AVERAGING PRINCIPLE AND NORMAL DEVIATIONS FOR MULTI-SCALE STOCHASTIC HYPERBOLIC-PARABOLIC EQUATIONS

MICHAEL RÖCKNER, LONGJIE XIE AND LI YANG

Abstract. We study the asymptotic behavior of stochastic hyperbolic-parabolic equations with slow and fast time scales. Both the strong and weak convergence in the averaging principe are established, which can be viewed as a functional law of large numbers. Then we study the stochastic fluctuations of the original system around its averaged equation. We show that the normalized difference converges weakly to the solution of a linear stochastic wave equation, which is a form of functional central limit theorem. We provide a unified proof for the above convergence by using the Poisson equation in Hilbert spaces. Moreover, sharp rates of convergence are obtained, which are shown not to depend on the regularity of the coefficients in the equation for the fast variable.

AMS 2010 Mathematics Subject Classification: 60H15, 60F05, 70K70.

Keywords and Phrases: Stochastic hyperbolic-parabolic equations; averaging principle; strong and weak convergence; homogenization.

1. Introduction

Let $T > 0$ and $D \subseteq \mathbb{R}^d$ ($d \geqslant 1$) be a bounded open set. Consider the following system of stochastic hyperbolic-parabolic equations:

$$\begin{cases} \dfrac{\partial^2 U_t^\varepsilon(\xi)}{\partial t^2} = \Delta U_t^\varepsilon(\xi) + f(U_t^\varepsilon(\xi), Y_t^\varepsilon(\xi)) + \dot{W}_t^1(\xi), & (t,\xi) \in (0,T] \times D, \\ \dfrac{\partial Y_t^\varepsilon(\xi)}{\partial t} = \dfrac{1}{\varepsilon}\Delta Y_t^\varepsilon(\xi) + \dfrac{1}{\varepsilon} g(U_t^\varepsilon(\xi), Y_t^\varepsilon(\xi)) + \dfrac{1}{\sqrt{\varepsilon}}\dot{W}_t^2(\xi), & (t,\xi) \in (0,T] \times D, \\ U_t^\varepsilon(\xi) = Y_t^\varepsilon(\xi) = 0, & (t,\xi) \in (0,T] \times \partial D, \\ U_0^\varepsilon(\xi) = u(\xi),\ \dfrac{\partial U_t^\varepsilon(\xi)}{\partial t}\Big|_{t=0} = v(\xi),\ Y_0^\varepsilon(\xi) = y(\xi), & \xi \in D, \end{cases} \tag{1.1}$$

where Δ is the Laplacian operator, ∂D denotes the boundary of the domain D, $f, g : \mathbb{R}^2 \to \mathbb{R}$ are measurable functions, W_t^1 and W_t^2 are two mutually independent Q_1- and Q_2-Wiener processes both defined on a complete probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geqslant 0}, \mathbb{P})$, and the small parameter $0 < \varepsilon \ll 1$ represents the separation of time scales between the

This work is supported by the DFG through CRC 1283 and NSFC (No. 12071186, 11931004).

'slow' process U_t^ε and the 'fast' motion Y_t^ε (with time order $1/\varepsilon$). Randomly perturbed hyperbolic partial differential equations are usually used to model wave propagation and mechanical vibration in a random medium. If these phenomena are temperature dependent or heat generating, then the underlying hyperbolic equation will be coupled with a stochastic parabolic equation, which leads to the mathematical description of slow-fast systems through (1.1), see e.g. [16, 30, 34, 39] and the references therein. In this respect, the question arises how a thermal environment at large time scales may influence the dynamics of the whole system.

In the mathematical literature, powerful averaging and homogenization methods have been developed to study the asymptotic behavior of multi-scale systems. The averaging principle can be viewed as a functional law of large numbers, which says that under certain regularity assumptions on the coefficients, the slow component will converge to the solution of the so-called averaged equation as $\varepsilon \to 0$. The averaged equation then captures the evolution of the original system over a long time scale, which does not depend on the fast variable any more. This theory was first studied by Bogoliubov [4] for deterministic ordinary differential equations, and extended to stochastic differential equations (SDEs for short) by Khasminskii [28], see also [1, 23, 24, 29, 37] and the references therein. Recently, the averaging principle for two time scale stochastic partial differential equations (SPDEs for short) has attracted considerable attention. In [12], Cerrai and Freidlin proved the averaging principle for slow-fast stochastic reaction-diffusion equations with noise only in the fast motion. Later, Cerrai [9, 11] generalized this result to more general reaction-diffusion equations, see also [2, 14, 36, 38] and the references therein for further developments. We also mention that Bréhier [5, 6] studied the rate of convergence in terms of $\varepsilon \to 0$ in the averaging principle for parabolic SPDEs and obtained the 1/2-order rate of strong convergence (in the mean-square sense) and the 1-order rate of weak convergence (in the distribution sense), which are known to be optimal. These rates of convergence are important for the study of other limit theorems in probability theory and numerical schemes, known as the Heterogeneous Multi-scale Method for the original multi-scale system, see e.g. [7, 20]. Concerning stochastic hyperbolic-parabolic equations, Fu ect. [22] established the strong convergence in the averaging principle for system (1.1) when $d = 1$ by the classical Khasminskii time discretization method, and obtained the 1/4-order rate of strong convergence. In [21], by using asymptotic expansion arguments, the authors studied the weak order convergence for system (1.1), but only in a not fully coupled case ($g(u, y) = g(y)$), i.e., the fast equation does not depend on the slow process.

In this paper, we shall first prove the strong and weak convergence in the averaging principle for the fully coupled system (1.1) with singular coefficients, see **Theorem 2.1**. Compared with [21, 22], we assume that the coefficients are only η-Hölder continuous with respect to the fast variable with any $\eta > 0$, and we obtain the optimal 1/2-order rate of strong convergence as well as the 1-order rate of weak convergence. Moreover, we find that both the strong and weak convergence rates do not depend on the regularity of the coefficients in the equation for the fast variable. This implies that the evolution of the multi-scale system (1.1) relies mainly on the slow variable, which coincides with the intuition since in the limit equation the fast component has been totally averaged out. Furthermore, the arguments we use are different from those in [5, 9, 11, 12, 21, 22]. Our method to establish the strong and weak convergence is based on the Poisson equation in Hilbert space, which is more unifying and much simpler.

The averaged equation for (1.1) is only valid in the limit when the time scale separation between the fast and slow variables is infinitely wide. Of course, the scale separation is never infinite in reality. For small but positive ε, the slow variable U_t^ε will experience fluctuations around its averaged motion $\bar{U}_t$. These small fluctuations can be captured by studying the functional central limit theorem. Namely, we are interested in the asymptotic behavior of the normalized difference

$$Z_t^\varepsilon := \frac{U_t^\varepsilon - \bar{U}_t}{\sqrt{\varepsilon}} \tag{1.2}$$

as ε tends to 0. Such result is known to be closely related to the homogenization behavior of singularly perturbed partial differential equations, which is of its own interest, see e.g. [25, 26]. For the study of the functional central limit theorem for finite dimensional multi-scale systems, we refer the reader to the fundamental paper by Khasminskii [28], see also [1, 15, 27, 32, 33, 35]. The infinite dimensional situation is more open and papers on this subject are very few. In [10], Cerrai studied the normal deviations for a deterministic reaction-diffusion equation with one dimensional space variable perturbed by a fast process, and proved the weak convergence to a Gaussian process, whose covariance is explicitly described. Later, this was generalized to general stochastic reaction-diffusion equations by Wang and Roberts [38]. In both papers, the methods of proof are based on Khasminskii's time discretization argument. Recently, we [36] studied the normal deviations for general slow-fast parabolic SPDEs by using the technique of Poisson equation.

In this paper, we further develop the argument used in [36] to study the functional central limit theorem for the stochastic hyperbolic-parabolic system (1.1) with Hölder continuous coefficients. We show that the normalized difference Z_t^ε, defined by (1.2), converges weakly as $\varepsilon \to 0$ to the solution of a linear stochastic wave equation, see **Theorem 2.3**. Moreover, the optimal 1/2-order rate of convergence is obtained. This rate also does not depend on the regularity of the coefficients in the equation for the fast variable, which again is natural since in the limit equation the fast component has been homogenized out. As far as we know, the result we obtained is completely new. It turns out that the argument we use to prove the functional central limit theorem is closely and universally connected with the proof of the strong and weak convergence in the averaging principle. We note that due to the model considered in this paper, the framework we deal with is different from [36]. Furthermore, we derive the higher order spatial-temporal convergence in the averaging principle and in the functional central limit theorem. Throughout our proof, several strong and weak fluctuation estimates will play an important role, see Lemmas 4.1, 4.2 and 5.2 below.

The rest of this paper is organized as follows. In Section 2, we first introduce some assumptions and state our main results. Section 3 is devoted to establish some preliminary estimates. Then we prove the strong and weak convergence results, Theorem 2.1, and the normal deviation result, Theorem 2.3, in Section 4 and Section 5, respectively.

