepsilon := \inf\{t \geq 0 : |Y_t^\varepsilon(\xi^\varepsilon)|^{2p} > r\} \wedge T$ . In view of (4.6), for any  $0 < \alpha < 1$  we have

$$\begin{aligned} \mathbb{E} \left( \sup_{t \in [0,T]} |Y_t^\varepsilon(\xi^\varepsilon)|^{2p} \right)^\alpha &= \alpha \int_0^\infty r^{\alpha-1} \mathbb{P} \left( \sup_{t \in [0,T]} |Y_t^\varepsilon(\xi^\varepsilon)|^{2p} > r \right) dr \\ &\leq \alpha \int_0^\infty r^{\alpha-1} (1 \wedge r^{-1} \mathbb{E}|Y_{\tau_r^\varepsilon}^\varepsilon(\xi^\varepsilon)|^{2p}) dr \\ &\leq \alpha \int_0^\infty r^{\alpha-1} (1 \wedge r^{-1} C_{p,T} (1 + \mathbb{E}|\xi^\varepsilon|^{2p})) dr \\ &\leq \alpha C_{p,T} (1 + \mathbb{E}|\xi^\varepsilon|^{2p})^\alpha \int_0^\infty \lambda^{\alpha-1} (1 \wedge \lambda^{-1}) d\lambda \\ &\leq C_{p,T} (1 + \mathbb{E}|\xi^\varepsilon|^{2p})^\alpha \end{aligned}$$

by the change of variables  $r \mapsto [C_{p,T} (1 + \mathbb{E}|\xi^\varepsilon|^{2p})] \lambda$ .

Note that

$$X_t^\varepsilon(\zeta^\varepsilon) = \zeta^\varepsilon + \int_0^t f_\varepsilon(s, X_s^\varepsilon(\zeta^\varepsilon), Y_s^\varepsilon(\xi^\varepsilon)) ds + \int_0^t \sigma_\varepsilon(s, X_s^\varepsilon(\zeta^\varepsilon)) dW_s^1.$$

It follows from Itô's formula and  $(\mathbf{H}_x^1)$  that

$$\begin{aligned} |X_t^\varepsilon(\zeta^\varepsilon)|^{2p} &\leq |\zeta^\varepsilon|^{2p} + p \int_0^t |X_s^\varepsilon(\zeta^\varepsilon)|^{2p-2} (K_4(1 + |X_s^\varepsilon(\zeta^\varepsilon)|^2) + K_5 |Y_s^\varepsilon(\xi)|^\theta) ds \\ &\quad + 2p \int_0^t |X_s^\varepsilon(\zeta^\varepsilon)|^{2p-2} \langle X_s^\varepsilon(\zeta^\varepsilon), \sigma_\varepsilon(s, X_s^\varepsilon(\zeta^\varepsilon)) \rangle dW_s^1 \\ &\quad + 2p(p-1) \int_0^t |X_s^\varepsilon(\zeta^\varepsilon)|^{2p-2} |\sigma_\varepsilon(s, X_s^\varepsilon(\zeta^\varepsilon))|^2 ds. \end{aligned}$$

Then by Burkholder-Davis-Gundy's inequality and Young's inequality, one sees that

$$\begin{aligned} &\mathbb{E} \left( \sup_{t \in [0,T]} |X_t^\varepsilon(\zeta^\varepsilon)|^{2p} \right) \\ &\leq \mathbb{E} |\zeta^\varepsilon|^{2p} + p \mathbb{E} \int_0^T |X_s^\varepsilon(\zeta^\varepsilon)|^{2p-2} (K_4(1 + |X_s^\varepsilon(\zeta^\varepsilon)|^2) + K_5 |Y_s^\varepsilon(\xi)|^\theta) ds \\ &\quad + \frac{1}{2} \mathbb{E} \left( \sup_{t \in [0,T]} |X_t^\varepsilon(\zeta^\varepsilon)|^{2p} \right) + C_T \mathbb{E} \int_0^T (|X_s^\varepsilon(\zeta^\varepsilon)|^{2p} + 1) ds \\ &\quad + 2p(p-1) \mathbb{E} \int_0^T |X_s^\varepsilon(\zeta^\varepsilon)|^{2p-2} (2L_\sigma^2 |X_s^\varepsilon(\zeta^\varepsilon)|^2 + C) ds, \end{aligned}$$

which by (4.2) and Gronwall's inequality implies that

$$\mathbb{E} \left( \sup_{t \in [0,T]} |X_t^\varepsilon(\zeta^\varepsilon)|^{2p} \right) \leq C_{p,T} (1 + \mathbb{E}|\zeta^\varepsilon|^{2p} + \mathbb{E}|\xi|^\theta).$$

Similarly, we have

$$\mathbb{E} \left( \sup_{t \in [0,T]} |\bar{X}_t(\zeta)|^{2p} \right) \leq C_{p,T} (1 + \mathbb{E}|\zeta|^{2p}).$$

□

Finally, we prove the Hölder continuity of  $X_t^\varepsilon, t \geq 0$ .

**Lemma 4.4.** *Assume that  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$  and  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^3)$  hold. If  $g \in C_b(\mathbb{R}^{d_1+d_2})$ , then there exists a constant  $C > 0$  such that for all  $0 \leq s \leq t \leq T$*

$$\mathbb{E}|X_t^\varepsilon(x) - X_s^\varepsilon(x)|^2 \leq C_{T,|x|,|y|}|t - s|. \quad (4.7)$$

*Proof.* It follows from Itô's formula, the Burkholder-Davis-Gundy inequality,  $(\mathbf{H}_x^2)$ ,  $(\mathbf{H}_x^3)$ , (4.2) and (4.3) that

$$\begin{aligned} & \mathbb{E}|X_t^\varepsilon(x) - X_s^\varepsilon(x)|^2 \\ & \leq 2\mathbb{E} \left| \int_s^t f_\varepsilon(r, X_r^\varepsilon(x), Y_r^\varepsilon(y)) dr \right|^2 + 2\mathbb{E} \left| \int_s^t \sigma_\varepsilon(r, X_r^\varepsilon(x)) dW_r^1 \right|^2 \\ & \leq |t - s| \int_s^t \mathbb{E} |f_\varepsilon(r, X_r^\varepsilon(x), Y_r^\varepsilon(y))|^2 dr + 4\mathbb{E} \int_s^t |\sigma_\varepsilon(r, X_r^\varepsilon(x))|_{HS}^2 dr \\ & \leq |t - s| C \int_s^t \mathbb{E} (1 + |X_r^\varepsilon(x)|^{2\theta_1} + |Y_r^\varepsilon(y)|^{2\theta_2}) dr + C\mathbb{E} \int_s^t (1 + |X_r^\varepsilon(x)|^2) dr \\ & \leq C_{T,|x|,|y|}|t - s|. \end{aligned}$$

□

**4.2. Proof of Theorem 2.3.** Now we are in a position to prove Theorem 2.3.

*Proof.* For brevity, we define  $X_t^\varepsilon := X_t^\varepsilon(x)$ ,  $Y_t^\varepsilon := Y_t^\varepsilon(y)$  and  $\bar{X}_t := \bar{X}_t(x)$  for all  $t \geq 0$  in this proof. By Itô's formula and the Burkholder-Davis-Gundy inequality, we have

$$\begin{aligned} & \mathbb{E}|X_t^\varepsilon - \bar{X}_t|^2 \\ & \leq 2\mathbb{E} \left| \int_0^t (f_\varepsilon(s, X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(\bar{X}_s)) ds \right|^2 + 2\mathbb{E} \left| \int_0^t (\sigma_\varepsilon(s, X_s^\varepsilon) - \bar{\sigma}(\bar{X}_s)) dW_s^1 \right|^2 \\ & \leq 4\mathbb{E} \left| \int_0^t (f_\varepsilon(s, X_s^\varepsilon, Y_s^\varepsilon) - \hat{f}_\varepsilon(s, X_s^\varepsilon)) ds \right|^2 + 4\mathbb{E} \left| \int_0^t (\hat{f}_\varepsilon(s, X_s^\varepsilon) - \bar{f}(\bar{X}_s)) ds \right|^2 \\ & \quad + 4\mathbb{E} \int_0^t |\sigma_\varepsilon(s, X_s^\varepsilon) - \bar{\sigma}(\bar{X}_s)|_{HS}^2 ds =: 4\mathfrak{J}_1(t, \varepsilon) + 4\mathfrak{J}_2(t, \varepsilon) + 4\mathfrak{J}_3(t, \varepsilon). \end{aligned} \quad (4.8)$$

First of all, we estimate  $\mathfrak{J}_1(t, \varepsilon)$ . Set  $\phi(t, x, y) := f(t, x, y) - \hat{f}(t, x)$ . Let  $\psi$  be the solution to the following Poisson equation

$$\mathfrak{L}_2\psi(t, x, y) = \phi(t, x, y), \quad y \in \mathbb{R}^{d_2},$$

where  $t \in \mathbb{R}$  and  $x \in \mathbb{R}^{d_1}$  are parameters. By Lemma 3.4, one sees that  $\psi \in C^{1,2,2}(\mathbb{R}^{1+d_1+d_2})$ . Then according to Itô's formula, we have for any  $t > 0$

$$\begin{aligned} & \psi(\varepsilon^{-\gamma}t, X_t^\varepsilon, Y_t^\varepsilon) \\ & = \psi(0, x, y) + \int_0^t (\varepsilon^{-\gamma}\partial_s + \mathfrak{L}_1^\varepsilon)\psi(\varepsilon^{-\gamma}s, X_s^\varepsilon, Y_s^\varepsilon) ds + \int_0^t \varepsilon^{-2\alpha}\mathfrak{L}_2\psi(\varepsilon^{-\gamma}s, X_s^\varepsilon, Y_s^\varepsilon) ds \\ & \quad + \int_0^t \varepsilon^{-\beta}\mathfrak{L}_3\psi(\varepsilon^{-\gamma}s, X_s^\varepsilon, Y_s^\varepsilon) ds + \int_0^t \sigma(\varepsilon^{-\gamma}s, X_s^\varepsilon) \cdot \partial_x\psi(\varepsilon^{-\gamma}s, X_s^\varepsilon, Y_s^\varepsilon) dW_s^1 \\ & \quad + \varepsilon^{-\alpha} \int_0^t g(X_s^\varepsilon, Y_s^\varepsilon) \cdot \partial_y\psi(\varepsilon^{-\gamma}s, X_s^\varepsilon, Y_s^\varepsilon) dW_s^2. \end{aligned}$$

Therefore,

$$\mathbb{E} \left| \int_0^t f_\varepsilon(s, X_s^\varepsilon, Y_s^\varepsilon) - \hat{f}_\varepsilon(s, X_s^\varepsilon) ds \right|^2$$

$$\begin{aligned} &\leq \varepsilon^{4\alpha} \mathbb{E} \left| \psi(\varepsilon^{-\gamma} t, X_t^\varepsilon, Y_t^\varepsilon) - \psi(0, x, y) - \int_0^t (\varepsilon^{-\gamma} \partial_s + \mathfrak{L}_1^\varepsilon) \psi(\varepsilon^{-\gamma} s, X_s^\varepsilon, Y_s^\varepsilon) ds \right. \\ &\quad \left. - \int_0^t \varepsilon^{-\beta} \mathfrak{L}_3 \psi(\varepsilon^{-\gamma} s, X_s^\varepsilon, Y_s^\varepsilon) ds - M_t^1 - \varepsilon^{-\alpha} M_t^2 \right|^2, \end{aligned}$$

where

$$\begin{aligned} M_t^1 &:= \int_0^t \sigma_\varepsilon(s, X_s^\varepsilon) \cdot \partial_x \psi(\varepsilon^{-\gamma} s, X_s^\varepsilon, Y_s^\varepsilon) dW_s^1, \\ M_t^2 &:= \int_0^t g(X_s^\varepsilon, Y_s^\varepsilon) \cdot \partial_y \psi(\varepsilon^{-\gamma} s, X_s^\varepsilon, Y_s^\varepsilon) dW_s^2. \end{aligned}$$

Then by Lemma 3.4,  $(\mathbf{H}_x^2)$ ,  $(\mathbf{H}_x^3)$  and  $(\mathbf{H}_y^3)$ , one sees that there exist  $m', m'' > 0$  such that

$$\begin{aligned} \mathfrak{I}_1(t, \varepsilon) &:= \mathbb{E} \left| \int_0^t f_\varepsilon(s, X_s^\varepsilon, Y_s^\varepsilon) - \hat{f}_\varepsilon(s, X_s^\varepsilon) ds \right|^2 \\ &\leq \varepsilon^{4\alpha} C \left( 1 + \mathbb{E} \left( \sup_{t \in [0, T]} |X_t^\varepsilon|^{m'} \right) + \mathbb{E} \left( \sup_{t \in [0, T]} |Y_t^\varepsilon|^{m''} \right) \right) \\ &\quad + (\varepsilon^{4\alpha-2\gamma} + \varepsilon^{4\alpha} + \varepsilon^{4\alpha-2\beta}) C \int_0^t \left( 1 + \mathbb{E} |X_s^\varepsilon|^{m'} + \mathbb{E} |Y_s^\varepsilon|^{m''} \right) ds \\ &\quad + 5\varepsilon^{4\alpha} \mathbb{E} |M_t^1|^2 + 5\varepsilon^{2\alpha} \mathbb{E} |M_t^2|^2. \end{aligned}$$

And it follows from the Burkholder-Davis-Gundy inequality, Lemma 3.4,  $(\mathbf{H}_x^3)$  and  $(\mathbf{H}_y^3)$  that there exist  $m', m'' > 0$  such that

$$\begin{aligned} \mathbb{E} |M_t^1|^2 + \mathbb{E} |M_t^2|^2 &\leq 2 \int_0^t \mathbb{E} |\sigma_\varepsilon(s, X_s^\varepsilon) \cdot \partial_x \psi(\varepsilon^{-\gamma} s, X_s^\varepsilon, Y_s^\varepsilon)|^2 ds \\ &\quad + 2 \int_0^t \mathbb{E} |g(X_s^\varepsilon, Y_s^\varepsilon) \cdot \partial_y \psi(\varepsilon^{-\gamma} s, X_s^\varepsilon, Y_s^\varepsilon)|^2 ds \\ &\leq C \int_0^t \left( 1 + \mathbb{E} |X_s^\varepsilon|^{m'} + \mathbb{E} |Y_s^\varepsilon|^{m''} \right) ds. \end{aligned}$$

Hence, letting  $\tilde{\alpha} := \min\{4\alpha - 2\gamma, 4\alpha - 2\beta, 2\alpha\}$ , by Lemma 4.3, one sees that

$$\mathfrak{I}_1(t, \varepsilon) \leq C_T \varepsilon^{\tilde{\alpha}}. \quad (4.9)$$

Now we estimate

$$\mathfrak{I}_2(t, \varepsilon) := \mathbb{E} \left| \int_0^t \left( \hat{f}_\varepsilon(s, X_s^\varepsilon) - \bar{f}(\bar{X}_s) \right) ds \right|^2.$$

Define  $\tau_n^\varepsilon := \inf \{t \geq 0 : |X_t^\varepsilon| + |\bar{X}_t| > n\} \wedge T$ . Then we have

$$\begin{aligned} \mathfrak{I}_2(t, \varepsilon) &\leq \mathbb{E} \left( \chi_{\{\tau_n^\varepsilon \geq t\}} \left| \int_0^t \left( \hat{f}_\varepsilon(s, X_s^\varepsilon) - \bar{f}(\bar{X}_s) \right) ds \right|^2 \right) \\ &\quad + \mathbb{E} \left( \chi_{\{\tau_n^\varepsilon \leq t\}} \left| \int_0^t \left( \hat{f}_\varepsilon(s, X_s^\varepsilon) - \bar{f}(\bar{X}_s) \right) ds \right|^2 \right) =: \mathcal{I}_1 + \mathcal{I}_2. \end{aligned} \quad (4.10)$$

For  $\mathcal{I}_2$ , by Hölder's inequality, Chebyshev's inequality,  $(\mathbf{H}_x^2)$ , Lemmas 4.1 and 4.3, we have

$$\begin{aligned} \mathcal{I}_2 &\leq \left( \mathbb{E} \chi_{\{\tau_n^\varepsilon \leq t\}}^2 \right)^{\frac{1}{2}} \left( \mathbb{E} \left| \int_0^t \left( \hat{f}_\varepsilon(s, X_s^\varepsilon) - \bar{f}(\bar{X}_s) \right) ds \right|^4 \right)^{\frac{1}{2}} \\ &\leq n^{-1} C_T \left( \mathbb{E} \left( \sup_{t \in [0, T]} |X_t^\varepsilon|^2 \right) + \mathbb{E} \left( \sup_{t \in [0, T]} |\bar{X}_t|^2 \right) \right)^{\frac{1}{2}} \end{aligned} \quad (4.11)$$

$$\times \left( \mathbb{E} \int_0^T (|X_s^\varepsilon|^{4\theta_1} + |\bar{X}_s|^{4\theta_1} + 1) ds \right)^{\frac{1}{2}} \leq C_T n^{-1}.$$