Notations. To end this section, we introduce some usual notations for convenience. Given Hilbert spaces H_1, H_2 and $\hat{H}$, we use $\mathscr{L}(H_1, H_2)$ to denote the space of all linear and bounded operators from H_1 to H_2. If $H_1 = H_2$, we write $\mathscr{L}(H_1) = \mathscr{L}(H_1, H_1)$ for simplicity. Recall that an operator $Q \in \mathscr{L}(\hat{H})$ is called Hilbert-Schmidt if

$$\|Q\|^2_{\mathscr{L}_2(\hat{H})} := Tr(QQ^*) < +\infty.$$

We shall denote the space of all Hilbert-Schmidt operators on $\hat{H}$ by $\mathscr{L}_2(\hat{H})$. Let $L^\infty_\ell(H_1 \times H_2, \hat{H})$ denote the space of all measurable maps $\phi : H_1 \times H_2 \to \hat{H}$ with linear growth, i.e.,

$$\|\phi\|_{L^\infty_\ell(\hat{H})} := \sup_{(x,y)\in H_1\times H_2} \frac{\|\phi(x, y)\|_{\hat{H}}}{1 + \|x\|_{H_1} + \|y\|_{H_2}} < \infty.$$

For $k \in \mathbb{N}$, the space $C^{k,0}_\ell(H_1 \times H_2, \hat{H})$ contains all $\phi \in L^\infty_\ell(H_1 \times H_2, \hat{H})$ such that ϕ has k times Gâteaux derivatives with respect to the x-variable satisfying

$$\|\phi\|_{C^{k,0}_\ell(\hat{H})} := \sup_{(x,y)\in H_1\times H_2} \frac{\sum_{\imath=1}^{k} \|D^i_x\phi(x, y)\|_{\mathscr{L}^i(H_1,\hat{H})}}{1 + \|x\|_{H_1} + \|y\|_{H_2}} < \infty.$$

Similarly, the space $C^{0,k}_\ell(H_1 \times H_2, \hat{H})$ contains all $\phi \in L^\infty_\ell(H_1 \times H_2, \hat{H})$ such that ϕ has k times Gâteaux derivatives with respect to the y-variable satisfying

$$\|\phi\|_{C^{0,k}_\ell(\hat{H})} := \sup_{(x,y)\in H_1\times H_2} \frac{\sum_{i=1}^{k} \|D^i_y\phi(x, y)\|_{\mathscr{L}^i(H_2,\hat{H})}}{1 + \|x\|_{H_1} + \|y\|_{H_2}} < \infty.$$

For $k, m \in \mathbb{N}$, let $C^{k,m}_\ell(H_1 \times H_2, \hat{H})$ be the space of all maps satisfying

$$\|\phi\|_{C^{k,m}_\ell(\hat{H})} := \|\phi\|_{L^\infty_\ell(\hat{H})} + \|\phi\|_{C^{k,0}_\ell(\hat{H})} + \|\phi\|_{C^{0,m}_\ell(\hat{H})} < \infty, \tag{1.3}$$

and for $\eta \in (0, 1)$, the space $C^{k,\eta}_\ell(H_1 \times H_2, \hat{H})$ consists of all $\phi \in C^{k,0}_\ell(H_1 \times H_2, \hat{H})$ satisfying

$$\|\phi(x, y_1) - \phi(x, y_2)\|_{\hat{H}} \leqslant C_0\|y_1 - y_2\|^\eta_{H_2}\big(1 + \|x\|_{H_1} + \|y_1\|_{H_2} + \|y_2\|_{H_2}\big).$$

The space $C^{k,\eta}_b(H_1 \times H_2, \hat{H})$ consists of all $\phi \in C^{k,\eta}_\ell(H_1 \times H_2, \hat{H})$ whose k times Gâteaux derivatives with respect to the first variable are bounded, and the space $C^{k,\eta}_B(H_1 \times H_2, \hat{H})$

consists of all maps in $C_b^{k,\eta}(H_1\times H_2,\hat{H})$ which are bounded themselves. We also introduce the space $\mathbb{C}_l^{k,k}(H_1\times H_2,\hat{H})$ consisting of all maps which have k times Fréchet derivatives with respect to both the first variable and the second variable and satisfy (1.3). The space $\mathbb{C}_b^{k,k}(H_1\times H_2,\hat{H})$ consists of all $\phi\in\mathbb{C}_l^{k,k}(H_1\times H_2,\hat{H})$ with all derivatives bounded. When $\hat{H}=\mathbb{R}$, we will omit the letter $\hat{H}$ for simplicity.

2. Assumptions and main results

Let $H:=L^2(D)$ be the usual space of square integrable functions on a bounded open domain D in $\mathbb{R}^d$ with scalar product and norm denoted by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively. Let A be the realization of the Laplacian with Dirichlet boundary conditions in H. It is known that there exists a complete orthonormal basis $\{e_n\}_{n\in\mathbb{N}}$ of H such that

$$Ae_n=-\lambda_n e_n,$$

with $0<\lambda_1\leqslant\lambda_2\leqslant\cdots\lambda_n\leqslant\cdots$. For $\alpha\in\mathbb{R}$, let $H^\alpha:=\mathcal{D}((-A)^{\frac{\alpha}{2}})$ be the Hilbert space endowed with the scalar product

$$\langle x,y\rangle_\alpha:=\langle(-A)^{\frac{\alpha}{2}}x,(-A)^{\frac{\alpha}{2}}y\rangle=\sum_{n=1}^{\infty}\lambda_n^\alpha\langle x,e_n\rangle\langle y,e_n\rangle,\quad\forall x,y\in H^\alpha,$$

and norm

$$\|x\|_\alpha:=\left(\sum_{n=1}^{\infty}\lambda_n^\alpha\langle x,e_n\rangle^2\right)^{\frac{1}{2}},\quad\forall x\in H^\alpha.$$

Then A can be regarded as an operator from H^α to $H^{\alpha-2}$. For the drift coefficients f and g given in system (1.1), we introduce two Nemytskii operators $F,G:H\times H\to H$ by

$$F(u,y)(\xi):=f(u(\xi),y(\xi)),\quad G(u,y)(\xi):=g(u(\xi),y(\xi)),\quad\xi\in D.\tag{2.1}$$

We remark that these operators are not Fréchet differentiable in H.

To give precise results, it is convenient to write system (1.1) in the following abstract formulation in H:

$$\begin{cases}\dfrac{\partial^2U_t^\varepsilon}{\partial t^2}=AU_t^\varepsilon+F(U_t^\varepsilon,Y_t^\varepsilon)+\dot{W}_t^1, & t\in(0,T],\\[2mm] \dfrac{\partial Y_t^\varepsilon}{\partial t}=\dfrac{1}{\varepsilon}AY_t^\varepsilon+\dfrac{1}{\varepsilon}G(U_t^\varepsilon,Y_t^\varepsilon)+\dfrac{1}{\sqrt{\varepsilon}}\dot{W}_t^2, & t\in(0,T],\\[2mm] U_0^\varepsilon=u,\ \dfrac{\partial U_t^\varepsilon}{\partial t}\Big|_{t=0}=v,\ Y_0^\varepsilon=y.\end{cases}\tag{2.2}$$

For $i=1,2$, we assume that Q_i are nonnegative, symmetric operators with respect to $\{e_n\}_{n\in\mathbb{N}}$, i.e.,

$$Q_ie_n=\beta_{i,n}e_n,\ \ \beta_{i,n}>0,n\in\mathbb{N}.$$

In addition, we assume that

$$Tr(Q_i) = \sum_{n\in\mathbb{N}} \beta_{i,n} < +\infty, \quad i = 1, 2. \tag{2.3}$$

Given $u \in H$, consider the following frozen equation:

$$\mathrm{d}Y_t^u = AY_t^u \mathrm{d}t + G(u, Y_t^u)\mathrm{d}t + \mathrm{d}W_t^2, \quad Y_0^u = y \in H. \tag{2.4}$$

Under our assumptions below, the process Y_t^u admits a unique invariant measure $\mu^u(\mathrm{d}y)$. Then, the averaged equation for system (2.2) is

$$\begin{cases} \dfrac{\partial^2 \bar{U}_t}{\partial t^2} = A\bar{U}_t + \bar{F}(\bar{U}_t) + \dot{W}_t^1, \quad t \in (0, T], \\ \bar{U}_0 = u, \ \dfrac{\partial \bar{U}_t}{\partial t}|_{t=0} = v, \end{cases} \tag{2.5}$$

where

$$\bar{F}(u) := \int_H F(u, y)\mu^u(\mathrm{d}y). \tag{2.6}$$

Let $\dot{U}_t^\varepsilon := \partial U_t^\varepsilon / \partial t$ and $\dot{\bar{U}}_t := \partial \bar{U}_t / \partial t$. The following is the first main result of this paper.

Theorem 2.1. *Let $T > 0$, $u \in H^1$ and $v, y \in H$. Assume that $f \in C_b^{2,\eta}(\mathbb{R} \times \mathbb{R}, \mathbb{R})$ and $g \in C_B^{2,\eta}(\mathbb{R} \times \mathbb{R}, \mathbb{R})$ with $\eta > 0$. Then we have:*
(i) (strong convergence) for any $q \geqslant 1$,

$$\sup_{t\in[0,T]} \mathbb{E} \left(\|U_t^\varepsilon - \bar{U}_t\|_1^2 + \|\dot{U}_t^\varepsilon - \dot{\bar{U}}_t\|^2 \right)^{q/2} \leqslant C_1\, \varepsilon^{q/2}; \tag{2.7}$$

(ii) (weak convergence) for any $\phi \in \mathbb{C}_b^3(H)$ and $\tilde{\phi} \in \mathbb{C}_b^3(H^{-1})$,

$$\sup_{t\in[0,T]} \left(\left|\mathbb{E}[\phi(U_t^\varepsilon)] - \mathbb{E}[\phi(\bar{U}_t)]\right| + \left|\mathbb{E}[\tilde{\phi}(\dot{U}_t^\varepsilon)] - \mathbb{E}[\tilde{\phi}(\dot{\bar{U}}_t)]\right| \right) \leqslant C_2\, \varepsilon, \tag{2.8}$$

where $C_1 = C(T, u, v, y)$ and $C_2 = C(T, u, v, y, \phi, \tilde{\phi})$ are positive constants independent of ε and η.

Remark 2.2. *(i) The $1/2$-order rate of strong convergence in (2.7) and the 1-order rate of weak convergence in (2.8) should be optimal, which coincides with the SDE case as well as the stochastic reaction-diffusion equation case. Moreover, we obtain that both the strong and weak convergence rates do not depend on the regularity of the coefficients in the equation for the fast variable. This coincides with the intuition, since in the limit equation the fast component has been averaged out.*

(ii) Note that the coefficients are assumed to be only η-Hölder continuous with respect to the fast variable, which is sufficient for us to prove the above convergence in the

averaging principle. However, the pathwise uniqueness of solutions for system (2.2) is not clear under such weak assumptions. In particular, if the system is not fully coupled in the sense that the fast motion does not depend on the slow variable (i.e., $g(u, y) = g(y)$ in (1.1)), then the well-posedness for the fast equation with only Hölder continuous coefficients has been proven in [19, Theorem 7] *by using the Zvonkin's transformation. This in turn implies the strong well-posedness of the whole system (1.1).*

Recall that Z_t^ε is defined by (1.2). In view of (2.2) and (2.5), we have

$$\begin{aligned}\frac{\partial^2 Z_t^\varepsilon}{\partial t^2} &= AZ_t^\varepsilon + \frac{1}{\sqrt{\varepsilon}}\Big[F(U_t^\varepsilon, Y_t^\varepsilon) - \bar{F}(\bar{U}_t)\Big]\\ &= AZ_t^\varepsilon + \frac{1}{\sqrt{\varepsilon}}\Big[\bar{F}(U_t^\varepsilon) - \bar{F}(\bar{U}_t)\Big] + \frac{1}{\sqrt{\varepsilon}}\delta F(U_t^\varepsilon, Y_t^\varepsilon),\end{aligned}$$

where

$$\delta F(u, y) := F(u, y) - \bar{F}(u).$$

To study the homogenization behavior of Z_t^ε, we consider the following Poisson equation:

$$\mathcal{L}_2(u, y)\Psi(u, y) = -\delta F(u, y), \tag{2.9}$$

where $\mathcal{L}_2$ is the generator of the frozen equation (2.4) given by

$$\mathcal{L}_2(u, y)\varphi(y) := \langle Ay + G(u, y), D_y\varphi(y)\rangle + \frac{1}{2}Tr\left(D_y^2\varphi(y)Q_2^{\frac{1}{2}}(Q_2^{\frac{1}{2}})^*\right), \ \forall \varphi \in C_\ell^2(H), \tag{2.10}$$

and $u \in H$ is regarded as a parameter. According to Theorem 3.1 below, there exists a unique solution Ψ to equation (2.9). Then, the limit process $\bar{Z}_t$ of Z_t^ε turns out to satisfy the following linear stochastic wave equation:

$$\begin{cases} \dfrac{\partial^2 \bar{Z}_t}{\partial t^2} = A\bar{Z}_t + D_u\bar{F}(\bar{U}_t).\bar{Z}_t + \sigma(\bar{U}_t)\dot{W}_t, \ \ t \in (0, T], \\ \bar{Z}_0 = 0, \ \dfrac{\partial \bar{Z}_t}{\partial t}|_{t=0} = 0, \end{cases} \tag{2.11}$$

where W_t is another cylindrical Wiener process independent of W_t^1, and σ is a Hilbert-Schmidt operator satisfying

$$\frac{1}{2}\sigma(u)\sigma^*(u) = \overline{\delta F \otimes \Psi}(u) := \int_H \big[\delta F(u, y) \otimes \Psi(u, y)\big]\mu^u(\mathrm{d}y).$$

Let $\dot{Z}_t^\varepsilon := \partial Z_t^\varepsilon/\partial t$ and $\dot{\bar{Z}}_t := \partial \bar{Z}_t/\partial t$. We have the following result.

Theorem 2.3 (Normal deviation). *Let $T > 0$, $f \in C_b^{2,\eta}(\mathbb{R}\times\mathbb{R}, \mathbb{R})$ and $g \in C_B^{2,\eta}(\mathbb{R}\times\mathbb{R}, \mathbb{R})$ with $\eta > 0$. Then for any $u \in H^1$, $v, y \in H$, $\phi \in \mathbb{C}_b^3(H)$ and $\tilde{\phi} \in \mathbb{C}_b^3(H^{-1})$, we have*

$$\sup_{t\in[0,T]}\Big(\big|\mathbb{E}[\phi(Z_t^\varepsilon)] - \mathbb{E}[\phi(\bar{Z}_t)]\big| + \big|\mathbb{E}[\tilde{\phi}(\dot{Z}_t^\varepsilon)] - \mathbb{E}[\tilde{\phi}(\dot{\bar{Z}}_t)]\big|\Big) \leqslant C_3\,\varepsilon^{\frac{1}{2}}, \tag{2.12}$$

where $C_3 = C(T, u, \dot{u}, y, \phi, \tilde{\phi}) > 0$ is a constant independent of ε and η.

Remark 2.4. *The 1/2-order rate of convergence in (2.12) coincides with the SDE case and should be optimal. Moreover, the convergence rate does not depend on the regularity of the coefficients in the equation for the fast variable.*

3. Preliminaries

3.1. **Poisson equation.** We will rewrite the system (2.2) as an abstract evolution equation. To this end, we first introduce some notations. For $\alpha \in \mathbb{R}$, by $\mathcal{H}^\alpha := H^\alpha \times H^{\alpha-1}$ we denote the Hilbert space endowed with the scalar product

$$\langle u, v\rangle_{\mathcal{H}^\alpha} := \langle u_1, v_1\rangle_\alpha + \langle u_2, v_2\rangle_{\alpha-1}, \quad \forall u = (u_1, u_2)^T, v = (v_1, v_2)^T \in \mathcal{H}^\alpha,$$

and norm

$$|||u|||_\alpha^2 := \|u_1\|_\alpha^2 + \|u_2\|_{\alpha-1}^2, \quad \forall u = (u_1, u_2)^T \in \mathcal{H}^\alpha.$$

For simplicity, we write $\mathcal{H} := H \times H^{-1}$. Let Π_1 be the canonical projection from $\mathcal{H}$ to H, and define

$$V_t^\varepsilon := \frac{\mathrm{d}}{\mathrm{d}t} U_t^\varepsilon \quad \text{and} \quad X_t^\varepsilon := (U_t^\varepsilon, V_t^\varepsilon)^T.$$

Then, the system (2.2) can be rewritten as

$$\begin{cases} \mathrm{d}X_t^\varepsilon = \mathcal{A}X_t^\varepsilon + \mathcal{F}(X_t^\varepsilon, Y_t^\varepsilon) + B\mathrm{d}W_t^1, \\ \mathrm{d}Y_t^\varepsilon = \varepsilon^{-1}AY_t^\varepsilon + \varepsilon^{-1}\mathcal{G}(X_t^\varepsilon, Y_t^\varepsilon) + \varepsilon^{-1/2}\mathrm{d}W_t^2, \\ X_0^\varepsilon = x,\ Y_0^\varepsilon = y, \end{cases} \tag{3.1}$$

where $x := (u, v)^T$, and

$$\mathcal{A} := \begin{pmatrix} 0 & I \\ A & 0 \end{pmatrix},\ \mathcal{F}(x, y) := \begin{pmatrix} 0 \\ F(\Pi_1(x), y) \end{pmatrix},\ \mathcal{G}(x, y) := G(\Pi_1(x), y),\ B\mathrm{d}W_t^1 := \begin{pmatrix} 0 \\ \mathrm{d}W_t^1 \end{pmatrix},$$

and F, G are defined by (2.1). Similarly, concerning the averaged equation (2.5), let

$$\bar{V}_t := \frac{\mathrm{d}}{\mathrm{d}t}\bar{U}_t \quad \text{and} \quad \bar{X}_t := (\bar{U}_t, \bar{V}_t)^T.$$

Then we can transfer (2.5) into a stochastic evolution equation:

$$\mathrm{d}\bar{X}_t = \mathcal{A}\bar{X}_t\mathrm{d}t + \bar{\mathcal{F}}(\bar{X}_t)\mathrm{d}t + B\mathrm{d}W_t^1, \quad \bar{X}_0 = x = (u, v)^T \in \mathcal{H}, \tag{3.2}$$

where

$$\bar{\mathcal{F}}(x) := \begin{pmatrix} 0 \\ \bar{F}(\Pi_1(x)) \end{pmatrix},$$

and $\bar{F}$ is defined by (2.6). It is known (see e.g. [3]) that $\mathcal{A}$ generates a strongly continuous group $\{e^{t\mathcal{A}}\}_{t\geqslant 0}$ which is given by

$$e^{t\mathcal{A}} = \begin{pmatrix} C_t & (-A)^{-\frac{1}{2}}S_t \\ -(-A)^{\frac{1}{2}}S_t & C_t \end{pmatrix}, \tag{3.3}$$

where $C_t := \cos((-A)^{\frac{1}{2}})t)$ and $S_t := \sin((-A)^{\frac{1}{2}})t)$. For any $x \in \mathcal{H}$, we have $|||e^{\mathcal{A}t}x|||_0 \leqslant |||x\,|||_0$. Moreover, under the assumptions on f and g, one can check that $F \in C_b^{2,\eta}(H \times H, H)$ and $G \in C_B^{2,\eta}(H \times H, H)$. By definition, we further have $\mathcal{F} \in C_b^{2,\eta}(\mathcal{H} \times H, \mathcal{H}^1)$ and $\mathcal{G} \in C_B^{2,\eta}(\mathcal{H} \times H, H)$. Furthermore, according to [36, Lemma 3.7], we also have that $\bar{\mathcal{F}} \in C_b^2(\mathcal{H}, \mathcal{H}^1)$.