Set  $\tilde{X}_s^\varepsilon := X_{k\delta}^\varepsilon$  for  $s \in [k\delta, (k+1)\delta)$ ,  $k \in \mathbb{N}$ . For  $\mathcal{I}_1$ , by Hölder's inequality, Lemmas 4.1 and 4.4, one sees that

$$\begin{aligned} \mathcal{I}_1 &\leq 4\mathbb{E} \left| \int_0^{t \wedge \tau_n^\varepsilon} (\hat{f}_\varepsilon(s, X_s^\varepsilon) - \hat{f}_\varepsilon(s, \tilde{X}_s^\varepsilon)) ds \right|^2 \\ &\quad + 4\mathbb{E} \left( \chi_{\{\tau_n^\varepsilon \geq t\}} \left| \int_0^t (\hat{f}_\varepsilon(s, \tilde{X}_s^\varepsilon) - \bar{f}(\tilde{X}_s^\varepsilon)) ds \right|^2 \right) \\ &\quad + 4\mathbb{E} \left| \int_0^{t \wedge \tau_n^\varepsilon} (\bar{f}(\tilde{X}_s^\varepsilon) - \bar{f}(X_s^\varepsilon)) ds \right|^2 + 4\mathbb{E} \left| \int_0^{t \wedge \tau_n^\varepsilon} (\bar{f}(X_s^\varepsilon) - \bar{f}(\bar{X}_s)) ds \right|^2 \\ &\leq C_T n^{2\theta_1} \int_0^T \mathbb{E} |X_s^\varepsilon - \tilde{X}_s^\varepsilon|^2 ds + C_T n^{2\theta_1} \int_0^T \mathbb{E} |X_s^\varepsilon - \bar{X}_s|^2 ds + \mathcal{I}_1^2 \\ &\leq C_T n^{2\theta_1} \delta + C_T n^{2\theta_1} \int_0^t \mathbb{E} |X_s^\varepsilon - \bar{X}_s|^2 ds + \mathcal{I}_1^2, \end{aligned} \quad (4.12)$$

where

$$\mathcal{I}_1^2 := 4\mathbb{E} \left( \chi_{\{\tau_n^\varepsilon \geq t\}} \left| \int_0^t (\hat{f}_\varepsilon(s, \tilde{X}_s^\varepsilon) - \bar{f}(\tilde{X}_s^\varepsilon)) ds \right|^2 \right).$$

Employing the technique of time discretization, a change of variables,  $(\mathbf{A}_f)$ , Remark 2.2 and (4.3), we have

$$\begin{aligned} \mathcal{I}_1^2 &\leq 8\mathbb{E} \left( \chi_{\{\tau_n^\varepsilon \geq t\}} \left| \sum_{k=0}^{t(\delta)} \int_{k\delta}^{(k+1)\delta} (\hat{f}_\varepsilon(s, X_{k\delta}^\varepsilon) - \bar{f}(X_{k\delta}^\varepsilon)) ds \right|^2 \right) \\ &\quad + 8\mathbb{E} \left( \chi_{\{\tau_n^\varepsilon \geq t\}} \left| \int_{t(\delta)\delta}^t (\hat{f}_\varepsilon(s, X_{t(\delta)\delta}^\varepsilon) - \bar{f}(X_{t(\delta)\delta}^\varepsilon)) ds \right|^2 \right) \\ &\leq C_T \left[ \delta (\omega_n^f(\delta/\varepsilon^\gamma))^2 + \delta^2 \right], \end{aligned} \quad (4.13)$$

where  $t(\delta) := \lceil \frac{t}{\delta} \rceil$ . Then (4.12) and (4.13) yield

$$\mathcal{I}_1 \leq C_T \left[ \delta (\omega_n^f(\delta/\varepsilon^\gamma))^2 + n^{2\theta_1} \delta \right] + C_T n^{2\theta_1} \int_0^t \mathbb{E} |X_s^\varepsilon - \bar{X}_s|^2 ds. \quad (4.14)$$

Combining (4.10), (4.11) and (4.14), we have

$$\mathfrak{I}_2(t, \varepsilon) \leq C_T \left[ \delta (\omega_n^f(\delta/\varepsilon^\gamma))^2 + n^{2\theta_1} \delta + n^{-1} \right] + C_T n^{2\theta_1} \int_0^t \mathbb{E} |X_s^\varepsilon - \bar{X}_s|^2 ds. \quad (4.15)$$

Similarly, for  $\mathfrak{I}_3(t, \varepsilon)$ , it follows from the Burkholder-Davis-Gundy inequality,  $(\mathbf{A}_\sigma)$ ,  $(\mathbf{H}_x^3)$ , Lemma 4.4 and (4.3) that

$$\begin{aligned} \mathfrak{I}_3(t, \varepsilon) &:= \mathbb{E} \left| \int_0^t (\sigma_\varepsilon(s, X_s^\varepsilon) - \bar{\sigma}(\bar{X}_s)) dW_s^1 \right|^2 \\ &\leq 4\mathbb{E} \int_0^t |\sigma_\varepsilon(s, X_s^\varepsilon) - \sigma_\varepsilon(s, \tilde{X}_s^\varepsilon)|_{HS}^2 ds + 4\mathbb{E} \int_0^t |\sigma_\varepsilon(s, \tilde{X}_s^\varepsilon) - \bar{\sigma}(\tilde{X}_s^\varepsilon)|_{HS}^2 ds \\ &\quad + 4\mathbb{E} \int_0^t |\bar{\sigma}(X_s^\varepsilon) - \bar{\sigma}(\tilde{X}_s^\varepsilon)|_{HS}^2 ds + 4\mathbb{E} \int_0^t L_\sigma |X_s^\varepsilon - \bar{X}_s|^2 ds \end{aligned} \quad (4.16)$$

$$\leq C_T (\delta + \omega^\sigma(\delta/\varepsilon^\gamma)) + 4\mathbb{E} \int_0^t L_\sigma |X_s^\varepsilon - \bar{X}_s|^2 ds.$$

In view of (4.8), (4.9), (4.15) and (4.16), we have

$$\begin{aligned} \sup_{t \in [0, T]} \mathbb{E} |X_t^\varepsilon - \bar{X}_t|^2 &\leq C\varepsilon^{\tilde{\alpha}} + C_T n^{2\theta_1} \int_0^T \sup_{0 \leq r \leq s} \mathbb{E} |X_r^\varepsilon - \bar{X}_r|^2 ds \\ &\quad + C_T \left[ \delta (\omega_n^f(\delta/\varepsilon^\gamma))^2 + n^{2\theta_1} \delta + \omega^\sigma(\delta/\varepsilon^\gamma) + n^{-1} \right], \end{aligned}$$

which by Gronwall's inequality implies

$$\begin{aligned} \sup_{t \in [0, T]} \mathbb{E} |X_t^\varepsilon - \bar{X}_t|^2 &\leq C_T \left( \varepsilon^{\tilde{\alpha}} + n^{2\theta_1} \delta \right) \exp(Cn^{2\theta_1} T) \\ &\quad + C_T \left( (\omega_n^f(\delta/\varepsilon^\gamma))^2 + n^{-1} + \omega^\sigma(\delta/\varepsilon^\gamma) \right) \exp(Cn^{2\theta_1} T). \end{aligned} \quad (4.17)$$

Let  $\delta = \varepsilon^{\gamma/2}$ . Taking  $\varepsilon \rightarrow 0$  and  $n \rightarrow \infty$ , we obtain

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

□

## 5. NORMAL DEVIATION

This section is dedicated to proving the normal deviation for slow-fast stochastic differential equations (2.2). In Section 5.1, we prove the optimal strong convergence rate for (2.2). Subsequently, we show the normal deviation in Section 5.2.

**5.1. The optimal strong convergence rate.** Before investigating the optimal strong convergence rate, let us first show that  $\bar{f}$  is monotone under some suitable conditions.

**Lemma 5.1.** *Assume that  $(\mathbf{H}_y^1)$ – $(\mathbf{H}_y^3)$  and  $(\mathbf{H}_x^4)$  hold. Then there exists a constant  $C > 0$  such that for any  $x_1, x_2 \in \mathbb{R}^{d_1}$*

$$\langle \bar{f}(x_1) - \bar{f}(x_2), x_1 - x_2 \rangle \leq C|x_1 - x_2|^2.$$

*Proof.* It follows from  $(\mathbf{H}_x^4)$ , Hölder's inequality, Corollary 3.2 and Lemma 3.3 that for all  $x_1, x_2 \in \mathbb{R}^{d_1}$

$$\begin{aligned} &\langle \bar{f}(x_1) - \bar{f}(x_2), x_1 - x_2 \rangle \\ &= \left\langle \int_{\mathbb{R}^{d_2}} f(x_1, y) \mu^{x_1}(dy) - \int_{\mathbb{R}^{d_2}} f(x_2, y) \mu^{x_2}(dy), x_1 - x_2 \right\rangle \\ &= \int_{\mathbb{R}^{d_2}} \langle f(x_1, y) - f(x_2, y), x_1 - x_2 \rangle \mu^{x_1}(dy) + \left\langle \int_{\mathbb{R}^{d_2}} f(x_2, y) (\mu^{x_1} - \mu^{x_2})(dy), x_1 - x_2 \right\rangle \\ &\leq |x_1 - x_2|^2 \int_{\mathbb{R}^{d_2}} M (1 + |y|^{\theta_2}) \mu^{x_1}(dy) + \left\langle \int_{\mathbb{R}^{d_2}} f(x_2, y) (\mu^{x_1} - \mu^{x_2})(dy), x_1 - x_2 \right\rangle \\ &\leq C|x_1 - x_2|^2 + C\mathbb{E} \left[ (1 + |Y_t^{x_1}|^{\theta_2} + |Y_t^{x_2}|^{\theta_2}) |Y_t^{x_1} - Y_t^{x_2}| \right] |x_1 - x_2| \\ &\leq C|x_1 - x_2|^2 + C \left[ \mathbb{E} (1 + |Y_t^{x_1}|^{2\theta_1} + |Y_t^{x_2}|^{2\theta_2}) \right]^{\frac{1}{2}} \left( \mathbb{E} |Y_t^{x_1} - Y_t^{x_2}|^2 \right)^{\frac{1}{2}} |x_1 - x_2| \\ &\leq C|x_1 - x_2|^2, \end{aligned}$$

where  $t \in \mathbb{R}$  and  $Y_t^{x_1}$  (respectively,  $Y_t^{x_2}$ ) is the stationary solution to (2.1) with frozen  $x_1$  (respectively,  $x_2$ ). □

Now we can give the proof of the optimal strong convergence rate for the first averaging principle.

*Proof of Theorem 2.5.* For simplicity, we define

$$X_t^\varepsilon := X_t^\varepsilon(x), \quad \bar{X}_t := \bar{X}_t(x), \quad Y_t^\varepsilon := Y_t^\varepsilon(y)$$

for any  $t \geq 0$  in this subsection. By Itô's formula,  $(\mathbf{H}_x^3)$  and Lemma 5.1, we have

$$\begin{aligned} |X_t^\varepsilon - \bar{X}_t|^2 &= \int_0^t \left( 2\langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(X_s^\varepsilon), X_s^\varepsilon - \bar{X}_s \rangle + 2\langle \bar{f}(X_s^\varepsilon) - \bar{f}(\bar{X}_s), X_s^\varepsilon - \bar{X}_s \rangle \right. \\ &\quad \left. + |\sigma(X_s^\varepsilon) - \sigma(\bar{X}_s)|_{HS}^2 \right) ds + 2 \int_0^t \langle X_s^\varepsilon - \bar{X}_s, (\sigma(X_s^\varepsilon) - \sigma(\bar{X}_s)) dW_s^1 \rangle \\ &\leq \int_0^t 2\langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(X_s^\varepsilon), X_s^\varepsilon - \bar{X}_s \rangle + C|X_s^\varepsilon - \bar{X}_s|^2 ds \\ &\quad + 2 \int_0^t \langle X_s^\varepsilon - \bar{X}_s, (\sigma(X_s^\varepsilon) - \sigma(\bar{X}_s)) dW_s^1 \rangle. \end{aligned} \quad (5.1)$$

Define  $\phi(x, y, \bar{x}) := 2\langle f(x, y), x - \bar{x} \rangle$ ,  $\bar{\phi}(x, \bar{x}) := 2\langle \bar{f}(x), x - \bar{x} \rangle$  for all  $(x, y, \bar{x}) \in \mathbb{R}^{d_1+d_2+d_1}$ . For any  $(x, \bar{x}) \in \mathbb{R}^{2d_1}$  let  $\Psi(x, \cdot, \bar{x})$  be the solution to

$$\mathfrak{L}_2 \Psi(x, y, \bar{x}) = \phi(x, y, \bar{x}) - \bar{\phi}(x, \bar{x}), \quad y \in \mathbb{R}^{d_2}.$$

By Lemma 3.4, one sees that  $\Psi \in C^{2,4,2}(\mathbb{R}^{d_1+d_2+d_1})$  and that there exist constants  $C, m_1, m_2, m_3 > 0$  such that

$$\begin{aligned} |\Psi(x, y, \bar{x})| &+ \sum_{i=1}^2 (|\partial_x^i \Psi(x, y, \bar{x})| + |\partial_{\bar{x}}^i \Psi(x, y, \bar{x})|) + |\partial_y \Psi(x, y, \bar{x})| \\ &\leq C(1 + |x|^{m_1} + |y|^{m_2} + |\bar{x}|^{m_3}). \end{aligned} \quad (5.2)$$

Applying Itô's formula to  $t \mapsto \Psi(X_t^\varepsilon, Y_t^\varepsilon, \bar{X}_t)$ , we have

$$\begin{aligned} &\Psi(X_t^\varepsilon, Y_t^\varepsilon, \bar{X}_t) - \Psi(x, y, \bar{x}) \\ &= \int_0^t (\mathfrak{L}_1 + \mathfrak{L}_{\bar{x}} + \varepsilon^{-1} \mathfrak{L}_2) \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) ds + \int_0^t \sigma(X_s^\varepsilon) \cdot \partial_x \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) dW_s^1 \\ &\quad + \int_0^t \sigma(\bar{X}_s) \cdot \partial_{\bar{x}} \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) dW_s^1 + \varepsilon^{-\frac{1}{2}} \int_0^t g(X_s^\varepsilon, Y_s^\varepsilon) \cdot \partial_y \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) dW_s^2, \end{aligned}$$

which implies that

$$\begin{aligned} &\int_0^t 2\langle (f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(X_s^\varepsilon)), X_s^\varepsilon - \bar{X}_s \rangle ds \\ &= \varepsilon \left( \Psi(X_t^\varepsilon, Y_t^\varepsilon, \bar{X}_t) - \Psi(x, y, \bar{x}) - \int_0^t (\mathfrak{L}_1 + \mathfrak{L}_{\bar{x}}) \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) ds \right) \\ &\quad - \varepsilon \int_0^t \sigma(X_s^\varepsilon) \cdot \partial_x \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) dW_s^1 - \varepsilon \int_0^t \sigma(\bar{X}_s) \cdot \partial_{\bar{x}} \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) dW_s^1 \\ &\quad - \sqrt{\varepsilon} \int_0^t g(X_s^\varepsilon, Y_s^\varepsilon) \cdot \partial_y \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) dW_s^2. \end{aligned} \quad (5.3)$$

Note that  $\Psi(x, y, \bar{x}) = 2\langle \psi(x, y), x - \bar{x} \rangle$ , where  $\psi$  is the solution to

$$\mathfrak{L}_2 \psi(x, y) = f(x, y) - \bar{f}(x).$$

Then in view of (5.1), (5.2), (5.3),  $(\mathbf{H}_x^2)$ ,  $(\mathbf{H}_x^3)$ , Lemma 4.1 and Young's inequality, there exist constants  $m'_1, m'_2, m'_3 > 0$  such that

$$\begin{aligned} &|X_t^\varepsilon - \bar{X}_t|^2 \\ &\leq \int_0^t C|X_s^\varepsilon - \bar{X}_s|^2 ds + 2 \int_0^t \langle X_s^\varepsilon - \bar{X}_s, (\sigma(X_s^\varepsilon) - \sigma(\bar{X}_s)) dW_s^1 \rangle \end{aligned}$$

$$\begin{aligned}
& + \varepsilon \left( \Psi(X_t^\varepsilon, Y_t^\varepsilon, \bar{X}_t) - \Psi(x, y, x) - \int_0^t (\mathfrak{L}_1 + \mathfrak{L}_{\bar{x}}) \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) ds \right) \\
& + \varepsilon \int_0^t \sigma(X_s^\varepsilon) \cdot \partial_x \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) dW_s^1 + \varepsilon \int_0^t \sigma(\bar{X}_s) \cdot \partial_{\bar{x}} \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) dW_s^1 \\
& + \sqrt{\varepsilon} 2 \int_0^t g(X_s^\varepsilon, Y_s^\varepsilon) \cdot \langle \partial_y \psi(X_s^\varepsilon, Y_s^\varepsilon), X_s^\varepsilon - \bar{X}_s \rangle dW_s^2 \\
& \leq \int_0^t C |X_s^\varepsilon - \bar{X}_s|^2 ds + 2 \int_0^t \langle X_s^\varepsilon - \bar{X}_s, (\sigma(X_s^\varepsilon) - \sigma(\bar{X}_s)) \rangle dW_s^1 \\
& + \varepsilon C \left( 1 + \sup_{t \in [0, T]} |X_t^\varepsilon|^{m'_1} + \sup_{t \in [0, T]} |Y_t^\varepsilon|^{m'_2} + \sup_{t \in [0, T]} |\bar{X}_t|^{m'_3} \right) \\
& + \varepsilon \int_0^t \sigma(X_s^\varepsilon) \cdot \partial_x \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) dW_s^1 + \varepsilon \int_0^t \sigma(\bar{X}_s) \cdot \partial_{\bar{x}} \Psi(X_s^\varepsilon, Y_s^\varepsilon, \bar{X}_s) dW_s^1 \\
& + \sqrt{\varepsilon} 2 \int_0^t g(X_s^\varepsilon, Y_s^\varepsilon) \cdot \langle \partial_y \psi(X_s^\varepsilon, Y_s^\varepsilon), X_s^\varepsilon - \bar{X}_s \rangle dW_s^2.
\end{aligned}$$