The Poisson equation will be the crucial tool in our paper. Recall that $\mathcal{L}_2(u, y)$ is defined by (2.10). If there is no confusion possible, we shall also write

$$\mathcal{L}_2\varphi(y) := \mathcal{L}_2(x, y)\varphi(y) := \mathcal{L}_2(\Pi_1(x), y)\varphi(y), \quad \forall \varphi \in C_\ell^2(H). \tag{3.4}$$

Consider the following Poisson equation:

$$\mathcal{L}_2(x, y)\psi(x, y) = -\phi(x, y), \tag{3.5}$$

where $x \in \mathcal{H}$ is regarded as a parameter, and $\phi : \mathcal{H} \times H \to \hat{H}$ is measurable. To be well-defined, it is necessary to make the following "centering" assumption on ϕ:

$$\int_H \phi(x, y)\mu^x(\mathrm{d}y) = 0, \quad \forall x \in \mathcal{H}. \tag{3.6}$$

The following result has been proven in [36, Theorem 3.2].

Theorem 3.1. *Let $\eta > 0$ and $k = 0, 1, 2$, and assume $\mathcal{G} \in C_B^{k,\eta}(\mathcal{H} \times H, H)$. Then for every $\phi(\cdot, \cdot) \in C_\ell^{k,\eta}(\mathcal{H} \times H, \hat{H})$ satisfying (3.6), there exists a unique solution $\psi(\cdot, \cdot) \in \psi \in C_\ell^{k,0}(\mathcal{H} \times H, \hat{H}) \cap \mathbb{C}_\ell^{0,2}(\mathcal{H} \times H, \hat{H})$ to equation (3.5) which is given by*

$$\psi(x, y) = \int_0^\infty \mathbb{E}\big[\phi(x, Y_t^x(y))\big]\mathrm{d}t,$$

where $Y_t^x(y) = Y_t^u(y)$ satisfies the frozen equation (2.4).

3.2. **Moment estimates.** We prove the following estimates for the solution X_t^ε and Y_t^ε of system (3.1).

Lemma 3.2. *Let $T > 0$, $x \in \mathcal{H}^1, y \in H$, and let $(X_t^\varepsilon, Y_t^\varepsilon)$ satisfy*

$$\begin{cases} X_t^\varepsilon = e^{t\mathcal{A}}x + \displaystyle\int_0^t e^{(t-s)\mathcal{A}}\mathcal{F}(X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s + \int_0^t e^{(t-s)\mathcal{A}}B\mathrm{d}W_s^1, \\ Y_t^\varepsilon = e^{\frac{t}{\varepsilon}A}y + \varepsilon^{-1}\displaystyle\int_0^t e^{\frac{t-s}{\varepsilon}A}\mathcal{G}(X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s + \varepsilon^{-1/2}\int_0^t e^{\frac{t-s}{\varepsilon}A}\mathrm{d}W_s^2. \end{cases} \tag{3.7}$$

Then for any $q \geqslant 1$, we have

$$\sup_{\varepsilon\in(0,1)} \mathbb{E}\Big(\sup_{t\in[0,T]} |||X_t^\varepsilon\,|||_1^{2q}\Big) \leqslant C_{T,q}\big(1 + |||x\,|||_1^{2q} + \|y\|^{2q}\big)$$

and

$$\sup_{\varepsilon\in(0,1)} \sup_{t\in[0,T]} \mathbb{E}\|Y_t^\varepsilon\|^{2q} + \sup_{\varepsilon\in(0,1)} \mathbb{E}\left(\int_0^T \|Y_t^\varepsilon\|_1^2\mathrm{d}t\right)^q \leqslant C_{T,q}(1 + \|y\|^{2q}), \tag{3.8}$$

where $C_{T,q} > 0$ is a constant.

Proof. Applying Itô's formula (see e.g. [31, Section 4.2]) to $\|Y_t^\varepsilon\|^{2q}$ and taking expectation, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\|Y_t^\varepsilon\|^{2q} = \frac{2q}{\varepsilon}\mathbb{E}\left[\|Y_t^\varepsilon\|^{2q-2}\langle AY_t^\varepsilon, Y_t^\varepsilon\rangle\right] + \frac{2q}{\varepsilon}\mathbb{E}\left[\|Y_t^\varepsilon\|^{2q-2}\langle \mathcal{G}(X_t^\varepsilon, Y_t^\varepsilon), Y_t^\varepsilon\rangle\right] \\ + \Big(\frac{q}{\varepsilon} + \frac{2q(q-1)}{\varepsilon}\Big)Tr(Q_2)\mathbb{E}\|Y_t^\varepsilon\|^{2q-2}.\end{aligned}$$

It follows from Poincaré inequality, Young's inequality and (2.3) that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\|Y_t^\varepsilon\|^{2q} &\leqslant -\frac{2q\lambda_1}{\varepsilon}\mathbb{E}\|Y_t^\varepsilon\|^{2q} + \frac{2qC_0}{\varepsilon}\mathbb{E}\|Y_t^\varepsilon\|^{2q-1} + \Big(\frac{q}{\varepsilon} + \frac{2q(q-1)}{\varepsilon}\Big)Tr(Q_2)\mathbb{E}\|Y_t^\varepsilon\|^{2q-2} \\ &\leqslant -\frac{qC_0}{\varepsilon}\mathbb{E}\|Y_t^\varepsilon\|^{2q} + \frac{C_0}{\varepsilon}.\end{aligned}$$

Using Gronwall's inequality, we obtain

$$\mathbb{E}\|Y_t^\varepsilon\|^{2q} \leqslant e^{-\frac{qC_0}{\varepsilon}t}\|y\|^{2q} + \frac{C_0}{\varepsilon}\int_0^t e^{-\frac{qC_0}{\varepsilon}(t-s)}\mathrm{d}s \leqslant C_0(1 + \|y\|^{2q}). \tag{3.9}$$

Furthermore, in view of [18, Theorem 5.3.5], the process $X_t^\varepsilon = (U_t^\varepsilon, V_t^\varepsilon)^T$ enjoys the following energy equality:

$$|||X_t^\varepsilon|||_1^2 = |||x|||_1^2 + 2\int_0^t \langle V_s^\varepsilon, F(U_s^\varepsilon, Y_s^\varepsilon)\rangle\mathrm{d}s + 2\int_0^t \langle U_t^\varepsilon, \mathrm{d}W_s^1\rangle + \int_0^t TrQ_1\mathrm{d}s.$$

Then it is easy to check that

$$|||X_t^\varepsilon|||_1^{2q} \leqslant C_0\bigg(1 + |||x|||_1^{2q} + \Big|\int_0^t \langle V_s^\varepsilon, F(U_s^\varepsilon, Y_s^\varepsilon)\rangle\mathrm{d}s\Big|^q + \Big|\int_0^t \langle U_t^\varepsilon, \mathrm{d}W_s^1\rangle\Big|^q\bigg). \tag{3.10}$$

On the one hand, note that

$$\begin{aligned}&\mathbb{E}\sup_{0\leqslant t\leqslant T}\Big|\int_0^t \langle V_s^\varepsilon, F(U_s^\varepsilon, Y_s^\varepsilon)\rangle\mathrm{d}s\Big|^q \\ &\leqslant C_1\mathbb{E}\Big(\int_0^T \|V_s^\varepsilon\|^2\mathrm{d}s\Big)^q + C_1\mathbb{E}\left(\int_0^T (1 + \|U_s^\varepsilon\|^2 + \|Y_s^\varepsilon\|^2)\mathrm{d}s\right)^q \\ &\leqslant C_1\mathbb{E}\left(\int_0^T |||X_s^\varepsilon|||_1^{2q}\,\mathrm{d}s\right) + C_1\mathbb{E}\left(\int_0^T (1 + \|Y_s^\varepsilon\|^{2q})\mathrm{d}s\right).\end{aligned} \tag{3.11}$$

On the other hand, in view of Burkholder-Davis-Gundy's inequality, we have

$$\mathbb{E}\sup_{0\leqslant t\leqslant T}\Big|\int_0^t \langle U_t^\varepsilon, \mathrm{d}W_s^1\rangle\Big|^q \leqslant C_2TrQ_1\mathbb{E}\left(\int_0^T \|U_s^\varepsilon\|^2\mathrm{d}s\right)^{\frac{q}{2}} \leqslant C_2\mathbb{E}\left(\int_0^T |||X_s^\varepsilon|||_1^{2q}\,\mathrm{d}s\right). \tag{3.12}$$

Combining (3.11) and (3.12) with (3.10), we get

$$\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}\|\|X_t^\varepsilon\|\|_1^{2q}\Big)\leqslant C_3(1+\|\|x\|\|_1^{2q})+C_3\mathbb{E}\left(\int_0^T\|\|X_s^\varepsilon\|\|_1^{2q}+\|Y_s^\varepsilon\|^{2q}\mathrm{d}s\right).$$

Thus, it follows from Gronwall's inequality that

$$\mathbb{E}(\sup_{0\leqslant t\leqslant T}\|\|X_t^\varepsilon\|\|_1^{2q})\leqslant C_4\left(1+\|\|x\|\|_1^{2q}+\int_0^T\mathbb{E}\|Y_s^\varepsilon\|^{2q}\mathrm{d}s\right),$$

which together with (3.9) yields

$$\mathbb{E}\Big(\sup_{0\leqslant t\leqslant T}\|\|X_t^\varepsilon\|\|_1^{2q}\Big)\leqslant C_5(1+\|\|x\|\|_1^{2q}+\|y\|^{2q}).$$