Then thanks to Lemma 4.3 and Burkholder-Davis-Gundy's inequality, we obtain

$$\begin{aligned}
& \mathbb{E} \left( \sup_{t \in [0, T]} |X_t^\varepsilon - \bar{X}_t|^2 \right) \\
& \leq C \int_0^T \mathbb{E} |X_s^\varepsilon - \bar{X}_s|^2 ds + 6 \mathbb{E} \left( \int_0^T |X_s^\varepsilon - \bar{X}_s|^2 |\sigma(X_s^\varepsilon) - \sigma(\bar{X}_s)|^2 ds \right)^{\frac{1}{2}} \\
& + \varepsilon C_T \left( 1 + \mathbb{E} \left( \sup_{t \in [0, T]} |X_t^\varepsilon|^{m'_1} \right) + \mathbb{E} \left( \sup_{t \in [0, T]} |Y_t^\varepsilon|^{m'_2} \right) + \mathbb{E} \left( \sup_{t \in [0, T]} |\bar{X}_t|^{m'_3} \right) \right) \\
& + \sqrt{\varepsilon} 6 \mathbb{E} \left( \int_0^T |g(X_s^\varepsilon, Y_s^\varepsilon)|^2 |\partial_y \psi(X_s^\varepsilon, Y_s^\varepsilon)|^2 |X_s^\varepsilon - \bar{X}_s|^2 ds \right)^{\frac{1}{2}} \\
& \leq \frac{1}{2} \mathbb{E} \left( \sup_{t \in [0, T]} |X_s^\varepsilon - \bar{X}_s|^2 \right) + C \int_0^T \mathbb{E} |X_s^\varepsilon - \bar{X}_s|^2 ds + C_T \varepsilon,
\end{aligned}$$

which by Gronwall's inequality implies that

$$\mathbb{E} \left( \sup_{t \in [0, T]} |X_s^\varepsilon - \bar{X}_s|^2 \right) \leq C_T \varepsilon.$$

□

**5.2. Proof of Theorem 2.6.** Prior to presenting the proof of the theorem regarding the normal deviation, we prove several lemmas.

**Lemma 5.2.** *Assume that  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$  and  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^3)$  hold. Let  $(X_t^\varepsilon(x), Y_t^\varepsilon(y))$  be the solution to (1.3) for any  $(x, y) \in \mathbb{R}^{d_1+d_2}$ . Furthermore, suppose that there exist constants  $c_1, c_2, C \geq 1$  such that for all  $(x, y) \in \mathbb{R}^{d_1+d_2}$*

$$|b(x, y)| + |B(x, y)| \leq C (1 + |x|^{c_1} + |y|^{c_2}).$$

*Then there exists a constant  $C_T > 0$  such that for any  $0 \leq s < t \leq T$  and  $0 < \varepsilon \leq 1$*

$$\varepsilon \mathbb{E} |Y_t^\varepsilon - Y_s^\varepsilon|^2 \leq C_T |t - s|. \tag{5.4}$$

*Proof.* It follows from Itô's formula, (4.3) and (4.2) that

$$\mathbb{E} |Y_t^\varepsilon - Y_s^\varepsilon|^2 = \frac{1}{\varepsilon} \mathbb{E} \int_s^t (2 \langle B(X_r^\varepsilon, Y_r^\varepsilon), Y_r^\varepsilon - Y_s^\varepsilon \rangle + |g(X_r^\varepsilon, Y_r^\varepsilon)|_{HS}^2) dr$$

$$\begin{aligned}
& + \frac{1}{\varepsilon^\beta} \mathbb{E} \int_s^t 2 \langle b(X_r^\varepsilon, Y_r^\varepsilon), Y_r^\varepsilon - Y_s^\varepsilon \rangle dr \\
& \leq \frac{1}{\varepsilon} \mathbb{E} \int_s^t C \left( 1 + \sup_{r \in [0, T]} |X_r^\varepsilon|^{2c_1} + \sup_{r \in [0, T]} |Y_r^\varepsilon|^{2c_2} \right) dr \leq \frac{1}{\varepsilon} C_T |t - s|.
\end{aligned}$$

□

**Lemma 5.3.** *If  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^4)$  and  $(\mathbf{H}_y^1)$ – $(\mathbf{H}_y^3)$  hold, then  $Z^\varepsilon$  is tight in  $C([0, T]; \mathbb{R}^{d_1})$ .*

*Proof.* According to (2.3) and the Arzela-Ascoli theorem, it suffices to show that there exists a constant  $C_T > 0$  such that for all  $0 \leq s \leq t \leq T$

$$\mathbb{E} |Z_t^\varepsilon - Z_s^\varepsilon| \leq C_T |t - s|^{\frac{1}{2}}.$$

To this end, by Itô's formula, the Burkholder-Davis-Gundy inequality, Hölder's inequality and (2.3), one sees that

$$\begin{aligned}
& \mathbb{E} |Z_t^\varepsilon - Z_s^\varepsilon| \\
& \leq \mathbb{E} \left| \int_s^t \frac{1}{\sqrt{\varepsilon}} (f(X_r^\varepsilon, Y_r^\varepsilon) - \bar{f}(\bar{X}_r)) dr \right| + \mathbb{E} \left| \int_s^t \frac{1}{\sqrt{\varepsilon}} (\sigma(X_r^\varepsilon) - \sigma(\bar{X}_r)) dW_r^1 \right| \\
& \leq \mathcal{J}_1 + \mathbb{E} \int_s^t |\nabla \bar{f}(\bar{X}_r + \iota(X_r^\varepsilon - \bar{X}_r))| |Z_r^\varepsilon| dr + C_T |t - s|^{\frac{1}{2}} \\
& \leq \mathcal{J}_1 + \left( \mathbb{E} \int_s^t |\nabla \bar{f}(\bar{X}_r + \iota(X_r^\varepsilon - \bar{X}_r))|^2 dr \right)^{\frac{1}{2}} \left( \mathbb{E} \int_s^t |Z_r^\varepsilon|^2 dr \right)^{\frac{1}{2}} + C_T |t - s|^{\frac{1}{2}},
\end{aligned} \tag{5.5}$$

where  $\iota \in [0, 1]$  and

$$\mathcal{J}_1 := \mathbb{E} \left| \int_s^t \frac{1}{\sqrt{\varepsilon}} (f(X_r^\varepsilon, Y_r^\varepsilon) - \bar{f}(X_r^\varepsilon)) dr \right|.$$

Employing Lemma 4.1, (4.3), (4.4) and (2.3), we obtain that there exists a constant  $C_T > 0$  such that

$$\begin{aligned}
& \left( \mathbb{E} \int_s^t |\nabla \bar{f}(\bar{X}_r + \iota(X_r^\varepsilon - \bar{X}_r))|^2 dr \right)^{\frac{1}{2}} \left( \mathbb{E} \int_s^t |Z_r^\varepsilon|^2 dr \right)^{\frac{1}{2}} \\
& \leq \left( \mathbb{E} \int_s^t (|X_r^\varepsilon|^{2\theta_1} + |\bar{X}_r|^{2\theta_1}) dr \right)^{\frac{1}{2}} C_T |t - s|^{\frac{1}{2}} \leq C_T |t - s|.
\end{aligned} \tag{5.6}$$

Now we estimate  $\mathcal{J}_1$ . Note that by the Burkholder-Davis-Gundy inequality, we get

$$\begin{aligned}
\mathcal{J}_1 & \leq \sqrt{\varepsilon} \mathbb{E} |\psi(X_t^\varepsilon, Y_t^\varepsilon) - \psi(X_s^\varepsilon, Y_s^\varepsilon)| + \sqrt{\varepsilon} \mathbb{E} \left| \int_s^t \mathfrak{L}_1 \psi(X_r^\varepsilon, Y_r^\varepsilon) dr \right| \\
& \quad + \sqrt{\varepsilon} \mathbb{E} \left| \int_s^t \sigma(X_r^\varepsilon) \cdot \partial_x \psi(X_r^\varepsilon, Y_r^\varepsilon) dW_r^1 \right| + \mathbb{E} \left| \int_s^t g(X_r^\varepsilon, Y_r^\varepsilon) \cdot \partial_y \psi(X_r^\varepsilon, Y_r^\varepsilon) dW_r^2 \right| \\
& \leq \sqrt{\varepsilon} \mathbb{E} |\psi(X_t^\varepsilon, Y_t^\varepsilon) - \psi(X_s^\varepsilon, Y_s^\varepsilon)| + \sqrt{\varepsilon} \mathbb{E} \left| \int_s^t \mathfrak{L}_1 \psi(X_r^\varepsilon, Y_r^\varepsilon) dr \right| \\
& \quad + \sqrt{\varepsilon} 3 \mathbb{E} \left( \int_s^t |\sigma(X_r^\varepsilon) \partial_x \psi(X_r^\varepsilon, Y_r^\varepsilon)|^2 dr \right)^{\frac{1}{2}} + 3 \mathbb{E} \left( \int_s^t |g(X_r^\varepsilon, Y_r^\varepsilon) \partial_y \psi(X_r^\varepsilon, Y_r^\varepsilon)|^2 dr \right)^{\frac{1}{2}},
\end{aligned}$$

where  $\psi$  is the solution to  $\mathfrak{L}_2 \psi(x, y) = f(x, y) - \bar{f}(x)$ . Then in view of Lemma 3.4,  $(\mathbf{H}_x^2)$ ,  $(\mathbf{H}_x^3)$ , Lemma 4.3 and Hölder's inequality, (4.7) and (5.4), there exist constants  $\iota_1, \iota_2 \in (0, 1)$ ,  $C > 0$  and  $p_1, p_2 > 0$  such that

$$\begin{aligned}
\mathcal{J}_1 & \leq \sqrt{\varepsilon} \mathbb{E} (|\partial_x \psi(X_s^\varepsilon + \iota_1(X_t^\varepsilon - X_s^\varepsilon), Y_t^\varepsilon)| |X_t^\varepsilon - X_s^\varepsilon|) \\
& \quad + \sqrt{\varepsilon} \mathbb{E} (|\partial_y \psi(X_s^\varepsilon, Y_s^\varepsilon + \iota_2(Y_t^\varepsilon - Y_s^\varepsilon))| |Y_t^\varepsilon - Y_s^\varepsilon|)
\end{aligned} \tag{5.7}$$

$$\begin{aligned}
& + \sqrt{\varepsilon} C \mathbb{E} \int_s^t (1 + |X_r^\varepsilon|^{p_1} + |Y_r^\varepsilon|^{p_2}) dr \\
& \leq C_T \left( \sqrt{\varepsilon} (\mathbb{E}|X_t^\varepsilon - X_s^\varepsilon|^2)^{\frac{1}{2}} + (\varepsilon \mathbb{E}|Y_t^\varepsilon - Y_s^\varepsilon|^2)^{\frac{1}{2}} + |t - s|^{\frac{1}{2}} \right) \leq C_T |t - s|^{\frac{1}{2}}.
\end{aligned}$$

Combining (5.5), (5.6) and (5.7), we obtain

$$\mathbb{E}|Z_t^\varepsilon - Z_s^\varepsilon| \leq C_T |t - s|^{\frac{1}{2}}.$$

□

**Lemma 5.4.** *Assume that  $(\mathbf{H}_y^1)$ – $(\mathbf{H}_y^5)$  hold and let  $\phi \in C^{2,4,2}(\mathbb{R}^{d_1+d_2+d_1})$  satisfying  $(\mathbf{H}_\phi^2)$ . Furthermore, suppose that  $\phi(x, y, \cdot) \in C_b^2(\mathbb{R}^{d_1})$  and  $\partial_y^j \partial_x^i \phi(x, y, \cdot) \in C_b(\mathbb{R}^{d_1})$  for all  $(x, y) \in \mathbb{R}^{d_1+d_2}$  and  $0 \leq 2i + j \leq 4$ . Then there exists  $C_T > 0$  such that for all  $0 \leq t \leq T$*

$$\mathbb{E} \left( \int_0^t \delta\phi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \right) \leq C_T \varepsilon^{\frac{1}{2}},$$

where  $\delta\phi(x, y, z) := \phi(x, y, z) - \int_{\mathbb{R}^{d_2}} \phi(x, y, z) \mu^x(dy)$ ,  $\forall (x, y, z) \in \mathbb{R}^{d_1+d_2+d_1}$ .

*Proof.* For any  $(x, z) \in \mathbb{R}^{2d_1}$ , let  $\psi$  be the solution to

$$\mathfrak{L}_2 \psi(x, y, z) = \delta\phi(x, y, z), \quad y \in \mathbb{R}^{d_2}.$$

It follows from Lemma 3.4 that  $\psi \in C^{2,2,2}(\mathbb{R}^{d_1+d_2+d_1})$  and that there exist constants  $C, m_1, m_2 > 0$  such that for all  $(x, y, z) \in \mathbb{R}^{d_1+d_2+d_1}$

$$\begin{aligned}
& |\psi(x, y, z)| + |\partial_x \psi(x, y, z)| + |\partial_x^2 \psi(x, y, z)| \\
& + |\partial_z \psi(x, y, z)| + |\partial_z^2 \psi(x, y, z)| + |\partial_y \psi(x, y, z)| \\
& \leq C (1 + |x|^{m_1} + |y|^{m_2}).
\end{aligned} \tag{5.8}$$

By Itô's formula, one sees that

$$\begin{aligned}
\mathbb{E}\psi(X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon) - \psi(x, y, 0) &= \mathbb{E} \int_0^t \mathfrak{L}_1 \psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds + \frac{1}{\varepsilon} \mathbb{E} \int_0^t \mathfrak{L}_2 \psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \\
& + \frac{1}{\sqrt{\varepsilon}} \mathbb{E} \int_0^t \langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(\bar{X}_s), \partial_z \psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) \rangle ds,
\end{aligned}$$

which by (5.8), Hölder's inequality,  $(\mathbf{H}_x^2)$  and  $(\mathbf{H}_x^3)$  implies that there exist  $p_1, p_2, p_3 > 0$  such that

$$\begin{aligned}
\mathbb{E} \int_0^t \delta\phi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds &= \varepsilon \left( \mathbb{E}\psi(X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon) - \psi(x, y, 0) - \mathbb{E} \int_0^t \mathfrak{L}_1 \psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \right) \\
& - \sqrt{\varepsilon} \mathbb{E} \int_0^t \langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(\bar{X}_s), \partial_z \psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) \rangle ds \\
& \leq \sqrt{\varepsilon} C \mathbb{E} \left( 1 + \sup_{t \in [0, T]} |X_t^\varepsilon|^{p_1} + \sup_{t \in [0, T]} |Y_t^\varepsilon|^{p_2} + \sup_{t \in [0, T]} |\bar{X}_t|^{p_3} \right).
\end{aligned} \tag{5.9}$$

Combining (5.9) and Lemma 4.3, we have

$$\mathbb{E} \int_0^t \delta\phi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \leq C_T \varepsilon^{\frac{1}{2}}.$$

□

Now we recall the following lemma, which is from [41, Proposition 3].

**Lemma 5.5.** *If  $\{Z^\varepsilon, \bar{Z} : 0 < \varepsilon \leq 1\} \subset C([0, T]; \mathbb{R}^{d_1})$  is tight, then for any  $\delta > 0$  there exist  $N \in \mathbb{N}$  and  $z^1, \dots, z^N \in C([0, T]; \mathbb{R}^{d_1})$  such that for any  $0 < \varepsilon \leq 1$*

$$\mathbb{P} \left( \bigcap_{k=1}^N \left\{ \sup_{t \in [0, T]} |Z_t^\varepsilon - z_t^k| > \delta \right\} \right) < \delta, \quad \mathbb{P} \left( \bigcap_{k=1}^N \left\{ \sup_{t \in [0, T]} |\bar{Z}_t - z_t^k| > \delta \right\} \right) < \delta.$$

**Remark 5.6.** (i) Note that we also have

$$GG^T(x) = \int_{\mathbb{R}^{d_2}} (f(x, y) - \bar{f}(x)) \psi(x, y) \mu^x(dy), \quad \forall x \in \mathbb{R}^{d_1}, \quad (5.10)$$

where  $\psi(x, y)$  is the solution to  $\mathfrak{L}_2(x, y)\psi(x, y) = f(x, y) - \bar{f}(x)$ ,  $y \in \mathbb{R}^{d_2}$ .

(ii) It follows from  $(\mathbf{H}_x^2)$ , (5.10) and Lemma 3.4 that there exist constants  $m > 0$  and  $C > 0$  such that for any  $x \in \mathbb{R}^{d_1}$

$$|G(x)| \vee |GG^T(x)| \vee |\nabla(GG^T(x))| \leq C(1 + |x|^m).$$

Therefore, we have for any  $x_1, x_2 \in \mathbb{R}^{d_1}$

$$|GG^T(x_1) - GG^T(x_2)| \leq C(1 + |x_1|^m + |x_2|^m)|x_1 - x_2|.$$

We are now in a position to prove Theorem 2.6.

*Proof of Theorem 2.6.* Note that it follows from Itô's formula that

$$\begin{aligned} \varphi(\bar{Z}_t) &= \varphi(0) + \int_0^t \langle \nabla \bar{f}(\bar{X}_s) \bar{Z}_s, \nabla \varphi(\bar{Z}_s) \rangle + \frac{1}{2} \text{Tr}[\nabla^2 \varphi(\bar{Z}_s) GG^T(\bar{X}_s)] ds \\ &\quad + \int_0^t \langle \nabla \varphi(\bar{Z}_s), G(\bar{X}_s) d\widetilde{W}_s^1 \rangle, \end{aligned} \quad (5.11)$$

$$\varphi(Z_t^\varepsilon) = \varphi(0) + \int_0^t \frac{1}{\sqrt{\varepsilon}} \langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(\bar{X}_s), \nabla \varphi(Z_s^\varepsilon) \rangle ds. \quad (5.12)$$

Combining (5.11) and (5.12), we have

$$\begin{aligned} &|\mathbb{E}\varphi(Z_t^\varepsilon) - \mathbb{E}\varphi(\bar{Z}_t)| \\ &= \left| \mathbb{E} \int_0^t \left( \frac{1}{\sqrt{\varepsilon}} \langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(\bar{X}_s), \nabla \varphi(Z_s^\varepsilon) \rangle - \langle \nabla \bar{f}(\bar{X}_s) \bar{Z}_s, \nabla \varphi(\bar{Z}_s) \rangle \right. \right. \\ &\quad \left. \left. - \frac{1}{2} \text{Tr}[\nabla_z^2 \varphi(\bar{Z}_s) GG^T(\bar{X}_s)] \right) ds \right| \leq \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3, \end{aligned}$$

where

$$\begin{aligned} \mathcal{I}_1 &:= \left| \mathbb{E} \int_0^t \frac{1}{\sqrt{\varepsilon}} \langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(\bar{X}_s), \nabla \varphi(Z_s^\varepsilon) \rangle - \frac{1}{2} \text{Tr}[\nabla^2 \varphi(Z_s^\varepsilon) GG^T(\bar{X}_s)] ds \right|, \\ \mathcal{I}_2 &:= \left| \mathbb{E} \int_0^t \frac{1}{\sqrt{\varepsilon}} \langle \bar{f}(X_s^\varepsilon) - \bar{f}(\bar{X}_s), \nabla \varphi(Z_s^\varepsilon) \rangle - \langle \nabla \bar{f}(\bar{X}_s) \bar{Z}_s, \nabla \varphi(\bar{Z}_s) \rangle ds \right| \\ &= \left| \mathbb{E} \int_0^t \langle \nabla \bar{f}(\bar{X}_s + \iota(X_s^\varepsilon - \bar{X}_s)) Z_s^\varepsilon, \nabla \varphi(Z_s^\varepsilon) \rangle - \langle \nabla \bar{f}(\bar{X}_s) \bar{Z}_s, \nabla \varphi(\bar{Z}_s) \rangle ds \right|, \\ \mathcal{I}_3 &:= \left| \mathbb{E} \int_0^t \frac{1}{2} \text{Tr}[\nabla^2 \varphi(Z_s^\varepsilon) GG^T(\bar{X}_s)] - \frac{1}{2} \text{Tr}[\nabla^2 \varphi(\bar{Z}_s) GG^T(\bar{X}_s)] ds \right| \end{aligned}$$

for some  $\iota \in [0, 1]$ . Thus, we only need to demonstrate that  $\lim_{\varepsilon \rightarrow 0} \mathcal{I}_1 = 0$ ,  $\lim_{\varepsilon \rightarrow 0} \mathcal{I}_2 = 0$ , and  $\lim_{\varepsilon \rightarrow 0} \mathcal{I}_3 = 0$  to establish our result. To this end, we will break the proof into three steps.

**(Step 1)** Let  $\phi(x, y, z) := \langle f(x, y), \nabla \varphi(z) \rangle$ ,  $\bar{\phi}(x, z) := \langle \bar{f}(x), \nabla \varphi(z) \rangle$ ,  $\forall (x, y, z) \in \mathbb{R}^{d_1+d_2+d_1}$ . It is obvious that  $\Psi(x, y, z) := -\langle \psi(x, y), \nabla \varphi(z) \rangle$  is the solution to

$$\mathfrak{L}_2 \Psi(x, y, z) = -(\phi(x, y, z) - \bar{\phi}(x, z)),$$

where  $\psi(x, y)$  is the solution to  $\mathfrak{L}_2 \psi(x, y) = (f(x, y) - \bar{f}(x))$ . Then by Lemma 3.4, one sees that  $\Psi \in C^{2,2,2}(\mathbb{R}^{d_1+d_2+d_1})$  and that there exist constants  $C, m_1, m_2 > 0$  such that

$$\begin{aligned} &|\Psi(x, y, z)| + |\partial_x \Psi(x, y, z)| + |\partial_x^2 \Psi(x, y, z)| \\ &\quad + |\partial_z \Psi(x, y, z)| + |\partial_z^2 \Psi(x, y, z)| + |\partial_y \Psi(x, y, z)| \end{aligned} \quad (5.13)$$

$$\leq C(1 + |x|^{m_1} + |y|^{m_2}).$$

Applying Itô's formula to  $t \mapsto \Psi(X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon)$ , we have

$$\begin{aligned} \mathbb{E}\Psi(X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon) - \mathbb{E}\Psi(x, y, 0) &= \mathbb{E} \int_0^t \left( \mathfrak{L}_1 \Psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) + \frac{1}{\varepsilon} \mathfrak{L}_2 \Psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) \right. \\ &\quad \left. + \frac{1}{\sqrt{\varepsilon}} \langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(\bar{X}_s), \partial_z \Psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) \rangle \right) ds. \end{aligned}$$

Therefore, by (5.13),  $(\mathbf{H}_x^2)$ ,  $(\mathbf{H}_x^3)$ , Hölder's inequality, Lemma 4.3 and Theorem 2.5 we obtain

$$\begin{aligned} \mathcal{J}_1 &= \left| \sqrt{\varepsilon} \left( \mathbb{E}\Psi(x, y, 0) - \mathbb{E}\Psi(X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon) + \mathbb{E} \int_0^t \mathfrak{L}_1 \Psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \right) \right. \\ &\quad \left. + \mathbb{E} \int_0^t \langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(\bar{X}_s), \partial_z \Psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) \rangle - \frac{1}{2} \text{Tr}[\nabla^2 \varphi(Z_s^\varepsilon) GG^T(X_s^\varepsilon)] ds \right| \\ &\leq C\sqrt{\varepsilon} + \left| \mathbb{E} \int_0^t \langle \bar{f}(X_s^\varepsilon) - \bar{f}(\bar{X}_s), \partial_z \Psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) \rangle \right| + \mathcal{J}_1^3 \\ &\leq C\sqrt{\varepsilon} + \left| \mathbb{E} \int_0^t |\nabla \bar{f}(\bar{X}_s + \iota(X_s^\varepsilon - \bar{X}_s))| |X_s^\varepsilon - \bar{X}_s| |\partial_z \Psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon)| ds \right| + \mathcal{J}_1^3 \\ &\leq C\sqrt{\varepsilon} + \mathcal{J}_1^3 \end{aligned} \tag{5.14}$$

for some  $\iota \in [0, 1]$ , where

$$\mathcal{J}_1^3 := \left| \mathbb{E} \int_0^t \langle f(X_s^\varepsilon, Y_s^\varepsilon) - \bar{f}(X_s^\varepsilon), \partial_z \Psi(X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) \rangle - \frac{1}{2} \text{Tr}[\nabla^2 \varphi(Z_s^\varepsilon) GG^T(X_s^\varepsilon)] ds \right|.$$

Note that

$$\begin{aligned} &\langle f(x, y) - \bar{f}(x), \partial_z \Psi(x, y, z) \rangle - \frac{1}{2} \text{Tr}[\nabla^2 \varphi(z) GG^T(x)] \\ &= \sum_{i,j=1}^{d_1} (f_i(x, y) - \bar{f}_i(x)) \psi_j(x, y) \partial_{z_j z_i}^2 \varphi(z) \\ &\quad - \sum_{i,j=1}^{d_1} \int_{\mathbb{R}^{d_2}} (f_i(x, y) - \bar{f}_i(x)) \psi_j(x, y) \partial_{z_j z_i}^2 \varphi(z) \mu^x(dy). \end{aligned}$$

By Lemma 5.4, one sees that  $\lim_{\varepsilon \rightarrow 0} \mathcal{J}_1^3 = 0$ , which by (5.14) implies that  $\lim_{\varepsilon \rightarrow 0} \mathcal{J}_1 = 0$ .

**(Step 2)** Now we show that  $\lim_{\varepsilon \rightarrow 0} \mathcal{J}_2 = 0$ . By Lemmas 5.3 and 5.5, for any  $\delta > 0$  there exist  $z^1, \dots, z^N \in C([0, T]; \mathbb{R}^{d_1})$  such that for any  $0 < \varepsilon \leq 1$

$$\mathbb{P} \left( \bigcap_{k=1}^N \left\{ \sup_{t \in [0, T]} |Z_t^\varepsilon - z_t^k| > \delta \right\} \right) < \delta, \tag{5.15}$$

$$\mathbb{P} \left( \bigcap_{k=1}^N \left\{ \sup_{t \in [0, T]} |\bar{Z}_t - z_t^k| > \delta \right\} \right) < \delta. \tag{5.16}$$

Define

$$\begin{aligned} \Omega_{1,k}^\varepsilon &:= \left\{ \sup_{t \in [0, T]} |Z_t^\varepsilon - z_t^k| < 2\delta, \sup_{t \in [0, T]} |\bar{Z}_t - z_t^k| < 2\delta \right\}, \\ \Omega_2^\varepsilon &:= \bigcap_{k=1}^N \left\{ \sup_{t \in [0, T]} |Z_t^\varepsilon - z_t^k| > \delta \right\}, \quad \Omega_3 := \bigcap_{k=1}^N \left\{ \sup_{t \in [0, T]} |\bar{Z}_t - z_t^k| > \delta \right\}. \end{aligned}$$

Let  $\tilde{\Omega}_{1,1}^\varepsilon = \Omega_{1,1}^\varepsilon$ ,  $\tilde{\Omega}_{1,k}^\varepsilon = \Omega_{1,k}^\varepsilon \setminus \left( \bigcup_{i=1}^{k-1} \tilde{\Omega}_{1,i}^\varepsilon \right)$ ,  $k = 2, \dots, N$ . It is obvious that for any  $0 < \varepsilon \leq 1$

$$\Omega = \bigcup_{k=1}^N \tilde{\Omega}_{1,k}^\varepsilon \cup \Omega_2^\varepsilon \cup \Omega_3, \quad \tilde{\Omega}_{1,i}^\varepsilon \cap \tilde{\Omega}_{1,j}^\varepsilon = \emptyset, i \neq j.$$

Therefore,

$$\begin{aligned} \mathcal{I}_2 &= \left| \mathbb{E} \int_0^t \langle \nabla \bar{f}(\bar{X}_s + \iota(X_s^\varepsilon - \bar{X}_s))Z_s^\varepsilon, \nabla \varphi(Z_s^\varepsilon) \rangle - \langle \nabla \bar{f}(\bar{X}_s)\bar{Z}_s, \nabla \varphi(\bar{Z}_s) \rangle ds \right| \\ &\leq \mathcal{I}_{2,1} + \mathcal{I}_{2,2} + \sum_{k=1}^N \mathcal{I}_{2,3}^k, \end{aligned} \quad (5.17)$$

where

$$\begin{aligned} \mathcal{I}_{2,1} &:= \left| \mathbb{E} \left[ \chi_{\Omega_2^\varepsilon} \int_0^t \langle \nabla \bar{f}(\bar{X}_s + \iota(X_s^\varepsilon - \bar{X}_s))Z_s^\varepsilon, \nabla \varphi(Z_s^\varepsilon) \rangle - \langle \nabla \bar{f}(\bar{X}_s)\bar{Z}_s, \nabla \varphi(\bar{Z}_s) \rangle ds \right] \right|, \\ \mathcal{I}_{2,2} &:= \left| \mathbb{E} \left[ \chi_{\Omega_3} \int_0^t \langle \nabla \bar{f}(\bar{X}_s + \iota(X_s^\varepsilon - \bar{X}_s))Z_s^\varepsilon, \nabla \varphi(Z_s^\varepsilon) \rangle - \langle \nabla \bar{f}(\bar{X}_s)\bar{Z}_s, \nabla \varphi(\bar{Z}_s) \rangle ds \right] \right|, \\ \mathcal{I}_{2,3}^k &:= \left| \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t \langle \nabla \bar{f}(\bar{X}_s + \iota(X_s^\varepsilon - \bar{X}_s))Z_s^\varepsilon, \nabla \varphi(Z_s^\varepsilon) \rangle - \langle \nabla \bar{f}(\bar{X}_s)\bar{Z}_s, \nabla \varphi(\bar{Z}_s) \rangle ds \right] \right|. \end{aligned}$$

In view of (5.11), Remark 5.6, the Burkholder-Davis-Gundy inequality and (4.4), for any  $p > 1$  there exists  $m_1 > 0$  such that

$$\begin{aligned} &\mathbb{E} \left| \int_0^t \langle \nabla \bar{f}(\bar{X}_s)\bar{Z}_s, \nabla \varphi(\bar{Z}_s) \rangle ds \right|^p \\ &= \mathbb{E} \left| \varphi(\bar{Z}_t) - \varphi(0) - \int_0^t \frac{1}{2} \text{Tr}[\nabla^2 \varphi(\bar{Z}_s)GG^T(\bar{X}_s)] ds - \int_0^t \langle \nabla \varphi(\bar{Z}_s), G(\bar{X}_s) d\tilde{W}_s^1 \rangle \right|^p \\ &\leq C_{T,p,\|\nabla \varphi\|_\infty} \left( \|\varphi\|_\infty^p + \int_0^T \mathbb{E} |\bar{X}_s|^{m_1} ds + 1 \right) \leq C_{p,T}. \end{aligned} \quad (5.18)$$

It follows from Hölder's inequality, (5.15) and (5.18) that

$$\begin{aligned} &\mathcal{I}_{2,1} + \mathcal{I}_{2,2} \\ &\leq \mathbb{E} \left[ (\chi_{\Omega_2^\varepsilon} + \chi_{\Omega_3}) \left| \int_0^t \langle \nabla \bar{f}(\bar{X}_s + \iota(X_s^\varepsilon - \bar{X}_s))Z_s^\varepsilon, \nabla \varphi(Z_s^\varepsilon) \rangle ds \right| \right] \\ &\quad + \mathbb{E} \left[ (\chi_{\Omega_2^\varepsilon} + \chi_{\Omega_3}) \left| \int_0^t \langle \nabla \bar{f}(\bar{X}_s)\bar{Z}_s, \nabla \varphi(\bar{Z}_s) \rangle ds \right| \right] \\ &\leq \left[ (\mathbb{P}(\{\Omega_2^\varepsilon\}))^{\frac{1}{3}} + (\mathbb{P}(\{\Omega_3\}))^{\frac{1}{3}} \right] \left( \mathbb{E} \left| \int_0^t \langle \nabla \bar{f}(\bar{X}_s + \iota(X_s^\varepsilon - \bar{X}_s))Z_s^\varepsilon, \nabla \varphi(Z_s^\varepsilon) \rangle ds \right|^{\frac{3}{2}} \right)^{\frac{2}{3}} \\ &\quad + \left[ (\mathbb{P}(\{\Omega_2^\varepsilon\}))^{\frac{1}{2}} + (\mathbb{P}(\{\Omega_3\}))^{\frac{1}{2}} \right] \left( \mathbb{E} \left| \int_0^t \langle \nabla \bar{f}(\bar{X}_s)\bar{Z}_s, \nabla \varphi(\bar{Z}_s) \rangle ds \right|^2 \right)^{\frac{1}{2}} \\ &\leq C_T \delta^{\frac{1}{3}} \left( \mathbb{E} \int_0^T |\nabla \bar{f}(\bar{X}_s + \iota(X_s^\varepsilon - \bar{X}_s))|^{\frac{3}{2}} |Z_s^\varepsilon|^{\frac{3}{2}} \|\nabla \varphi\|_\infty^{\frac{3}{2}} ds \right)^{\frac{2}{3}} + C_T \delta^{\frac{1}{2}}, \end{aligned}$$

which by Hölder's inequality and  $(\mathbf{H}_x^2)$  implies that

$$\mathcal{I}_{2,1} + \mathcal{I}_{2,2} \leq C_T \delta^{\frac{1}{3}} \left( \mathbb{E} \int_0^T |Z_s^\varepsilon|^2 + |\bar{X}_s|^{4\theta_1} + |X_s^\varepsilon|^{4\theta_1} ds \right)^{\frac{2}{3}} + C_T \delta^{\frac{1}{2}}. \quad (5.19)$$