In order to prove estimate (3.8), we deduce that

$$\begin{aligned}\mathbb{E}\left(\int_0^T\|Y_t^\varepsilon\|_1^2\mathrm{d}t\right)^q\leqslant{}& C_q\left(\int_0^T\left\|e^{\frac{t}{\varepsilon}A}y\right\|_1^2\mathrm{d}t\right)^q\\&+C_q\,\mathbb{E}\left(\int_0^T\left\|\varepsilon^{-1}\int_0^t e^{\frac{t-s}{\varepsilon}A}\mathcal{G}((X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right\|_1^2\mathrm{d}t\right)^q\\&+C_q\,\mathbb{E}\left(\int_0^T\left\|\varepsilon^{-1/2}\int_0^t e^{\frac{t-s}{\varepsilon}A}\mathrm{d}W_s^2\right\|_1^2\mathrm{d}t\right)^q=:\sum_{i=1}^3\mathscr{Y}_i(T,\varepsilon).\end{aligned}$$

For the first term, we have

$$\begin{aligned}\mathscr{Y}_1(T,\varepsilon)&\leqslant C_6\left(\int_0^{T/\varepsilon}\sum_{k=1}^\infty\lambda_k e^{-2\lambda_k t}\langle y,e_k\rangle^2\mathrm{d}t\right)^q\\&\leqslant C_6\left(\sum_{k=1}^\infty(1-e^{\frac{-2\lambda_k T}{\varepsilon}})\langle y,e_k\rangle^2\right)^q\leqslant C_6\|y\|^{2q}.\end{aligned}$$

Note that

$$\begin{aligned}\left\|\varepsilon^{-1}\int_0^t e^{\frac{t-s}{\varepsilon}A}\mathcal{G}(X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\right\|_1&\leqslant C_7\varepsilon^{-1}\int_0^t\left(\frac{t-s}{\varepsilon}\right)^{-1/2}e^{-\frac{\lambda_1(t-s)}{2\varepsilon}}\|\mathcal{G}(X_s^\varepsilon,Y_s^\varepsilon)\|\mathrm{d}s\\&\leqslant C_7\int_0^{t/\varepsilon}\frac{e^{-\frac{\lambda_1 s}{2}}}{s^{1/2}}\mathrm{d}s\leqslant C_7,\end{aligned}$$

which implies that

$$\mathscr{Y}_2(T,\varepsilon)\leqslant C_8.$$

For the last term, by Minkowski's inequality, Burkholder-Davis-Gundy's inequality and (2.3), we deduce that

$$\begin{aligned}
\mathscr{Y}_3(T,\varepsilon) &\leqslant C_9\Big\{ \int_0^T \Big(\mathbb{E}\Big\|\varepsilon^{-1/2}\int_0^t e^{\frac{t-s}{\varepsilon}A}\mathrm{d}W_s^2\Big\|_1^{2q}\Big)^{1/q}\mathrm{d}t\Big\}^q \\
&\leqslant C_9\Big\{ \int_0^T \Big(\mathbb{E}\Big(\varepsilon^{-1}\int_0^t \sum_{k=1}^\infty \lambda_k e^{-2\lambda_k\frac{t-s}{\varepsilon}}\langle Q_2 e_k, e_k\rangle \mathrm{d}s\Big)^q\Big)^{1/q}\mathrm{d}t\Big\}^q \\
&\leqslant C_9\Big\{ \int_0^T \Big(\mathbb{E}\Big(\int_0^{t/\varepsilon} \sum_{k=1}^\infty \lambda_k e^{-2\lambda_k s}\langle Q_2 e_k, e_k\rangle \mathrm{d}s\Big)^q\Big)^{1/q}\mathrm{d}t\Big\}^q \leqslant C_9.
\end{aligned}$$

Combining the above computations, we get the desired result. □

We also need the following estimate for $\mathcal{A}X_t^\varepsilon$.

Lemma 3.3. *Let $T>0$, $x=(u,v)^T\in\mathcal{H}^1$ and $y\in H$. Then for any $q\geqslant 1$ and $t\in[0,T]$, we have*

$$\mathbb{E}\,|||\,\mathcal{A}X_t^\varepsilon|||_0^q \leqslant C_{T,q}(1+|||x\,|||_1^q+\|y\|^q),$$

where $C_{T,q}>0$ is a constant.

Proof. By definition, we have

$$\mathcal{A}X_t^\varepsilon = \begin{pmatrix} 0 & I \\ A & 0\end{pmatrix}\begin{pmatrix} U_t^\varepsilon \\ V_t^\varepsilon\end{pmatrix} = \begin{pmatrix} V_t^\varepsilon \\ AU_t^\varepsilon\end{pmatrix}.$$

Thus, we deduce that

$$\begin{aligned}
|||\mathcal{A}X_t^\varepsilon|||_0^q &\leqslant C_q\left(\|V_t^\varepsilon\|^q + \|AU_t^\varepsilon\|_{-1}^q\right) \\
&= C_q\left(\|V_t^\varepsilon\|^q + \|(-A)^{\frac12}U_t^\varepsilon\|^q\right).
\end{aligned}$$

It then follows from (3.7) that

$$\begin{aligned}
\mathbb{E}\|(-A)^{\frac12}U_t^\varepsilon\|^q \leqslant & C_q\big(\|(-A)^{\frac12}C_t u\| + \|S_t v\|\big)^q + C_q\mathbb{E}\Big\|\int_0^t S_{t-s}F(U_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s\Big\|^q \\
&+ C_q\mathbb{E}\Big\|\int_0^t S_{t-s}\mathrm{d}W_s^1\Big\|^q := \sum_{i=1}^3 \mathscr{U}_i(t,\varepsilon).
\end{aligned}$$

For the first term, we have

$$\mathscr{U}_1(t,\varepsilon)\leqslant C_1\,|||\,x\,|||_1^q\,.$$

To control the second term, by Minkowski's inequality and Lemma 3.2, we get

$$\mathscr{U}_2(t,\varepsilon)\leqslant C_2\Big(\int_0^t\big(1+\mathbb{E}\|U_s^\varepsilon\|^q+\mathbb{E}\|Y_s^\varepsilon\|^q\big)^{1/q}\mathrm{d}s\Big)^q \leqslant C_2(1+|||x\,|||_1^q+\|y\|^q).$$

Finally, by Burkholder-Davis-Gundy's inequality, we obtain

$$\mathscr{U}_3(t,\varepsilon) \leqslant C_3.$$

Combining the above estimates, we have

$$\mathbb{E}\|(-A)^{\frac{1}{2}}U_t^\varepsilon\|^q \leqslant C_4(1+ |||x|||_1^q + \|y\|^q).$$

Note that

$$V_t^\varepsilon = -(-A)^{\frac{1}{2}}S_t u + C_t v + \int_0^t C_{t-s}F(U_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s + \int_0^t C_{t-s}\mathrm{d}W_s^1.$$

In a similar way, we can prove that

$$\mathbb{E}\|V_t^\varepsilon\|^q \leqslant C_5(1+ |||x|||_1^q + \|y\|^q).$$

Combining the above, we get the desired result. □

The following estimates for the solution of the averaged equation (3.2) can be proved in a similar way as Lemmas 3.2 and 3.3, hence we omit the details here.

Lemma 3.4. *Let $T>0$ and $x \in \mathcal{H}^1$. The averaged equation (3.2) admits a unique mild solution $\bar{X}_t$ such that for all $t \geqslant 0$,*

$$\bar{X}_t = e^{t\mathcal{A}}x + \int_0^t e^{(t-s)\mathcal{A}}\bar{\mathcal{F}}(\bar{X}_s)\mathrm{d}s + \int_0^t e^{(t-s)\mathcal{A}}B\mathrm{d}W_s^1. \tag{3.13}$$

Moreover, for any $q \geqslant 1$ we have

$$\sup_{\varepsilon\in(0,1)} \mathbb{E}\Big(\sup_{t\in[0,T]} |||\bar{X}_t|||_1^{2q}\Big) \leqslant C_{T,q}(1+|||x|||_1^{2q})$$

and

$$\mathbb{E}\, |||\mathcal{A}\bar{X}_t|||_0^q \leqslant C_{T,q}(1+|||x|||_1^q),$$

where $C_{T,q}>0$ is a constant.

4. Strong and weak convergence in the averaging principle

4.1. **Galerkin approximation.** Itô's formula will be used frequently below in the proof of the main result. However, due to the persence of unbounded operators in the equation, we can not apply Itô's formula for SPDE (3.1) directly. For this reason, we use the following Galerkin approximation scheme, which reduces the infinite dimensional setting to a finite dimensional one. For every $n \in \mathbb{N}$, let $H^n = span\{e_1, e_2, \cdots, e_n\}$. Denote the projection of H onto H^n by P_n, and set

$$\mathcal{F}_n(x,y) := \begin{pmatrix} 0 \\ P_nF(\Pi_1(x),y) \end{pmatrix}, \quad \mathcal{G}_n(x,y) := P_nG(\Pi_1(x),y).$$