Combining (5.19), (4.3), (4.4) and (2.3), we get

$$\mathcal{I}_{2,1} + \mathcal{I}_{2,2} \leq C_T \delta^{\frac{1}{3}}. \quad (5.20)$$

Note that for all  $1 \leq k \leq N$

$$\begin{aligned} \mathcal{J}_{2,3}^k &:= \left| \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t \langle (\nabla \bar{f}(\bar{X}_s + \iota(X_s^\varepsilon - \bar{X}_s)) - \nabla \bar{f}(\bar{X}_s)) Z_s^\varepsilon, \nabla \varphi(Z_s^\varepsilon) \rangle ds \right] \right| \\ &\quad + \left| \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t \langle \nabla \bar{f}(\bar{X}_s) Z_s^\varepsilon, \nabla \varphi(Z_s^\varepsilon) - \nabla \varphi(\bar{Z}_s) \rangle ds \right] \right| \\ &\quad + \left| \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t \langle \nabla \bar{f}(\bar{X}_s) (Z_s^\varepsilon - \bar{Z}_s), \nabla \varphi(\bar{Z}_s) \rangle ds \right] \right| =: \sum_{i=1}^3 \mathfrak{J}_i. \end{aligned} \quad (5.21)$$

First of all, by Hölder's inequality,  $(\mathbf{H}_x^2)$ , Lemma 4.3 and Theorem 2.5 we have

$$\begin{aligned} \mathfrak{J}_1 &\leq \|\nabla \varphi\|_\infty \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t |\nabla^2 \bar{f}(\bar{X}_s + \iota'(X_s^\varepsilon - \bar{X}_s))| |X_s^\varepsilon - \bar{X}_s| |Z_s^\varepsilon| \|\nabla \varphi\|_\infty ds \right] \\ &\leq \|\nabla \varphi\|_\infty \left( \mathbb{E} \int_0^T |Z_s^\varepsilon|^2 ds \right)^{\frac{1}{2}} \left( \mathbb{E} \int_0^T |X_s^\varepsilon - \bar{X}_s|^2 ds \right)^{\frac{1}{4}} \\ &\quad \times \left( \mathbb{E} \int_0^T |\nabla^2 \bar{f}(\bar{X}_s + \iota'(X_s^\varepsilon - \bar{X}_s))|^4 (|X_s^\varepsilon| + |\bar{X}_s|)^2 ds \right)^{\frac{1}{4}} \leq C_T \varepsilon^{\frac{1}{4}}, \end{aligned} \quad (5.22)$$

where  $\iota' \in [0, 1]$ . And in view of Hölder's inequality we have

$$\begin{aligned} \mathfrak{J}_2 + \mathfrak{J}_3 &\leq \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t |\nabla \bar{f}(\bar{X}_s)| (|Z_s^\varepsilon| \|\nabla^2 \varphi\|_\infty + \|\nabla \varphi\|_\infty) |Z_s^\varepsilon - \bar{Z}_s| ds \right] \\ &\leq C \delta \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^T |\nabla \bar{f}(\bar{X}_s)| (|Z_s^\varepsilon| + 1) ds \right]. \end{aligned} \quad (5.23)$$

Therefore, by (5.17), (5.20), (5.21), (5.22), (5.23), Hölder's inequality,  $(\mathbf{H}_x^2)$ , Lemma 4.3 and Theorem 2.5 one sees that

$$\mathcal{J}_2 \leq C_T \left( \delta^{\frac{1}{3}} + N \varepsilon^{\frac{1}{4}} \right) + \delta C \mathbb{E} \int_0^T |\nabla \bar{f}(\bar{X}_s)| (|Z_s^\varepsilon| + 1) ds \leq C_T \left( \delta^{\frac{1}{4}} + N \varepsilon^{\frac{1}{4}} \right),$$

which implies that  $\lim_{\varepsilon \rightarrow 0} \mathcal{J}_2 = 0$  by first letting  $\varepsilon \rightarrow 0$  and then letting  $\delta \rightarrow 0$ .

**(Step 3)** Now we estimate  $\mathcal{J}_3$ :

$$\begin{aligned} \mathcal{J}_3 &:= \left| \mathbb{E} \int_0^t \frac{1}{2} \text{Tr}[\nabla^2 \varphi(Z_s^\varepsilon) G G^T(X_s^\varepsilon)] - \frac{1}{2} \text{Tr}[\nabla^2 \varphi(\bar{Z}_s) G G^T(\bar{X}_s)] ds \right| \\ &\leq \left| \mathbb{E} \left[ \chi_{\Omega_2^\varepsilon} \int_0^t \frac{1}{2} \text{Tr}[\nabla^2 \varphi(Z_s^\varepsilon) G G^T(X_s^\varepsilon)] - \frac{1}{2} \text{Tr}[\nabla^2 \varphi(\bar{Z}_s) G G^T(\bar{X}_s)] ds \right] \right| \\ &\quad + \left| \mathbb{E} \left[ \chi_{\Omega_3} \int_0^t \frac{1}{2} \text{Tr}[\nabla^2 \varphi(Z_s^\varepsilon) G G^T(X_s^\varepsilon)] - \frac{1}{2} \text{Tr}[\nabla^2 \varphi(\bar{Z}_s) G G^T(\bar{X}_s)] ds \right] \right| \\ &\quad + \sum_{k=1}^N \left| \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t \frac{1}{2} \text{Tr}[\nabla^2 \varphi(Z_s^\varepsilon) G G^T(X_s^\varepsilon)] - \frac{1}{2} \text{Tr}[\nabla^2 \varphi(\bar{Z}_s) G G^T(\bar{X}_s)] ds \right] \right| \\ &=: \mathcal{J}_{3,1} + \mathcal{J}_{3,2} + \sum_{k=1}^N \mathcal{J}_{3,3}^k. \end{aligned} \quad (5.24)$$

Thanks to Hölder's inequality, (5.15), (5.16), Remark 5.6, (4.3) and (4.4), we have

$$\mathcal{J}_{3,1} + \mathcal{J}_{3,2} \leq C_T \delta^{1/2}. \quad (5.25)$$

And it follows from Remark 5.6, Hölder's inequality, Lemma 4.3 and Theorem 2.5 that there exists  $m > 0$  such that

$$\begin{aligned}
\mathcal{I}_{3,3}^k &: \leq \left| \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t \frac{1}{2} \text{Tr} [(\nabla^2 \varphi(Z_s^\varepsilon) - \nabla^2 \varphi(z_s^k)) GG^T(X_s^\varepsilon)] ds \right] \right| \\
&+ \left| \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t \frac{1}{2} \text{Tr} [\nabla^2 \varphi(z_s^k) GG^T(X_s^\varepsilon)] - \frac{1}{2} \text{Tr} [\nabla^2 \varphi(z_s^k) GG^T(\bar{X}_s)] ds \right] \right| \\
&+ \left| \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t \frac{1}{2} \text{Tr} [(\nabla^2 \varphi(\bar{Z}_s) - \nabla^2 \varphi(z_s^k)) GG^T(\bar{X}_s)] ds \right] \right| \\
&\leq C \left| \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t \|\nabla^3 \varphi\|_\infty |Z_s^\varepsilon - z_s^k| (1 + |X_s^\varepsilon|^m) ds \right] \right| \tag{5.26} \\
&+ \left| \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t \|\nabla^2 \varphi\|_\infty (1 + |X_s^\varepsilon|^m + |\bar{X}_s|^m) |X_s^\varepsilon - \bar{X}_s| ds \right] \right| \\
&+ \left| \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^t \|\nabla^3 \varphi\|_\infty |\bar{Z}_s - z_s^k| (1 + |\bar{X}_s|^m) ds \right] \right| \\
&\leq C_T \delta \left( \mathbb{E} \left[ \chi_{\tilde{\Omega}_{1,k}^\varepsilon} \int_0^T (1 + |X_s^\varepsilon|^m + |\bar{X}_s|^m) ds \right] \right) + C_T \varepsilon^{\frac{1}{2}}.
\end{aligned}$$

Therefore, employing (5.24), (5.25), (5.26), Hölder's inequality,  $(\mathbf{H}_x^2)$  and Lemma 4.3, we have

$$\mathcal{I}_3 \leq C_T \left( \delta^{1/2} + N \varepsilon^{\frac{1}{2}} \right) + C_T \delta \left( \mathbb{E} \int_0^T (1 + |X_s^\varepsilon|^m + |\bar{X}_s|^m) ds \right) \leq C_T (\delta^{1/2} + N \varepsilon^{1/2}).$$

Therefore, firstly letting  $\varepsilon \rightarrow 0$ , then taking  $\delta \rightarrow 0$ , we obtain  $\lim_{\varepsilon \rightarrow 0} \mathcal{I}_3 = 0$ .  $\square$

## 6. THE GLOBAL AVERAGING PRINCIPLE AND SECOND AVERAGING PRINCIPLE

In Section 6.1, we begin with recalling some well-known definitions and results about autonomous and nonautonomous dynamical systems (see e.g. [33]). Subsequently, we prove Theorem 2.8 in Section 6.2. Finally, we investigate the second averaging principle in Section 6.3.

**6.1. Preliminaries for dynamical systems.** Let  $(\mathcal{X}, d)$  be a complete metric space, and  $(\mathcal{P}, d_{\mathcal{P}})$  be a metric space.

**Definition 6.1.** Let  $T = \mathbb{R}$  or  $\mathbb{Z}$ . A *semi-dynamical system* is defined as a continuous function  $\phi : T^+ \times \mathcal{X} \rightarrow \mathcal{X}$  that satisfies  $\phi(0, x) = x$ ,  $\forall x \in \mathcal{X}$ , and  $\phi(t + s, x) = \phi(t, \phi(s, x))$  for all  $s, t \in T^+$  and  $x \in \mathcal{X}$ .

**Definition 6.2.** We say a nonempty compact subset  $A \subset \mathcal{X}$  is a *global attractor* of a semi-dynamical system  $\phi$  on  $\mathcal{X}$  if it is  $\phi$ -invariant and attracts bounded sets, i.e.  $\phi(t, A) = A$  for all  $t \in T^+$  and  $\lim_{t \rightarrow \infty} \text{dist}(\phi(t, D), A) = 0$  for any bounded subset  $D \subset \mathcal{X}$ .

**Definition 6.3.** A *nonautonomous dynamical system*  $(\sigma, \varphi)$  (in short,  $\varphi$ ) comprises two components:

- (i) A *dynamical system*  $\sigma$  on  $\mathcal{P}$  with time set  $T = \mathbb{Z}$  or  $\mathbb{R}$ , i.e.  $\sigma_0(\cdot) = \text{Id}_{\mathcal{P}}$ ,  $\sigma_{t+s}(p) = \sigma_t(\sigma_s(p))$  for all  $t, s \in T$  and  $p \in \mathcal{P}$ , and the mapping  $(t, p) \mapsto \sigma_t(p)$  is continuous.
- (ii) A *cocycle*  $\varphi : T^+ \times \mathcal{P} \times \mathcal{X} \rightarrow \mathcal{X}$  satisfies
  - (1)  $\varphi(0, p, x) = x$  for all  $(p, x) \in \mathcal{P} \times \mathcal{X}$ ,
  - (2)  $\varphi(t + s, p, x) = \varphi(t, \sigma_s(p), \varphi(s, p, x))$  for all  $s, t \in T^+$  and  $(p, x) \in \mathcal{P} \times \mathcal{X}$ ,
  - (3) the mapping  $(t, p, x) \mapsto \varphi(t, p, x)$  is continuous.

Here  $\mathcal{P}$  is called the *base or parameter space* and  $\mathcal{X}$  is the *fiber or state space*. For convenience, we also write  $\sigma_t(p)$  as  $\sigma_t p$ .

Furthermore, if  $\sigma_t(p)(\cdot) := p(t + \cdot)$  then  $(\mathcal{P}, \mathbb{R}, \sigma)$  is called a shift dynamical system or Bebutov shift flow.

**Definition 6.4.** The autonomous semi-dynamical system  $\pi$  on  $\mathcal{P} \times \mathcal{X}$  defined by

$$\pi : T^+ \times \mathcal{P} \times \mathcal{X} \rightarrow \mathcal{P} \times \mathcal{X}, \quad \pi(t, (p, x)) := (\theta_t(p), \varphi(t, p, x))$$

is called the *skew product flow* associated with the nonautonomous dynamical system  $(\theta, \varphi)$ .

**Definition 6.5.** Let  $(\sigma, \phi)$  be a skew product flow on a metric phase space  $(\mathcal{X}, d)$  with base set  $\mathcal{P}$ . A subset  $\mathcal{M}$  of the extended phase space  $\mathcal{P} \times \mathcal{X}$  is called a *nonautonomous set*, and for each  $p \in \mathcal{P}$ , the set  $M_p := \{x \in \mathcal{X} : (p, x) \in \mathcal{M}\}$  is called the *p-fiber* of  $\mathcal{M}$ . In general,  $\mathcal{M}$  is said to have a topological property (such as compactness or closeness) if each fiber of  $\mathcal{M}$  has this property.

**Definition 6.6.** Let  $(\sigma, \phi)$  be a skew product flow on a metric space  $(\mathcal{X}, d)$  with base set  $\mathcal{P}$ . A family  $\mathcal{A} = \{A_p\}_{p \in \mathcal{P}}$  of nonempty subsets of  $\mathcal{X}$  is called *invariant* w.r.t.  $(\sigma, \phi)$ , or  *$\phi$ -invariant*, if  $\phi(t, p, A_p) = A_{\sigma_t(p)}$  for all  $t \geq 0$  and  $p \in \mathcal{P}$ .

**Remark 6.7.** The compact set-valued mapping  $t \mapsto A_{\sigma_t(p)}$ , induced by a  $\phi$ -invariant family  $(A_p)_{p \in \mathcal{P}}$  of compact subsets, is continuous in  $t \in \mathbb{R}$  w.r.t. the Hausdorff metric for each fixed  $p \in \mathcal{P}$ .

**Definition 6.8.** Let  $(\sigma, \phi)$  be a skew product flow. A nonempty, compact and invariant nonautonomous set  $\mathcal{A}$  is called a *pullback attractor* if the pullback convergence

$$\lim_{t \rightarrow \infty} \text{dist}_{\mathcal{X}}(\phi(t, \sigma_{-t}(p), D), A_p) = 0$$

holds for every nonempty bounded subset  $D$  of  $\mathcal{X}$  and  $p \in \mathcal{P}$ .

**Lemma 6.9.** Assume that a semi-dynamical system  $\phi$  on  $\mathcal{X}$  has an absorbing set  $B$ , i.e. for any bounded subset  $D \subset \mathcal{X}$ , there exists  $T = T(D) \in T^+$  such that  $\phi(t, D) \subset B$  for all  $t \geq T$ . If  $B$  is positively invariant, i.e.  $\phi(t, B) \subset B$  for all  $t \in T^+$ , then  $\phi$  has a unique attractor  $A \subset B$  defined by  $A = \bigcap_{t \geq 0} \phi(t, B)$ .

**Definition 6.10.** Let  $(\sigma, \phi)$  be a skew product flow on  $\mathcal{X}$ . A nonempty compact subset  $B$  of  $\mathcal{X}$  is called *pullback absorbing* if for each  $p \in \mathcal{P}$  and every bounded subset  $D$  of  $\mathcal{X}$ , there exists a  $T = T(p, D) > 0$  such that  $\phi(t, \sigma_{-t}(p), D) \subset B$  for all  $t \geq T$ .

**Lemma 6.11.** Let  $(\sigma, \phi)$  be a skew product flow on  $\mathcal{X}$  with a compact pullback absorbing set  $B$  such that  $\phi(t, p, B) \subset B$  for all  $t \geq 0$  and  $p \in \mathcal{P}$ . Then there exists a pullback attractor  $\mathcal{A}$  with fibers in  $B$  uniquely determined by  $A_p = \bigcap_{\tau \geq 0} \overline{\bigcup_{t \geq \tau} \phi(t, \sigma_{-t}(p), B)}$  for all  $p \in \mathcal{P}$ .

**6.2. Proof of Theorem 2.8.** In order to prove Theorem 2.8, we need the following decay estimates of solutions to (1.3) and (1.4) under the dissipativity condition  $(\mathbf{H}_x^5)$ .

**Lemma 6.12.** Assume that  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$  and  $(\mathbf{H}_x^2)$ - $(\mathbf{H}_x^5)$  hold. If  $g \in C_b(\mathbb{R}^{d_1+d_2})$ , then for any  $(\zeta^\varepsilon, \xi^\varepsilon) \in \mathcal{L}^2(\Omega, \mathcal{F}_s, \mathbb{P}; \mathbb{R}^{d_1+d_2})$  there exists a unique solution  $V_{s,t}^\varepsilon(\zeta^\varepsilon, \xi^\varepsilon) := (X_{s,t}^\varepsilon(\zeta^\varepsilon), Y_{s,t}^\varepsilon(\xi^\varepsilon))$  to (2.4). Furthermore, there exists  $\varepsilon_0 > 0$  such that for all  $p \geq 1$  and  $0 < \varepsilon \leq \varepsilon_0$

$$\mathbb{E}|Y_{s,t}^\varepsilon(\xi^\varepsilon)|^{2p} \leq \mathbb{E}|\xi^\varepsilon|^{2p} e^{-\frac{1}{4}\varepsilon^{-2\alpha}\eta p(t-s)} + \varpi(p) \left(1 - e^{-\frac{1}{4}\varepsilon^{-2\alpha}\eta p(t-s)}\right), \quad (6.1)$$

and

$$\begin{aligned} \mathbb{E}|X_{s,t}^\varepsilon(\zeta^\varepsilon)|^2 &\leq \mathbb{E}|\zeta^\varepsilon|^2 e^{-\lambda_1(t-s)} + \frac{K_5 \varpi(\theta/2) + K_4}{\lambda_1} \left(1 - e^{-\lambda_1(t-s)}\right) \\ &\quad + K_5 \mathbb{E}|\xi^\varepsilon|^\theta \left( \left[ \lambda_1^{-1} \left(1 - e^{-\lambda_1(t-s)}\right) \right] \wedge e^{-\lambda_1(t-s)} \right), \end{aligned}$$

where  $\varpi(p) := (8K_1 + 8(p-1)\|g\|_\infty)^p \eta^{-p}$ ,  $\forall p \geq 1$ .

Moreover, for any  $1 < p < \frac{\lambda_1}{2L^2} + 1$  there exists a constant  $C > 0$  such that

$$\mathbb{E}|X_{s,t}^\varepsilon(\zeta^\varepsilon)|^{2p} \leq \mathbb{E}|\zeta^\varepsilon|^{2p} e^{-\frac{p}{2}(\lambda_1 - 2(p-1)L^2)(t-s)} + C(\mathbb{E}|\xi^\varepsilon|^{\theta p} + 1). \quad (6.2)$$

*Proof.* By (4.5) and Gronwall's inequality, one sees that there exists a  $\varepsilon_0 > 0$  such that for all  $0 < \varepsilon \leq \varepsilon_0$

$$\begin{aligned} \mathbb{E}|Y_{s,t}^\varepsilon(\xi^\varepsilon)|^{2p} &\leq \mathbb{E}|\xi^\varepsilon|^{2p} e^{-\frac{1}{4}\varepsilon^{-2\alpha}\eta p(t-s)} \\ &\quad + (8K_1 + 8(p-1)\|g\|_\infty)^p \eta^{-p} \left(1 - e^{-\frac{1}{4}\varepsilon^{-2\alpha}\eta p(t-s)}\right). \end{aligned}$$

It follows from Itô's formula,  $(\mathbf{H}_x^3)$ ,  $(\mathbf{H}_x^5)$  and Young's inequality that

$$\begin{aligned} &\mathbb{E}|X_{s,t}^\varepsilon(\zeta^\varepsilon)|^{2p} \\ &\leq \mathbb{E}|\zeta^\varepsilon|^{2p} + p\mathbb{E} \int_s^t |X_{s,\tau}^\varepsilon(\zeta^\varepsilon)|^{2p-2} \left(-\lambda_1 |X_{s,\tau}^\varepsilon(\zeta^\varepsilon)|^2 + K_5 |Y_{s,\tau}^\varepsilon(\xi)|^\theta + K_4\right) d\tau \\ &\quad + 2p(p-1)\mathbb{E} \int_s^t |X_{s,\tau}^\varepsilon(\zeta^\varepsilon)|^{2p-2} \left(L_\sigma^2 |X_{s,\tau}^\varepsilon(\zeta^\varepsilon)|^2 + K_7^2 + 2L_\sigma K_7 |X_{s,\tau}^\varepsilon(\zeta^\varepsilon)|\right) d\tau. \end{aligned}$$

If  $p = 1$ , in view of Gronwall's inequality and (6.1), one sees that there exists  $\varepsilon_0 > 0$  such that for all  $0 < \varepsilon \leq \varepsilon_0$

$$\mathbb{E}|X_{s,t}^\varepsilon(\zeta^\varepsilon)|^2 \leq \mathbb{E}|\zeta^\varepsilon|^2 e^{-\lambda_1(t-s)} + \frac{1}{\lambda_1} \left(K_5 (\mathbb{E}|\xi^\varepsilon|^\theta + \varpi(\theta/2)) + K_4\right) \left(1 - e^{-\lambda_1(t-s)}\right).$$

On the other hand, in view of Gronwall's inequality and (6.1), we obtain that there exists  $\varepsilon_0 > 0$  such that for all  $0 < \varepsilon \leq \varepsilon_0$

$$\begin{aligned} \mathbb{E}|X_{s,t}^\varepsilon(\zeta^\varepsilon)|^2 &\leq \mathbb{E}|\zeta^\varepsilon|^2 e^{-\lambda_1(t-s)} + \frac{K_5 \varpi(\theta/2) + K_4}{\lambda_1} \left(1 - e^{-\lambda_1(t-s)}\right) \\ &\quad + K_5 \mathbb{E}|\xi^\varepsilon|^\theta e^{-\lambda_1(t-s)} \left(\frac{1}{8}\varepsilon^{-2\alpha\eta\theta} - \lambda_1\right)^{-1} \left(1 - \exp\left\{\left(\lambda_1 - \frac{\eta\theta}{8\varepsilon^{2\alpha}}\right)(t-s)\right\}\right) \\ &\leq \mathbb{E}|\zeta^\varepsilon|^2 e^{-\lambda_1(t-s)} + K_5 \mathbb{E}|\xi^\varepsilon|^\theta e^{-\lambda_1(t-s)} + \frac{K_5 \varpi(\theta/2) + K_4}{\lambda_1} \left(1 - e^{-\lambda_1(t-s)}\right). \end{aligned}$$

In the case where  $1 < p < \frac{\lambda_1}{2L_\sigma^2} + 1$ , by Young's inequality we have

$$\begin{aligned} &\mathbb{E}|X_{s,t}^\varepsilon(\zeta^\varepsilon)|^{2p} \\ &\leq \mathbb{E}|\zeta^\varepsilon|^{2p} + p\mathbb{E} \int_s^t \left(-\frac{1}{2}(\lambda_1 - 2(p-1)L_\sigma^2) |X_{s,\tau}^\varepsilon(\zeta^\varepsilon)|^{2p} + C |Y_{s,\tau}^\varepsilon(\xi^\varepsilon)|^{\theta p} + C\right) d\tau, \end{aligned}$$

which by Gronwall's inequality and (6.1) implies that

$$\mathbb{E}|X_{s,t}^\varepsilon(\zeta^\varepsilon)|^{2p} \leq \mathbb{E}|\zeta^\varepsilon|^{2p} e^{-\frac{p}{2}(\lambda_1 - 2(p-1)L_\sigma^2)(t-s)} + C(\mathbb{E}|\xi^\varepsilon|^{\theta p} + 1).$$

□

**Lemma 6.13.** *Assume that  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^2)$ ,  $(\mathbf{H}_x^5)$  and  $(\mathbf{A}_f)$  hold. Then there exists a constant  $K_8 > 0$  such that for all  $x \in \mathbb{R}^{d_1}$*

$$2\langle \bar{f}(x), x \rangle + |\bar{\sigma}(x)|^2 \leq -\lambda_1 |x|^2 + K_8.$$

*Proof.* Let  $K' := \sup_{x \in \mathbb{R}^{d_1}} \int_{\mathbb{R}^{d_2}} |y|^\theta \mu^x(dy) < \infty$ . It follows from  $(\mathbf{H}_x^3)$ ,  $(\mathbf{H}_x^5)$ , Hölder's inequality and Young's inequality that for any  $x \in \mathbb{R}^{d_1}$

$$\begin{aligned} &2\langle \bar{f}(x), x \rangle + |\bar{\sigma}(x)|_{HS}^2 \\ &= 2 \left\langle \bar{f}(x) - \frac{1}{T} \int_0^T \hat{f}(s, x) ds, x \right\rangle + 2 \left\langle \frac{1}{T} \int_0^T \hat{f}(s, x) ds, x \right\rangle + \left| \frac{1}{T} \int_0^T \sigma(s, x) ds \right|^2 \\ &\quad + \left| \bar{\sigma}(x) - \frac{1}{T} \int_0^T \sigma(s, x) ds \right|^2 + 2 \left\langle \bar{\sigma}(x) - \frac{1}{T} \int_0^T \sigma(s, x) ds, \frac{1}{T} \int_0^T \sigma(s, x) ds \right\rangle \end{aligned}$$

$$\begin{aligned}
&\leq \frac{2}{T} \left| \int_0^T (\bar{f}(x) - \hat{f}(s, x)) ds \right| \cdot |x| + \frac{1}{T} \int_0^T \int_{\mathbb{R}^{d_2}} (2 \langle f(s, x, y), x \rangle + |\sigma(s, x)|^2) \mu^x(dy) ds \\
&\quad + \frac{1}{T} \int_0^T |\bar{\sigma}(x) - \sigma(s, x)|^2 ds + 2 \left| \bar{\sigma}(x) - \frac{1}{T} \int_0^T \sigma(s, x) ds \right| \left| \frac{1}{T} \int_0^T \sigma(s, x) ds \right| \\
&\leq -\lambda_1 |x|^2 + K_4 + K'K_5 + 1 + \frac{2}{T} \left| \int_0^T (\bar{f}(x) - \hat{f}(s, x)) ds \right| \cdot |x| \\
&\quad + \frac{1}{T} \int_0^T |\bar{\sigma}(x) - \sigma(s, x)|^2 ds + \frac{1}{T} \int_0^T |\bar{\sigma}(x) - \sigma(s, x)|^2 ds (2L_\sigma^2 |x|^2 + 2K_7^2),
\end{aligned}$$

which by  $(\mathbf{A}_f)$  and  $(\mathbf{A}_\sigma)$  implies that

$$2 \langle \bar{f}(x), x \rangle + |\bar{\sigma}(x)|_{HS}^2 \leq -\lambda_1 |x|^2 + K_4 + K'K_5 + 1$$

by letting  $T \rightarrow \infty$ .  $\square$

Similarly to the estimates provided in Lemma 6.12, the following lemma can be derived by applying the same methodology. Therefore, we omit the proof.

**Lemma 6.14.** *Assume that  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$  and  $(\mathbf{H}_x^2)$ – $(\mathbf{H}_x^5)$  hold. If  $g \in C_b(\mathbb{R}^{d_1+d_2})$ , then for any  $\zeta \in \mathcal{L}^2(\Omega, \mathcal{F}_s, \mathbb{P}; \mathbb{R}^{d_1})$ , there exists a unique solution  $\bar{X}_{s,t}(\zeta)$  to (1.4). Furthermore, there exists  $\varepsilon_0 > 0$  such that for all  $0 < \varepsilon \leq \varepsilon_0$*

$$\mathbb{E}|\bar{X}_{s,t}(\zeta)|^2 \leq \mathbb{E}|\zeta|^2 e^{-\lambda_1(t-s)} + \frac{K_8}{\lambda_1} (1 - e^{-\lambda_1(t-s)}).$$

Moreover, if  $\zeta \in \mathcal{L}^{2p}(\Omega, \mathcal{F}_s, \mathbb{P}; \mathbb{R}^{d_1})$  for any  $1 < p < \frac{\lambda_1}{2L_\sigma^2} + 1$ , then there exists a constant  $C > 0$  such that

$$\mathbb{E}|\bar{X}_{s,t}(\zeta)|^{2p} \leq \mathbb{E}|\zeta|^{2p} e^{-\frac{p}{2}(\lambda_1 - 2(p-1)L_\sigma^2)(t-s)} + C.$$

**Remark 6.15.** Assume that  $f$  and  $\sigma$  satisfy  $(\mathbf{A}_f)$  and  $(\mathbf{A}_\sigma)$ . Note that by [9, Lemma 5.10] one sees that for any  $\tilde{\mathbb{F}} := (\tilde{F}, \tilde{\sigma}) \in \mathcal{H}(\mathbb{F})$ ,  $\tilde{F}$  and  $\tilde{\sigma}$  satisfy  $(\mathbf{A}_f)$  and  $(\mathbf{A}_\sigma)$  provided  $\mathcal{H}(\mathbb{F})$  is compact.

Following a similar approach as in the proof of [8, Theorem 3.1], we establish the following lemma for the continuous dependence of solutions to (2.4) on initial values and coefficients. For convenience, we assume  $\varepsilon = 1$  in this lemma without loss of generality.

**Lemma 6.16.** *Assume that  $\xi^n, \xi \in \mathcal{L}^2(\Omega, \mathcal{F}_s, \mathbb{P}; \mathbb{R}^{d_1+d_2})$ . For any  $n \in \mathbb{N}$  let  $V_{s,t}^n(\xi^n)$  satisfy*

$$V_{s,t}^n(\xi^n) = \xi^n + \int_s^t F^n(r, V_{s,r}^n(\xi^n)) dr + \int_s^t G^n(r, V_{s,r}^n(\xi^n)) dW_r, \quad t \geq s,$$

and  $V_{s,t}(\xi)$  satisfy

$$V_{s,t}(\xi) = \xi + \int_s^t F(r, V_{s,r}(\xi)) dr + \int_s^t G(r, V_{s,r}(\xi)) dW_r, \quad t \geq s.$$

Suppose that  $F^n, G^n, F$  and  $G$  satisfy  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$  and  $(\mathbf{H}_x^2)$ – $(\mathbf{H}_x^5)$ . Furthermore, assume that  $\lim_{n \rightarrow \infty} F^n(t, v) = F(t, v)$  and  $\lim_{n \rightarrow \infty} G^n(t, v) = G(t, v)$  for all  $t \in \mathbb{R}$  and  $v \in \mathbb{R}^{d_1+d_2}$ . If  $\lim_{n \rightarrow \infty} d_{BL}(\mathcal{L}(\xi^n), \mathcal{L}(\xi)) = 0$ , then for any  $t \geq s$

$$\lim_{n \rightarrow \infty} \sup_{r \in [s, t]} d_{BL}(\mathcal{L}(V_{s,r}^n(\xi^n)), \mathcal{L}(V_{s,r}(\xi))) = 0.$$

**Lemma 6.17.** *Suppose that  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$  and  $(\mathbf{H}_x^2)$ – $(\mathbf{H}_x^5)$  hold. If  $g \in C_b(\mathbb{R}^{d_1+d_2})$ , then for any  $0 < \varepsilon \leq 1$ ,  $(\sigma, P_\varepsilon^*)$  is a skew product flow on the phase space  $(\mathcal{P}_{2,\theta}(\mathbb{R}^{d_1+d_2}), d_{BL})$  with base space  $\mathcal{H}(\mathbb{F}_\varepsilon)$ , where  $\sigma : \mathbb{R}_+ \times \mathcal{H}(\mathbb{F}_\varepsilon) \rightarrow \mathcal{H}(\mathbb{F}_\varepsilon)$  is defined by  $\sigma_t(\tilde{\mathbb{F}}_\varepsilon) := \tilde{\mathbb{F}}_\varepsilon(t+, \cdot)$  for all  $(t, \tilde{\mathbb{F}}_\varepsilon) \in \mathbb{R}_+ \times \mathcal{H}(\tilde{\mathbb{F}}_\varepsilon)$ .*

*Proof.* By Remark 2.7 and Lemma 6.12, one sees that

$$P_\varepsilon^*(t, \tilde{\mathbb{F}}_\varepsilon, \cdot) : \mathcal{P}_{2,\theta}(\mathbb{R}^{d_1+d_2}) \rightarrow \mathcal{P}_{2,\theta}(\mathbb{R}^{d_1+d_2})$$

for any  $t \geq 0$ ,  $\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)$  and  $0 < \varepsilon \leq \varepsilon_0$ . Employing Lemma 6.16, the uniqueness in law of the solutions to (2.4) follows, which completes the proof.  $\square$

With the help of the aforementioned results, we are now in a position to prove Theorem 2.8.