It is easy to check that $\mathcal{F}_n$ and $\mathcal{G}_n$ satisfy the same conditions as $\mathcal{F}$ and $\mathcal{G}$ with bounds which are uniform with respect to n. Consider the following finite dimensional system:

$$\begin{cases} \mathrm{d}X_t^{n,\varepsilon} = \mathcal{A}X_t^{n,\varepsilon}\mathrm{d}t + \mathcal{F}_n(X_t^{n,\varepsilon}, Y_t^{n,\varepsilon})\mathrm{d}t + P_n\mathrm{d}W_t^1, \\ \mathrm{d}Y_t^{n,\varepsilon} = \varepsilon^{-1}AY_t^{n,\varepsilon}\mathrm{d}t + \varepsilon^{-1}\mathcal{G}_n(X_t^{n,\varepsilon}, Y_t^{n,\varepsilon})\mathrm{d}t + \varepsilon^{-1/2}P_n\mathrm{d}W_t^2, \end{cases} \tag{4.1}$$

with initial values $X_0^{n,\varepsilon} = x^n \in H^n \times H^n$ and $Y_0^{n,\varepsilon} = y^n \in H^n$. The corresponding averaged equation for system (5.1) is given by

$$\mathrm{d}\bar{X}_t^n = \mathcal{A}\bar{X}_t^n\mathrm{d}t + \bar{\mathcal{F}}_n(\bar{X}_t^n)\mathrm{d}t + P_n\mathrm{d}W_t^1, \quad \bar{X}_0^n = x^n \in H^n \times H^n, \tag{4.2}$$

where

$$\bar{\mathcal{F}}_n(x) := \int_{H^n} \mathcal{F}_n(x,y)\mu_n^x(\mathrm{d}y), \tag{4.3}$$

and $\mu_n^x(\mathrm{d}y)$ is the invariant measure associated with the transition semigroup of the process $Y_t^{x,n}(y)$ which satisfies the frozen equation

$$\mathrm{d}Y_t^{x,n} = AY_t^{x,n}\mathrm{d}t + \mathcal{G}_n(x^n, Y_t^{x,n})\mathrm{d}t + P_n\mathrm{d}W_t^2, \ \ Y_0^{x,n} = y^n \in H^n.$$

Recall that $Y_t^u(y)$ satisfies (2.4) and note that $\mathcal{G}(x,y) = G(u,y)$. We know that $Y_t^{x,n}(y^n)$ converges strongly to $Y_t^x(y) := Y_t^u(y)$. Let $T > 0, x \in \mathcal{H}^1$ and $y \in H$. Then as shown in the proof of [17, Lemma 3.1], for any $q \geqslant 1$ and $t \in [0,T]$, we have

$$\lim_{n\to\infty} \mathbb{E} \,|\!|\!| \, X_t^\varepsilon - X_t^{n,\varepsilon} |\!|\!|_1^q = 0.$$

Furthermore, in view of (3.7), (3.13) and (3.3) we deduce that

$$\begin{aligned} \mathbb{E} \,|\!|\!| \, \bar{X}_t^n - \bar{X}_t |\!|\!|_1^q \leqslant{}& \mathbb{E}\Big|\!\Big|\!\Big| \int_0^t e^{(t-s)\mathcal{A}}(I - P_n)B\mathrm{d}W_s^1 \Big|\!\Big|\!\Big|_1^q \\ &+ \mathbb{E}\left(\int_0^t \left(\left\|(-A)^{-\frac{1}{2}}S_{t-s}(\bar{F}(\bar{U}_s) - \bar{F}_n(\bar{U}_s))\right\|_1 + \left\|C_{t-s}(\bar{F}(\bar{X}_s) - \bar{F}_n(\bar{U}_s)\right\|\right)\mathrm{d}s\right)^q \\ &+ \mathbb{E}\left(\int_0^t \left(\left\|(-A)^{-\frac{1}{2}}S_{t-s}(\bar{F}_n(\bar{U}_s) - \bar{F}_n(\bar{U}_s^n))\right\|_1 + \left\|C_{t-s}(\bar{F}_n(\bar{U}_s) - \bar{F}_n(\bar{U}_s^n))\right\|\right)\mathrm{d}s\right)^q. \end{aligned}$$

Since $\|\bar{F}_n - \bar{F}\| \to 0$ as $n \to \infty$ (see e.g. [5, (4.4)]), the first two terms go to 0 as $n \to \infty$ by the dominated convergence theorem. For the last term, we have

$$\begin{aligned} &\mathbb{E}\left(\int_0^t \left(\left\|(-A)^{-\frac{1}{2}}S_{t-s}\big(\bar{F}_n(\bar{U}_s) - \bar{F}_n(\bar{U}_s^n)\big)\right\|_1 + \left\|C_{t-s}(\bar{F}_n(\bar{U}_s) - \bar{F}_n(\bar{U}_s^n))\right\|\right)\mathrm{d}s\right)^q \\ &\leqslant C_1 \,\mathbb{E}\left(\int_0^t \|\bar{U}_s - \bar{U}_s^n\|_1 \mathrm{d}s\right)^q \leqslant C_1 \,\mathbb{E}\left(\int_0^t |\!|\!|\bar{X}_s - \bar{X}_s^n \,|\!|\!|_1 \,\mathrm{d}s\right)^q, \end{aligned}$$

which in turn yields by Gronwall's inequality that

$$\lim_{n\to\infty} \mathbb{E} \,|\!|\!| \, \bar{X}_t^n - \bar{X}_t |\!|\!|_1^q = 0.$$

Therefore, in order to prove Theorem 2.1, we only need to show that for any $q \geqslant 1$,

$$\sup_{t\in[0,T]} \mathbb{E} \, ||| \, X_t^{n,\varepsilon} - \bar{X}_t^n |||_1^q \leqslant C_T \, \varepsilon^{q/2}, \tag{4.4}$$

and for every $\varphi \in \mathbb{C}_b^3(\mathcal{H})$,

$$\sup_{t\in[0,T]} \left| \mathbb{E}[\varphi(X_t^{n,\varepsilon})] - \mathbb{E}[\varphi(\bar{X}_t^n)] \right| \leqslant C_T \, \varepsilon, \tag{4.5}$$

where $C_T > 0$ **is a constant independent of** n. In the rest of this section, we shall only work with the approximating system (4.1), and prove bounds that are uniform with respect to n. To simplify the notations, we omit the index n. In particular, the space H^n is denoted by H.

4.2. **Proof of Theorem 2.1 (strong convergence).** For simplicity, let

$$\begin{aligned}\mathcal{L}_1\varphi(x) := \mathcal{L}_1(x,y)\varphi(x) : &= \langle \mathcal{A}x + \mathcal{F}(x,y), D_x\varphi(x)\rangle_{\mathcal{H}} \\ &\quad + \frac{1}{2} Tr\left(D_x^2\varphi(x)(BQ_1)^{\frac{1}{2}}((BQ_1)^{\frac{1}{2}})^*\right), \quad \forall \varphi \in C_\ell^2(\mathcal{H}). \end{aligned} \tag{4.6}$$

As shown in Subsection 4.1, to prove the strong convergence result (2.7), we only need to prove (4.4). To this end, we first establish the following fluctuation estimate for an integral functional of $(X_s^\varepsilon, Y_s^\varepsilon)$ over time interval $[0,t]$, which will play an important role in proving (4.4).

Lemma 4.1 (Strong fluctuation estimate)**.** *Let $T, \eta > 0, x = (u,v)^T \in \mathcal{H}^1$ and $y \in H$. Assume that $\mathcal{F} \in C_b^{2,\eta}(\mathcal{H}\times H, \mathcal{H}^1)$ and $\mathcal{G} \in C_B^{2,\eta}(\mathcal{H}\times H, H)$. Then for any $t \in [0,T]$, $q \geqslant 1$ and every $\tilde{\phi}(x,y) := \begin{pmatrix} 0 \\ \phi(u,y) \end{pmatrix}$ satisfying (3.6) with $\phi \in C_b^{2,\eta}(H\times H, H)$, we have*

$$\mathbb{E}\left|\left|\left| \int_0^t e^{(t-s)\mathcal{A}} \tilde{\phi}(X_s^\varepsilon, Y_s^\varepsilon) \mathrm{d}s \right|\right|\right|_1^q \leqslant C_{T,q} \, \varepsilon^{q/2},$$

where $C_{T,q} > 0$ is a constant independent of ε, η and n.

Proof. Let ψ solve the Poisson equation,

$$\mathcal{L}_2(u,y)\psi(u,y) = -\phi(u,y),$$

and define

$$\tilde{\psi}_t(s,x,y) := e^{(t-s)\mathcal{A}}\tilde{\psi}(x,y) := e^{(t-s)\mathcal{A}} \begin{pmatrix} 0 \\ \psi(u,y) \end{pmatrix}.$$

Since $\mathcal{L}_2$ is an operator with respect to the y-variable, one can check that

$$\mathcal{L}_2\tilde{\psi}_t(s,x,y) = -e^{(t-s)\mathcal{A}}\tilde{\phi}(x,y). \tag{4.7}$$

Applying Itô's formula to $\tilde{\psi}_t(t, X_t^\varepsilon, Y_t^\varepsilon)$, we get

$$\tilde{\psi}_t(t, X_t^\varepsilon, Y_t^\varepsilon) = \tilde{\psi}_t(0,x,y) + \int_0^t (\partial_s + \mathcal{L}_1)\tilde{\psi}_t(s, X_s^\varepsilon, Y_s^\varepsilon)\mathrm{d}s$$

$$+\frac{1}{\varepsilon}\int_0^t \mathcal{L}_2\tilde{\psi}_t(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s+M_t^1+\frac{1}{\sqrt{\varepsilon}}M_t^2, \tag{4.8}$$

where M_t^1 and M_t^2 are defined by

$$M_t^1:=\int_0^t D_x\tilde{\psi}_t(s,X_s^\varepsilon,Y_s^\varepsilon)B\mathrm{d}W_s^1 \quad \text{and} \quad M_t^2:=\int_0^t D_y\tilde{\psi}_t(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}W_s^2.$$

Multiplying both sides of (4.8) by ε and using (4.7), we obtain

$$\begin{aligned}\int_0^t e^{(t-s)\mathcal{A}}\tilde{\phi}(X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s&=-\int_0^t \mathcal{L}_2\tilde{\psi}_t(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\\ &=\varepsilon\big[\tilde{\psi}_t(0,x,y)-\tilde{\psi}_t(t,X_t^\varepsilon,Y_t^\varepsilon)\big]+\varepsilon\int_0^t\partial_s\tilde{\psi}_t(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\\ &\quad+\varepsilon\int_0^t\mathcal{L}_1\tilde{\psi}_t(s,X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s+\varepsilon M_t^1+\sqrt{\varepsilon}M_t^2=:\sum_{i=1}^5\mathscr{J}_i(t,\varepsilon).\end{aligned} \tag{4.9}$$

According to Theorem 3.1 , we have that $\psi\in C_\ell^{2,2}(H\times H,H)$ and hence

$$\begin{aligned}|||e^{(t-s)\mathcal{A}}\tilde{\psi}(x,y)|||_1&=\Big|\Big|\Big|\begin{pmatrix}(-A)^{-\frac{1}{2}}S_{t-s}\psi(u,y)\\ C_{t-s}\psi(u,y)\end{pmatrix}\Big|\Big|\Big|_1\\ &=\|(-A)^{-\frac{1}{2}}S_{t-s}\psi(u,y)\|_1+\|C_{t-s}\psi(u,y)\|\\ &\leqslant 2\|\psi(u,y)\|\leqslant C_1(1+\|u\|+\|y\|).\end{aligned}$$

As a result, by Lemma 3.2 we get

$$\mathbb{E}\,|||\,\mathscr{J}_1(t,\varepsilon)|||_1^q\leqslant C_1\,\varepsilon^q(1+\mathbb{E}\|U_t^\varepsilon\|^q+\mathbb{E}\|Y_t^\varepsilon\|^q)\leqslant C_1\,\varepsilon^q.$$

Note that

$$\partial_s\tilde{\psi}_t(s,x,y)=-\mathcal{A}e^{(t-s)\mathcal{A}}\tilde{\psi}(x,y),$$

and that

$$\begin{aligned}|||\mathcal{A}e^{(t-s)\mathcal{A}}\tilde{\psi}(x,y)|||_1&=\Big|\Big|\Big|\begin{pmatrix}C_{t-s}\psi(u,y)\\ -(-\mathcal{A})^{\frac{1}{2}}S_{t-s}\psi(u,y)\end{pmatrix}\Big|\Big|\Big|_1\\ &=\|C_{t-s}\psi(u,y)\|_1+\|-(-A)^{\frac{1}{2}}S_{t-s}\psi(u,y)\|\\ &\leqslant 2\|\psi(u,y)\|_1\leqslant C_2(1+\|u\|_1^2+\|y\|_1^2),\end{aligned}$$

where the last inequality can be obtained as in [13, (2.16)]. Thus, using Minkowski's inequality and Lemma 3.2 again, we have

$$\mathbb{E}\,|||\,\mathscr{J}_2(t,\varepsilon)|||_1^q\leqslant C_2\,\varepsilon^q\Big(\int_0^T\big(1+\mathbb{E}\|U_s^\varepsilon\|_1^{2q}\big)^{1/q}\mathrm{d}t\Big)^q+C_2\,\varepsilon^q\,\mathbb{E}\Big(\int_0^T\|Y_s^\varepsilon\|_1^2\mathrm{d}s\Big)^q\leqslant C_2\,\varepsilon^q.$$

For the third term, we have

$$|\mathcal{L}_1\tilde{\psi}_t(s,X_s^\varepsilon,Y_s^\varepsilon)|\leqslant C_3\big(1+|||\mathcal{A}X_s^\varepsilon\,|||_0^2+|||\,X_s^\varepsilon\,|||_1^2+\|Y_s^\varepsilon\|^2\big),$$

which together with Minkowski's inequality, Lemmas 3.2 and 3.3 yields that

$$\mathbb{E} \,|||\, \mathscr{J}_3(t,\varepsilon)|||_1^q \leqslant C_3\,\varepsilon^q\Big(\int_0^t \Big(\mathbb{E}\big(1+|||\mathcal{A}X_s^\varepsilon\,|||_0^2+|||\,X_s^\varepsilon\,|||_1^2+\|Y_s^\varepsilon\|^2\big)^q\Big)^{1/q}\mathrm{d}s\Big)^q \leqslant C_3\,\varepsilon^q.$$

Finally, by Burkholder-Davis-Gundy's inequality, Theorem 3.1, Lemma 3.2 and (2.3), we have

$$\begin{aligned}\mathbb{E}\,|||\,\mathscr{J}_4(t,\varepsilon)|||_1^q &\leqslant C_4\,\varepsilon^q\Big(\int_0^T \mathbb{E}\big\|e^{(t-s)\mathcal{A}}D_x\tilde{\psi}(X_s^\varepsilon,Y_s^\varepsilon)BQ_1^{\frac{1}{2}}\big\|^2_{\mathscr{L}_2(\mathcal{H}^1)}\mathrm{d}s\Big)^{q/2}\\ &\leqslant C_4\,\varepsilon^q\left(\int_0^T(1+\mathbb{E}\,|||\,X_s^\varepsilon\,|||_1^2+\mathbb{E}\|Y_s^\varepsilon\|^2)\mathrm{d}s\right)^{q/2}\leqslant C_4\,\varepsilon^q,\end{aligned}$$

and similarly,

$$\mathbb{E}\,|||\,\mathscr{J}_5(t,\varepsilon)|||_1^q \leqslant C_5\,\varepsilon^{q/2}.$$

Combining the above inequalities with (4.9), we get the desired estimate. □

We are now in the position to give:

Proof of estimate (4.4). Fix $T>0$ below. In view of (3.7) and (3.13), for every $t\in[0,T]$ we have

$$X_t^\varepsilon-\bar{X}_t=\int_0^t e^{(t-s)\mathcal{A}}\big[\bar{\mathcal{F}}(X_s^\varepsilon)-\bar{\mathcal{F}}(\bar{X}_s)\big]\mathrm{d}s+\int_0^t e^{(t-s)\mathcal{A}}\delta\mathcal{F}(X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s,$$

where $\delta\mathcal{F}$ is defined by

$$\delta\mathcal{F}(x,y):=\mathcal{F}(x,y)-\bar{\mathcal{F}}(x)=\begin{pmatrix}0\\ \delta F(\Pi_1(x),y)\end{pmatrix}. \tag{4.10}$$

Thus, we have for any $q\geqslant 1$,

$$\begin{aligned}\mathbb{E}\,|||\,X_t^\varepsilon-\bar{X}_t|||_1^q &\leqslant C_0\,\mathbb{E}\Big|\Big|\Big|\int_0^t e^{(t-s)\mathcal{A}}\big[\bar{\mathcal{F}}(X_s^\varepsilon)-\bar{\mathcal{F}}(\bar{X}_s)\big]\mathrm{d}s\Big|\Big|\Big|_1^q\\ &\quad+C_0\,\mathbb{E}\Big|\Big|\Big|\int_0^t e^{(t-s)\mathcal{A}}\delta\mathcal{F}(X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\Big|\Big|\Big|_1^q=:\mathscr{I}_1(t,\varepsilon)+\mathscr{I}_2(t,\varepsilon).\end{aligned}$$

Since $\bar{\mathcal{F}}\in C_b^2(\mathcal{H},\mathcal{H}^1)$, by Minkowski's inequality we deduce that

$$\mathscr{I}_1(t,\varepsilon)\leqslant C_1\mathbb{E}\Big(\int_0^t |||\bar{\mathcal{F}}(X_s^\varepsilon)-\bar{\mathcal{F}}(\bar{X}_s)\,|||_1\,\mathrm{d}s\Big)^q\leqslant C_1\int_0^t\mathbb{E}\,|||\,X_s^\varepsilon-\bar{X}_s\,|||_1^q\,\mathrm{d}s.$$

For the second term, noting that $\delta\mathcal{F}(x,y)$ satisfies the centering condition (3.6), it follows by Lemma 4.1 directly that

$$\mathscr{I}_2(t,\varepsilon)\leqslant C_2\,\varepsilon^{q/2}.$$

Thus, we arrive at

$$\mathbb{E} \, ||| \, X_t^\varepsilon - \bar{X}_t |||_1^q \leqslant C_3 \, \varepsilon^{q/2} + C_3 \int_0^t \mathbb{E} \, ||| \, X_s^\varepsilon - \bar{X}_s \, |||_1^q \, \mathrm{d}s,$$

which together with Gronwall's inequality yields the desired result. □

4.3. **Proof of Theorem 2.1 (weak convergence).** As in the previous subsection, to prove the weak convergence result in Theorem 2.1 , we only need to show (4.5). The main reason for the difference between the strong and weak convergence rates in the averaging principle can be seen through the following estimate.

Lemma 4.2 (Weak fluctuation estimate)**.** *Let $T, \eta > 0, x = (u, v)^T \in \mathcal{H}^1$ and $y \in H$. Assume that $\mathcal{F} \in C_b^{2,\eta}(\mathcal{H} \times H, \mathcal{H}^1)$ and $\mathcal{G} \in C_B^{2,\eta}(\mathcal{H} \times H, H)$. Then for any $t \in [0, T]$, $\phi \in C_\ell^{1,2,\eta}([0, T] \times \mathcal{H} \times H)$ satisfying (3.6) and*

$$|\partial_t \phi(t, x, y)| \leqslant C_0(1 + |||x \, |||_1^2 + \|y\|^2), \tag{4.11}$$

we have

$$\mathbb{E} \left(\int_0^t \phi(s, X_s^\varepsilon, Y_s^\varepsilon) \mathrm{d}s \right) \leqslant C_T \, \varepsilon,$$

where $C_T > 0$ is a constant independent of ε, η and n.