*Proof of Theorem 2.8.* (i) For any  $R > 0$  and  $r \geq 1$ , define

$$\begin{aligned} \mathcal{D}_R^{1,r} &:= \left\{ \mu \in \mathcal{P}_r(\mathbb{R}^{d_1}) : \int_{\mathbb{R}^{d_1}} |x|^r \mu(dx) \leq R \right\}, \\ \mathcal{D}_R^{2,r} &:= \left\{ \mu \in \mathcal{P}_r(\mathbb{R}^{d_2}) : \int_{\mathbb{R}^{d_2}} |y|^r \mu(dy) \leq R \right\}. \end{aligned}$$

Set  $\varpi_1 := \lambda_1^{-1} (K_5 (2\varpi(\theta/2) + 1) + K_4) + 1$  and  $\varpi_2 := \varpi(\theta/2) + 1$ . Let  $\varpi := \varpi_1 + \varpi_2$ . Define

$$B := \overline{\left\{ m \in \mathcal{P}_{2,\theta}(\mathbb{R}^{d_1+d_2}) : \int_{\mathbb{R}^{d_1+d_2}} |z|^2 m(dz) \leq \varpi, m \circ \pi_1^{-1} \in \mathcal{D}_{\varpi_1}^{1,2}, m \circ \pi_2^{-1} \in \mathcal{D}_{\varpi_2}^{2,\theta} \right\}}. \quad (6.3)$$

It can be verified that  $B$  is compact. By Lemma 6.12, one sees that  $B \subset \mathcal{P}_{2,\theta}(\mathbb{R}^{d_1+d_2})$  is a pullback absorbing set such that  $P_\varepsilon^*(t, \tilde{\mathbb{F}}_\varepsilon, B) \subset B$  for all  $t \geq 0$  and  $\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)$ . It follows from Lemma 6.11 that  $(\sigma, P_\varepsilon^*)$  has a pullback attractor  $\mathcal{A}^\varepsilon$  with component subsets

$$\mathcal{A}_{\tilde{\mathbb{F}}_\varepsilon} := \bigcap_{\tau \geq 0} \overline{\bigcup_{t \geq \tau} P_\varepsilon^*(t, \sigma_{-t}(\tilde{\mathbb{F}}_\varepsilon), B)} \subset B, \quad \tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon).$$

(ii) Let

$$B_1 := \overline{\left\{ \mu \in \mathcal{P}_2(\mathbb{R}^{d_1}) : \int_{\mathbb{R}^{d_1}} |x|^2 \mu(dx) \leq K_8 \lambda_1^{-1} + 1 \right\}}.$$

Employing Lemma 6.14, we show that  $B_1$  is a positively invariant and absorbing set. Thanks to Lemma 6.9,  $\bar{P}^*$  admits a global attractor  $\bar{\mathcal{A}}$ , defined by

$$\bar{\mathcal{A}} := \bigcap_{t \geq 0} \bar{P}^*(t, B_1).$$

(iii) For any  $\delta > 0$  and bounded subset  $D \subset \mathcal{P}_2(\mathbb{R}^{d_1})$ , since  $\bar{\mathcal{A}}$  is the attractor of  $\bar{P}^*$ , there exists a  $T > 0$  such that for all  $t \geq T$

$$\bar{P}^*(t, D) \subset \mathcal{O}_{\delta/2}(\bar{\mathcal{A}}). \quad (6.4)$$

In view of (4.17), we have

$$\sup_{0 \leq t \leq 2T} d(\Pi_1 P_\varepsilon^*(t, \tilde{\mathbb{F}}_\varepsilon, m), \bar{P}^*(t, m \circ \pi_1^{-1})) < \eta(T, B)(\varepsilon) \quad (6.5)$$

for all  $m \in B$  and  $\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)$ . And there exists  $\varepsilon_0 > 0$  such that  $\eta(T, D)(\varepsilon) < \delta/2$  for all  $0 < \varepsilon \leq \varepsilon_0$ .

It follows from (6.4) and (6.5) that

$$\bigcup_{\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)} \Pi_1 P_\varepsilon^*(t, \tilde{\mathbb{F}}_\varepsilon, B) \subset \mathcal{O}_\delta(\bar{\mathcal{A}})$$

for all  $T \leq t \leq 2T$ . Taking some  $t_0 \in [T, 2T]$ , then we have

$$\bigcup_{\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)} P_\varepsilon^*(t_0, \sigma_{-t_0}(\tilde{\mathbb{F}}_\varepsilon), B) \subset \mathcal{O}_\delta(\bar{\mathcal{A}}).$$

It follows from the  $P_\varepsilon^*$ -invariance that  $P_\varepsilon^*(t_0, \sigma_{-t_0}(\tilde{\mathbb{F}}_\varepsilon), \mathcal{A}_{\sigma_{-t_0}(\tilde{\mathbb{F}}_\varepsilon)}) = \mathcal{A}_{\tilde{\mathbb{F}}_\varepsilon}$ , which implies that for all  $0 < \varepsilon \leq \varepsilon_0$  and  $\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)$

$$\mathcal{A}_{\tilde{\mathbb{F}}_\varepsilon} \subset \mathcal{O}_\delta(\bar{\mathcal{A}}),$$

because  $\mathcal{A}_{\sigma_{-t_0}(\mathbb{F}_\varepsilon)} \subset B$ .  $\square$

**6.3. Proof of Theorem 2.12.** In this subsection, we will give the proof of the second Bogolyubov theorem. To this end, we shall show the uniqueness and existence of bounded solutions to (1.3). We say that the solution  $V_t^\varepsilon := (X_t^\varepsilon, Y_t^\varepsilon)$ ,  $t \in \mathbb{R}$ , of (1.3) is  $\mathcal{L}^{2p}(\Omega, \mathbb{P}; \mathbb{R}^{d_1+d_2})$ -bounded if  $\sup_{t \in \mathbb{R}} \mathbb{E}|V_t^\varepsilon|^{2p} < \infty$ .

**Proposition 6.18.** *Assume that  $B(x, y) \equiv B(y)$  for all  $(x, y) \in \mathbb{R}^{d_1+d_2}$  and  $\beta \leq \alpha$ . Furthermore, suppose that  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$ ,  $(\mathbf{H}_y^6)$  and  $(\mathbf{H}_x^2)$ - $(\mathbf{H}_x^6)$  hold. If  $\beta = \alpha$  and  $\lambda_1 > \frac{L_b^2}{\eta}$ , then there exists  $\varepsilon_0 > 0$  such that for all  $0 < \varepsilon \leq \varepsilon_0$*

$$\mathbb{E}|V_{s,t}^\varepsilon(\xi_1) - V_{s,t}^\varepsilon(\xi_2)|^2 \leq \mathbb{E}|\xi_1 - \xi_2|^2 \exp\left(-\left(\frac{\lambda_1}{2} - \frac{L_b^2}{2\eta}\right)(t-s)\right). \quad (6.6)$$

If  $\beta < \alpha$  then there exists  $\varepsilon_0 > 0$  such that for all  $0 < \varepsilon \leq \varepsilon_0$

$$\mathbb{E}|V_{s,t}^\varepsilon(\xi_1) - V_{s,t}^\varepsilon(\xi_2)|^2 \leq \mathbb{E}|\xi_1 - \xi_2|^2 \exp\left(-\frac{\lambda_1}{2}(t-s)\right). \quad (6.7)$$

Moreover, for any  $0 < \varepsilon \leq \varepsilon_0$  there exists a unique solution  $V_t^\varepsilon$ ,  $t \in \mathbb{R}$ , to (1.3) such that

$$\sup_{t \in \mathbb{R}} \mathbb{E}|V_t^\varepsilon|^{2p} < \infty,$$

where  $1 \leq p < \frac{\lambda_1}{2L_b^2} + 1$ .

*Proof.* In view of Itô's formula, we have

$$\begin{aligned} & \mathbb{E}|V_{s,t}^\varepsilon(\xi_1) - V_{s,t}^\varepsilon(\xi_2)|^2 \\ &= \mathbb{E}|\xi_1 - \xi_2|^2 + \mathbb{E} \int_s^t \left( 2\langle F_\varepsilon(\tau, V_{s,\tau}^\varepsilon(\xi_1)) - F_\varepsilon(\tau, V_{s,\tau}^\varepsilon(\xi_2)), V_{s,\tau}^\varepsilon(\xi_1) - V_{s,\tau}^\varepsilon(\xi_2) \rangle \right. \\ & \quad \left. + |G_\varepsilon(\tau, V_{s,\tau}^\varepsilon(\xi_1)) - G_\varepsilon(\tau, V_{s,\tau}^\varepsilon(\xi_2))|_{HS}^2 \right) d\tau. \end{aligned} \quad (6.8)$$

By  $(\mathbf{H}_x^3)$ ,  $(\mathbf{H}_x^6)$  and  $(\mathbf{H}_y^6)$ , for all  $t \in \mathbb{R}$  and  $v_1 := (x_1, y_1)^T$ ,  $v_2 := (x_2, y_2)^T \in \mathbb{R}^{d_1+d_2}$  we have

$$\begin{aligned} & 2\langle F_\varepsilon(t, v_1) - F_\varepsilon(t, v_2), v_1 - v_2 \rangle + |G_\varepsilon(t, v_1) - G_\varepsilon(t, v_2)|_{HS}^2 \\ &= 2\langle f_\varepsilon(t, x_1, y_1) - f_\varepsilon(t, x_2, y_2), x_1 - x_2 \rangle + |\sigma_\varepsilon(t, x_1) - \sigma_\varepsilon(t, x_2)|_{HS}^2 \\ & \quad + \varepsilon^{-2\alpha} 2\langle B(y_1) - B(y_2), y_1 - y_2 \rangle + \varepsilon^{-2\alpha} |g(y_1) - g(y_2)|_{HS}^2 \\ & \quad + \varepsilon^{-\beta} 2\langle b(x_1, y_1) - b(x_2, y_2), y_1 - y_2 \rangle \\ & \leq -\lambda_1 |x_1 - x_2|^2 + \lambda_2 |y_1 - y_2|^2 - \varepsilon^{-2\alpha} \eta |y_1 - y_2|^2 \\ & \quad + \varepsilon^{-\beta} 2L_b(|x_1 - x_2| + |y_1 - y_2|)|y_1 - y_2|. \end{aligned} \quad (6.9)$$

If  $\beta = \alpha$  and  $\lambda_1 > L_b^2/\eta$ , then it follows from (6.9) and Young's inequality that there exists a constant  $\varepsilon_0 > 0$  such that for all  $0 < \varepsilon \leq \varepsilon_0$

$$\begin{aligned} & 2\langle F_\varepsilon(t, v_1) - F_\varepsilon(t, v_2), v_1 - v_2 \rangle + |G_\varepsilon(t, v_1) - G_\varepsilon(t, v_2)|_{HS}^2 \\ & \leq -\left(\frac{\lambda_1}{2} - \frac{L_b^2}{2\eta}\right) |x_1 - x_2|^2 \\ & \quad - \varepsilon^{-2\alpha} \left( \eta \left(1 - \frac{2L_b^2}{\lambda_1 \eta + L_b^2}\right) - \varepsilon^{2\alpha} \lambda_2 - \varepsilon^{2\alpha-\beta} 2L_b \right) |y_1 - y_2|^2 \\ & \leq -\left(\frac{\lambda_1}{2} - \frac{L_b^2}{2\eta}\right) |v_1 - v_2|^2, \end{aligned}$$

which by (6.8) and Gronwall's inequality implies that

$$\mathbb{E}|V_{s,t}^\varepsilon(\xi_1) - V_{s,t}^\varepsilon(\xi_2)|^2 \leq \mathbb{E}|\xi_1 - \xi_2|^2 \exp\left(-\left(\frac{\lambda_1}{2} - \frac{L_b^2}{2\eta}\right)(t-s)\right).$$

If  $\beta < \alpha$  then based on (6.9) and Young's inequality there exists  $\varepsilon_0 > 0$  such that for all  $0 < \varepsilon \leq \varepsilon_0$

$$\begin{aligned} & 2\langle F_\varepsilon(t, v_1) - F_\varepsilon(t, v_2), v_1 - v_2 \rangle + |G_\varepsilon(t, v_1) - G_\varepsilon(t, v_2)|_{HS}^2 \\ & \leq -\frac{\lambda_1}{2}|x_1 - x_2|^2 - \left( \varepsilon^{-2\alpha}\eta - \varepsilon^{-2\beta}\frac{2L_b^2}{\lambda_1} - \varepsilon^{-\beta}2L_b - \lambda_2 \right) |y_1 - y_2|^2 \leq -\frac{\lambda_1}{2}|v_1 - v_2|^2, \end{aligned}$$

which by (6.8) and Gronwall's inequality implies that

$$\mathbb{E}|V_{s,t}^{\varepsilon,1}(\xi_1) - V_{s,t}^{\varepsilon,2}(\xi_2)|^2 \leq \mathbb{E}|\xi_1 - \xi_2|^2 \exp\left(-\frac{\lambda_1}{2}(t-s)\right).$$

For any  $n \in \mathbb{N}$ , let  $V_{-n,t}^\varepsilon := (X_{-n,t}^\varepsilon, Y_{-n,t}^\varepsilon)$ ,  $t \geq -n$ , be the solution to (1.3). Thanks to (6.6), (6.7) and the classical pullback absorbing method (see e.g. [9, Theorem 3.6] for more details), there exists a  $\mathcal{L}^2(\Omega, \mathbb{P}; \mathbb{R}^{d_1+d_2})$ -bounded solution  $V_t^\varepsilon := (X_t^\varepsilon, Y_t^\varepsilon)$ ,  $t \in \mathbb{R}$ , of (1.3), which is the limit of  $V_t^\varepsilon(-n, 0)$  in  $\mathcal{L}^2(\Omega, \mathbb{P}; \mathbb{R}^{d_1+d_2})$  as  $n \rightarrow \infty$ . Moreover, by (6.2) and (6.1), one sees that  $\sup_{t \in \mathbb{R}} \mathbb{E}|V_t^\varepsilon|^{2p} \leq \infty$ , where  $1 \leq p < \frac{\lambda_1}{2L_b^2} + 1$ .  $\square$

**Remark 6.19.** Note that Proposition 6.18 implies that the bounded solution of (1.3) is globally asymptotically stable in the square-mean sense.

Now we can show that the bounded solution inherits the quasi-periodic property of the coefficients, and establish the second Bogolyubov theorem.

*Proof of Theorem 2.12.* Similar to the proof of Theorem 3.14 in [9], we show that if  $f$  and  $g$  are quasi-periodic then for any  $0 < \varepsilon \leq \varepsilon_0$ ,  $V_t^\varepsilon$ ,  $t \in \mathbb{R}$ , is quasi-periodic in distribution.

Note that it can be verified that for all  $x_1, x_2 \in \mathbb{R}^{d_1}$

$$2\langle \bar{f}(x_1) - \bar{f}(x_2), x_1 - x_2 \rangle + |\bar{\sigma}(x_1) - \bar{\sigma}(x_2)| \leq -\lambda_1|x_1 - x_2|^2,$$

which implies that the global attractor  $\bar{\mathcal{A}}$  of  $\bar{P}^*$  is a singleton set and

$$\bar{\mathcal{A}} = \{\mathcal{L}(\bar{X}_t)\}, \tag{6.10}$$

where  $\bar{X}_t$ ,  $t \in \mathbb{R}$  is the stationary solution to (1.4).

For any  $\tilde{\mathbb{F}}_\varepsilon \in \mathcal{H}(\mathbb{F}_\varepsilon)$  define

$$\mathfrak{B}_{\tilde{\mathbb{F}}_\varepsilon} := \left\{ \mathcal{L}(V_t^{\tilde{\mathbb{F}}_\varepsilon}) : t \in \mathbb{R} \right\},$$

where  $V_t^{\tilde{\mathbb{F}}_\varepsilon}$  is the bounded solution to (1.3) with  $\tilde{\mathbb{F}}_\varepsilon$  replacing  $\mathbb{F}_\varepsilon$ . Note that

$$P_\varepsilon^*(t, \sigma_{-t}\tilde{\mathbb{F}}_\varepsilon, \mathfrak{B}_{\sigma_{-t}\tilde{\mathbb{F}}_\varepsilon}) = \mathfrak{B}_{\tilde{\mathbb{F}}_\varepsilon}$$

and  $\mathfrak{B}_{\sigma_{-t}\tilde{\mathbb{F}}_\varepsilon} \subset B$ , where  $B$  is defined by (6.3). Then we have

$$\mathfrak{B}_{\tilde{\mathbb{F}}_\varepsilon} \subset \overline{\bigcap_{t \geq 0} \bigcup_{\tau \geq t} P_\varepsilon^*(t, \sigma_{-t}\tilde{\mathbb{F}}_\varepsilon, \mathfrak{B}_{\sigma_{-t}\tilde{\mathbb{F}}_\varepsilon})} \subset \overline{\bigcap_{t \geq 0} \bigcup_{\tau \geq t} P_\varepsilon^*(t, \sigma_{-t}\tilde{\mathbb{F}}_\varepsilon, B)} = \mathfrak{A}_{\tilde{\mathbb{F}}_\varepsilon},$$

which by (6.10) and Theorem 2.8 implies that

$$\limsup_{\varepsilon \rightarrow 0} \sup_{t \in \mathbb{R}} d_{BL}(\mathcal{L}(X_t^\varepsilon), \mathcal{L}(\bar{X}_t)) = 0.$$

$\square$

## APPENDIX A.

In this section, we will show the existence and uniqueness of solutions to (1.3).

**Lemma A.1.** *Suppose that  $(\mathbf{H}_y^1)$ ,  $(\mathbf{H}_y^3)$  and  $(\mathbf{H}_x^1)$ – $(\mathbf{H}_x^3)$  hold. Then for any  $(x, y) \in \mathbb{R}^{d_1+d_2}$  there exists a unique solution  $(X_t^\varepsilon(x), Y_t^\varepsilon(y))$  to (1.3) provided  $0 < \varepsilon \leq \sqrt[2\alpha]{\eta'/K_5} \wedge 1$ .*

*Proof.* Let  $v := (x, y)^T$ ,

$$F_\varepsilon(t, v) := (f_\varepsilon(t, x, y), \varepsilon^{-2\alpha}B(x, y) + \varepsilon^{-\beta}b(x, y))^T, \quad G_\varepsilon(t, v) := (\sigma_\varepsilon(t, x), \varepsilon^{-\alpha}g(x, y))^T$$

and  $W := (W^1, W^2)^T$ . Then equation (1.3) can be written as

$$dV_t^\varepsilon = F_\varepsilon(t, V_t^\varepsilon)dt + G_\varepsilon(t, V_t^\varepsilon)dW_t.$$

Note that by  $(\mathbf{H}_x^1)$  and  $(\mathbf{H}_y^1)$ , we have for any  $t \in \mathbb{R}$  and  $v := (x, y) \in \mathbb{R}^{d_1+d_2}$

$$\begin{aligned} & 2\langle F_\varepsilon(t, v), v \rangle + |G_\varepsilon(t, v)|_{HS}^2 \\ &= 2\langle f_\varepsilon(t, x, y), x \rangle + |\sigma_\varepsilon(t, x)|_{HS}^2 + \varepsilon^{-2\alpha} (2\langle B(x, y), y \rangle + |g(x, y)|_{HS}^2) + \varepsilon^{-\beta} 2\langle b(x, y), y \rangle \\ &\leq K_4|x|^2 + (-\varepsilon^{-2\alpha}\eta + \varepsilon^{-\beta}\tilde{\eta})|y|^2 + (K_5 - \varepsilon^{-2\alpha}\eta')|y|^\theta + K_4 + (\varepsilon^{-2\alpha} + \varepsilon^{-\beta})K_1. \end{aligned}$$

If  $\theta = 2$  or  $K_5 = 0$ , then for all  $0 < \varepsilon \leq 1$

$$2\langle F_\varepsilon(t, v), v \rangle + |G_\varepsilon(t, v)|_{HS}^2 \leq C_{K_4, \varepsilon}|v|^2 + C_{\varepsilon, K_1, K_4}, \quad \forall (t, v) \in \mathbb{R}^{1+d_1+d_2}.$$

Otherwise, for all  $0 < \varepsilon \leq \sqrt[2\alpha]{\eta'/K_5}$

$$2\langle F_\varepsilon(t, v), v \rangle + |G_\varepsilon(t, v)|_{HS}^2 \leq K_4|v|^2 + C_{\varepsilon, K_1, K_4}, \quad \forall (t, v) \in \mathbb{R}^{1+d_1+d_2}.$$

By  $(\mathbf{H}_x^2)$ ,  $(\mathbf{H}_x^3)$  and  $(\mathbf{H}_y^3)$ , one sees that

$$\begin{aligned} & 2\langle F_\varepsilon(t, v_1) - F_\varepsilon(t, v_2), v_1 - v_2 \rangle + |G_\varepsilon(t, v_1) - G_\varepsilon(t, v_2)|_{HS}^2 \\ &= 2\langle f_\varepsilon(t, x_1, y_1) - f_\varepsilon(t, x_2, y_2), x_1 - x_2 \rangle + |\sigma_\varepsilon(t, x_1) - \sigma_\varepsilon(t, x_2)|_{HS}^2 \\ &\quad + \varepsilon^{-2\alpha} 2\langle B(x_1, y_1) - B(x_2, y_2), y_1 - y_2 \rangle + \varepsilon^{-2\alpha} 2|g(x_1, y_1) - g(x_2, y_2)|_{HS}^2 \\ &\quad + \varepsilon^{-\beta} 2\langle b(x_1, y_1) - b(x_2, y_2), y_1 - y_2 \rangle \\ &\leq 2|f_\varepsilon(t, x_1, y_1) - f_\varepsilon(t, x_2, y_2)||x_1 - x_2| + L_\sigma^2|x_1 - x_2|^2 \\ &\quad + \varepsilon^{-2\alpha} 2\langle B(x_1, y_1) - B(x_1, y_2), y_1 - y_2 \rangle + \varepsilon^{-2\alpha} 2(L_g|x_1 - x_2| + L_g|y_1 - y_2|)^2 \\ &\quad + \varepsilon^{-2\alpha} 2|B(x_1, y_2) - B(x_2, y_2)||y_1 - y_2| + \varepsilon^{-\beta} 2|b(x_1, y_1) - b(x_2, y_2)||y_1 - y_2| \\ &\leq 2K_7(1 + |x_1|^{\theta_1} + |x_2|^{\theta_1} + |y_1|^{\theta_2} + |y_2|^{\theta_2})(|x_1 - x_2| + |y_1 - y_2|)|x_1 - x_2| + L_\sigma^2|x_1 - x_2|^2 \\ &\quad - \varepsilon^{-2\alpha}\eta|y_1 - y_2|^2 - \varepsilon^{-2\alpha}\eta'|y_1 - y_2|^\theta + \varepsilon^{-2\alpha}4L_g^2(|x_1 - x_2|^2 + |y_1 - y_2|^2) \\ &\quad + 2\varepsilon^{-2\alpha}K_3(1 + |y_2|^{\kappa_2})|x_1 - x_2||y_1 - y_2| \\ &\quad + 2\varepsilon^{-\beta}K_4(1 + |x_1|^{\kappa_1} + |x_2|^{\kappa_1} + |y_1|^{\kappa_2} + |y_2|^{\kappa_2})(|x_1 - x_2| + |y_1 - y_2|)|y_1 - y_2| \\ &\leq C_{K_4, K_7, L_\sigma, L_g, \varepsilon}(1 + |x_1|^{2\theta_1 \vee 2\kappa_1} + |x_2|^{2\theta_1 \vee 2\kappa_1} + |y_1|^{2\theta_2 \vee 2\kappa_2} + |y_2|^{2\theta_2 \vee 2\kappa_2})|x_1 - x_2|^2 \\ &\quad + C_{K_4, K_7, L_\sigma, L_g, \varepsilon, \tilde{\eta}}(1 + |x_1|^{2\theta_2 \vee 2\kappa_2} + |x_2|^{2\theta_2 \vee 2\kappa_2} + |y_1|^{2\theta_2 \vee 2\kappa_2} + |y_2|^{2\theta_2 \vee 2\kappa_2})|y_1 - y_2|^2 \end{aligned}$$

for any  $t \in \mathbb{R}$  and  $v_1 := (x_1, y_1), v_2 := (x_2, y_2) \in \mathbb{R}^{d_1+d_2}$ . Therefore, it follows from [35, Theorem 3.1.1] that for any  $(x, y) \in \mathbb{R}^{d_1+d_2}$  there exists a unique solution  $(X_t^\varepsilon(x), Y_t^\varepsilon(y))$  to (1.3) for all  $0 < \varepsilon \leq \sqrt[2\alpha]{\eta'/K_5} \wedge 1$ .  $\square$

#### ACKNOWLEDGEMENTS

The first author would like to acknowledge the warm hospitality of Bielefeld University. The second author was supported by NSFC Grants 11871132, 11925102, and Dalian High-level Talent Innovation Project (Grant 2020RD09). The third author was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 1283/2 2021 - 317210226.

## REFERENCES

- [1] ARNOLD, L. (2001). Hasselmann’s program revisited: the analysis of stochasticity in deterministic climate models. *Stochastic climate models* 141–157.
- [2] BAKHTIN, V. and KIFER, Y. (2004). Diffusion approximation for slow motion in fully coupled averaging. *Probab. Theory Related Fields* **129** 157–181.
- [3] BALL, K., KURTZ, T. G., POPOVIC, L. and REMPALA, G. (2006). Asymptotic analysis of multi-scale approximations to reaction networks. *Ann. Appl. Probab.* **16** 1925–1961.
- [4] BOGOLYUBOV, N. N. and MITROPOLSKY, Y. A. (1961). *Asymptotic Methods in the Theory of Non-linear Oscillations*. Hindustan Publishing Corp., Delhi, Gordon and Breach Science Publishers, New York.
- [5] CERRAI, S. (2009). A Khasminskii type averaging principle for stochastic reaction-diffusion equations. *Ann. Appl. Probab.* **19** 899–948.
- [6] CERRAI, S. (2009). Normal deviations from the averaged motion for some reaction-diffusion equations with fast oscillating perturbation. *J. Math. Pures Appl.* **91** 614–647.
- [7] CERRAI, S., and FREIDLIN, M. (2009). Averaging principle for a class of stochastic reaction-diffusion equations. *Probab. Theory Related Fields* **144** 137–177.
- [8] CHENG, M. and LIU, Z. (2021). Periodic, almost periodic and almost automorphic solutions for SPDEs with monotone coefficients. *Discrete Contin. Dyn. Syst. Ser. B* **26** 6425–6462.
- [9] CHENG, M. and LIU, Z. (2023). The second Bogolyubov theorem and global averaging principle for SPDEs with monotone coefficients. *SIAM J. Math. Anal.* **55** 1100–1144.
- [10] CHENG, M. and LIU, Z. (2023). Averaging principle for stochastic complex Ginzburg-Landau equations. *J. Differential Equations* **368** 58–104.
- [11] CRISAN, D., DOBSON, P., GODDARD, B. and SOUTTAR, I. (2022). Poisson equations with locally-Lipschitz coefficients and uniform in time averaging for stochastic differential equations via strong exponential stability. *arXiv:2204.02679v1*.
- [12] DONG, Z., SUN, X., XIAO, H. and ZHAI, J. (2018). Averaging principle for one dimensional stochastic Burgers equation. *J. Differential Equations* **265** 4749–4797.
- [13] DUAN, J. and WANG, W. (2014). *Effective Dynamics of Stochastic Partial Differential Equations*. Elsevier, Amsterdam.
- [14] E, W., LIU, D. and VANDEN-ELJNDEN, E. (2005). Analysis of multi-scale methods for stochastic differential equations. *Comm. Pure Appl. Math.* **58** 1544–1585.
- [15] FENG, J., FOUQUE, J.-P. and KUMAR, R. (2012). Small-time asymptotics for fast mean-reverting stochastic volatility models. *Ann. Appl. Probab.* **22** 1541–1575.
- [16] FREIDLIN, M. I. and WENTZELL, A. D. (2006). Long-time behavior of weakly coupled oscillators. *J. Stat. Phys.* **123** 1311–1337.
- [17] FREIDLIN, M. I. and WENTZELL, A. D. (2012). *Random Perturbations of Dynamical Systems*, 3rd ed. Springer, Heidelberg.
- [18] GAO, P. (2019). Averaging principle for stochastic Korteweg–de Vries equation. *J. Differential Equations* **267** 6872–6909.
- [19] HAIRER, M. and LI, X.-M. (2020). Averaging dynamics driven by fractional Brownian motion. *Ann. Probab.* **48** 1826–1860.
- [20] HALE, J. K. and VERDUYN LUNEL, S. M. (1990). Averaging in infinite dimensions. *J. Integral Equations Appl.* **2** 463–494.
- [21] HONG, W., LI, S. and LIU, W. (2021). Freidlin-Wentzell type large deviation principle for multi-scale locally monotone SPDEs. *SIAM J. Math. Anal.* **53** 6517–6561.
- [22] W. HU and C. LI, A convergence analysis of the perturbed compositional gradient flow: averaging principle and normal deviations. *Discrete Contin. Dyn. Syst.* **38** (2018), 4951–4977.  
HU, W. and LI, C. (2018). A convergence analysis of the perturbed compositional gradient flow: averaging principle and normal deviations. *Discrete Contin. Dyn. Syst.* **38** 4951–4977.
- [23] ILYIN, A. A. (1996). Averaging of dissipative dynamical systems with rapidly oscillating right-hand sides. *Mat. Sb.* **187** 15–58.
- [24] ILYIN, A. A. (1998). Global averaging of dissipative dynamical systems. *Rend. Accad. Naz. Sci. XL Mem. Mat. Appl. (5)* **22** 165–191.
- [25] KANG, H.-W. and KURTZ, T. G. (2013). Separation of time-scales and model reduction for stochastic reaction networks. *Ann. Appl. Probab.* **23** 529–583.
- [26] KHASMINSKII, R. Z. (1996). On stochastic processes defined by differential equations with a small parameter. *Theory Probab. Appl.* **11** 211–228.
- [27] KHASMINSKII, R. Z. (1968). On the principle of averaging the Itô’s stochastic differential equations. *Kybernetika (Prague)* **4** 260–279.
- [28] KHASMINSKII, R. Z. and YIN, G. (2004). On averaging principles: an asymptotic expansion approach. *SIAM J. Math. Anal.* **35** 1534–1560.
- [29] KIFER, Y. (2001). Averaging and climate models. *Stochastic climate models* **49** 171–188.

- [30] KIFER, Y. and VARADHAN, S. R. S. (2014). Nonconventional limit theorems in discrete and continuous time via martingales. *Ann. Probab.* **42** 649–688.
- [31] KONDA, V. R. and TSITSIKLIS, J. N. (2004). Convergence rate of linear two-time-scale stochastic approximation. *Ann. Appl. Probab.* **14** 796–819.
- [32] KONSTANTINOS, S. (2014). Fluctuation analysis and short time asymptotics for multiple scales diffusion processes. *Stoch. Dyn.* **14** 1350026, 22 pp.
- [33] KLOEDEN, P. E. and RASMUSSEN, M. (2011). *Nonautonomous Dynamical Systems*. American Mathematical Society, Providence.
- [34] KRYLOV, N. and BOGOLYUBOV, N. N. (1943). *Introduction to Non-Linear Mechanics*. Princeton University Press, Princeton.
- [35] LIU, W. and RÖCKNER, M. (2015). *Stochastic Partial Differential Equations: an Introduction*. Springer, Cham.
- [36] LIU, W., RÖCKNER, M., SUN, X. and XIE, Y. (2020). Averaging principle for slow-fast stochastic differential equations with time dependent locally Lipschitz coefficients. *J. Differential Equations* **268** 2910–2948.
- [37] MAJDA, A. J., TIMOFEYEV, I. and VANDEN EIJNDEN, E. (2001). A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.* **54** 891–974.
- [38] MASLOWSKI, B., SEIDLER, J. and VRKOČ, I. (1991). An averaging principle for stochastic evolution equations. II. *Math. Bohem.* **116** 191–224.
- [39] PAKDAMAN, K., THIEULLEN, M. and WAINRIB, G. (2012). Asymptotic expansion and central limit theorem for multi-scale piecewise-deterministic Markov processes. *Stochastic Process. Appl.* **122** 2292–2318.
- [40] RÖCKNER, M. and XIE, L. (2021). Averaging principle and normal deviations for multi-scale stochastic systems. *Comm. Math. Phys.* **383** 1889–1937.
- [41] PARDOUX, È. and VERETENNIKOV, A. YU. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29** 1061–1085.
- [42] PARDOUX, È. and VERETENNIKOV, A. YU. (2003). On Poisson equation and diffusion approximation. II. *Ann. Probab.* **31** 1166–1192.
- [43] SHI, Y., SUN, X., WANG, L. and XIE, Y. (2022). Asymptotic behavior for multi-scale SDEs with monotonicity coefficients driven by Lévy processes. *arXiv:2208.07560v1*
- [44] VERETENNIKOV, A. YU (1990). On an averaging principle for systems of stochastic differential equations. *Mat. Sb.* **181** 256–268.
- [45] WANG, W. and ROBERTS, A. J. (2012). Average and deviation for slow-fast stochastic partial differential equations. *J. Differential Equations* **253** 1265–1286.
- [46] XU, Y., DUAN, J. and XU, W. (2011). An averaging principle for stochastic dynamical systems with Lévy noise. *Phys. D* **240** 1395–1401.
- [47] ZELIK, S. (2006). Global averaging and parametric resonances in damped semilinear wave equations. *Proc. Roy. Soc. Edinburgh Sect. A* **136** 1053–1097.

M. CHENG: SCHOOL OF MATHEMATICS AND STATISTICS, BEIJING INSTITUTE OF TECHNOLOGY, BEIJING 100081, P. R. CHINA

*E-mail address:* mengyu.cheng@hotmail.com; mcheng@bit.edu.cn

Z. LIU: SCHOOL OF MATHEMATICAL SCIENCES, DALIAN UNIVERSITY OF TECHNOLOGY, DALIAN 116024, P. R. CHINA

*E-mail address:* zxliu@dlut.edu.cn

M. RÖCKNER: FAKULTÄT FÜR MATHEMATIK, UNIVERSITÄT BIELEFELD, D-33501 BIELEFELD, GERMANY

*E-mail address:* roeckner@math.uni-bielefeld.de