Proof. Let ψ solve the Poisson equation

$$\mathcal{L}_2 \psi(t, x, y) = -\phi(t, x, y), \tag{4.12}$$

where $\mathcal{L}_2$ is given by (3.4). According to Theorem 3.1, we can apply Itô's formula to $\psi(t, X_t^\varepsilon, Y_t^\varepsilon)$ to get that

$$\begin{aligned} \mathbb{E}[\psi(t, X_t^\varepsilon, Y_t^\varepsilon)] = \psi(0, x, y) + \mathbb{E} \left(\int_0^t (\partial_s + \mathcal{L}_1) \psi(s, X_s^\varepsilon, Y_s^\varepsilon) \mathrm{d}s \right) \\ + \frac{1}{\varepsilon} \mathbb{E} \left(\int_0^t \mathcal{L}_2 \psi(s, X_s^\varepsilon, Y_s^\varepsilon) \mathrm{d}s \right). \end{aligned}$$

Combining this with (4.12), we obtain

$$\begin{aligned} &\mathbb{E} \left(\int_0^t \phi(s, X_s^\varepsilon, Y_s^\varepsilon) \mathrm{d}s \right) \\ &= \varepsilon \mathbb{E} \big[\psi(0, x, y) - \psi(t, X_t^\varepsilon, Y_t^\varepsilon) \big] + \varepsilon \mathbb{E} \left(\int_0^t \mathcal{L}_1 \psi(s, X_s^\varepsilon, Y_s^\varepsilon) \mathrm{d}s \right) \\ &\quad + \varepsilon \mathbb{E} \left(\int_0^t \partial_s \psi(s, X_s^\varepsilon, Y_s^\varepsilon) \mathrm{d}s \right) =: \sum_{i=1}^3 \mathscr{W}_i(t, \varepsilon). \end{aligned}$$

By using exactly the same arguments as in the proof of Lemma 4.1, we can get that

$$\mathscr{W}_1(t, \varepsilon) + \mathscr{W}_2(t, \varepsilon) \leqslant C_1 \, \varepsilon.$$

To control the third term, note that

$$\mathcal{L}_2 \partial_t \psi(t,x,y) = -\partial_t \phi(t,x,y).$$

In view of condition (4.11), we have

$$|\partial_t \psi(t,x,y)| \leqslant C_0(1 + |||x|||_1^2 + \|y\|^2),$$

which together with Lemma 3.2 implies that

$$\mathscr{W}_3(t,\varepsilon) \leqslant C_2\, \varepsilon \mathbb{E}\left(\int_0^t (1 + |||X_s^\varepsilon|||_1^2 + \|Y_s^\varepsilon\|^2) \mathrm{d}s\right) \leqslant C_2\, \varepsilon.$$

Combining the above estimates, we get the desired result. $\square$

Given $T > 0$, consider the following Cauchy problem on $[0,T] \times \mathcal{H}$:

$$\begin{cases} \partial_t \bar{u}(t,x) = \bar{\mathcal{L}}_1 \bar{u}(t,x), \quad t \in (0,T], \\ \bar{u}(0,x) = \varphi(x), \end{cases} \tag{4.13}$$

where $\varphi : \mathcal{H} \to \mathbb{R}$ is measurable and $\bar{\mathcal{L}}_1$ is formally the infinitesimal generator of the process $\bar{X}_t$ given by

$$\begin{aligned} \bar{\mathcal{L}}_1 \varphi(x) &= \langle \mathcal{A}x + \bar{\mathcal{F}}(x), D_x \varphi(x) \rangle_{\mathcal{H}} \\ &\quad + \frac{1}{2} Tr\left(D_x^2 \varphi(x) (BQ_1)^{\frac{1}{2}} ((BQ_1)^{\frac{1}{2}})^* \right), \quad \forall \varphi \in C_\ell^2(\mathcal{H}). \end{aligned} \tag{4.14}$$

The following result has been proven in [21, Lemmas A.3-A.5 and 4.3].

Lemma 4.3. *For every $\varphi \in \mathbb{C}_b^3(\mathcal{H})$, there exists a solution $\bar{u} \in C_b^{1,3}([0,T] \times \mathcal{H})$ to equation (4.13) which is given by*

$$\bar{u}(t,x) = \mathbb{E}\big[\varphi(\bar{X}_t(x))\big].$$

Moreover, for any $t \in [0,T]$ and $x, h \in \mathcal{H}^1$, we have

$$|\partial_t D_x \bar{u}(t,x).h| \leqslant C_T \,|||\, h \,|||_1\, (1 + |||x|||_1),$$

where $C_T > 0$ is a constant.

Now, we are in the position to give:

Proof of estimate (4.5). Given $T > 0$ and $\varphi \in \mathbb{C}_b^3(\mathcal{H})$, let $\bar{u}$ solve the Cauchy problem (4.13). For any $t \in [0,T]$ and $x \in \mathcal{H}^1$, define

$$\tilde{u}(t,x) := \bar{u}(T-t,x).$$

Then one can check that

$$\tilde{u}(T,x) = \bar{u}(0,x) = \varphi(x) \quad \text{and} \quad \tilde{u}(0,x) = \bar{u}(T,x) = \mathbb{E}[\varphi(\bar{X}_T(x))].$$

Using Itô's formula and taking expectation, we deduce that

$$\mathbb{E}[\varphi(X_T^\varepsilon)] - \mathbb{E}[\varphi(\bar{X}_T)] = \mathbb{E}[\tilde{u}(T, X_T^\varepsilon) - \tilde{u}(0,x)]$$

$$= \mathbb{E}\left(\int_0^T \left(\partial_t + \mathcal{L}_1\right)\tilde{u}(t, X_t^\varepsilon)\mathrm{d}t\right)$$
$$= \mathbb{E}\left(\int_0^T [\mathcal{L}_1\tilde{u}(t, X_t^\varepsilon) - \bar{\mathcal{L}}_1\tilde{u}(t, X_t^\varepsilon)]\mathrm{d}t\right)$$
$$= \mathbb{E}\left(\int_0^T \langle\delta\mathcal{F}(X_t^\varepsilon, Y_t^\varepsilon), D_x\tilde{u}(t, X_t^\varepsilon)\rangle_{\mathcal{H}}\mathrm{d}t\right).$$

Note that the function

$$\phi(t, x, y) := \langle\delta\mathcal{F}(x, y), D_x\tilde{u}(t, x)\rangle_{\mathcal{H}}$$

satisfies the centering condition (3.6). Moreover, by Lemma 4.3 we have

$$\partial_t\phi(t, x, y) = \langle\delta\mathcal{F}(x, y), \partial_t D_x\bar{u}(T - t, x)\rangle_{\mathcal{H}} \leqslant C_0(1 + |||x\,|||_1^2 + \|y\|^2).$$

As a result of Lemma 4.2, we have

$$\mathbb{E}[\varphi(X_T^\varepsilon)] - \mathbb{E}[\varphi(\bar{X}_T)] \leqslant C_1\,\varepsilon,$$

which completes the proof. □

5. Normal deviations

5.1. **Cauchy problem.** Define

$$\mathcal{Z}_t^\varepsilon := \frac{X_t^\varepsilon - \bar{X}_t}{\sqrt{\varepsilon}}.$$

In view of (3.1) and (3.2), we consider the process $(X_t^\varepsilon, Y_t^\varepsilon, \bar{X}_t, \mathcal{Z}_t^\varepsilon)$ as the solution to the following system of equations:

$$\begin{cases} \mathrm{d}X_t^\varepsilon = \mathcal{A}X_t^\varepsilon\mathrm{d}t + \mathcal{F}(X_t^\varepsilon, Y_t^\varepsilon)\mathrm{d}t + B\mathrm{d}W_t^1, & X_0^\varepsilon = x, \\ \mathrm{d}Y_t^\varepsilon = \varepsilon^{-1}AY_t^\varepsilon\mathrm{d}t + \varepsilon^{-1}\mathcal{G}(X_t^\varepsilon, Y_t^\varepsilon)\mathrm{d}t + \varepsilon^{-1/2}\mathrm{d}W_t^2, & Y_0^\varepsilon = y, \\ \mathrm{d}\bar{X}_t = \mathcal{A}\bar{X}_t\mathrm{d}t + \bar{\mathcal{F}}(\bar{X}_t)\mathrm{d}t + B\mathrm{d}W_t^1, & \bar{X}_0 = x, \\ \mathrm{d}\mathcal{Z}_t^\varepsilon = \mathcal{A}\mathcal{Z}_t^\varepsilon\mathrm{d}t + \varepsilon^{-1/2}[\bar{\mathcal{F}}(X_t^\varepsilon) - \bar{\mathcal{F}}(\bar{X}_t)]\mathrm{d}t + \varepsilon^{-1/2}\delta\mathcal{F}(X_t^\varepsilon, Y_t^\varepsilon)\mathrm{d}t, & Z_0^\varepsilon = 0, \end{cases} \tag{5.1}$$

where $\delta\mathcal{F}$ is defined by (4.10). As a result of Theorem 2.1, we have that for any $q \geqslant 1$,

$$\sup_{0\leqslant t\leqslant T} \mathbb{E}\,|||\,\mathcal{Z}_t^\varepsilon|||_1^q \leqslant C_T < \infty. \tag{5.2}$$

Furthermore, note that

$$\mathcal{A}\mathcal{Z}_t^\varepsilon = \begin{pmatrix} 0 & I \\ A & 0 \end{pmatrix}\begin{pmatrix} \frac{U_t^\varepsilon - \bar{U}_t}{\sqrt{\varepsilon}} \\ \frac{V_t^\varepsilon - \bar{V}_t}{\sqrt{\varepsilon}} \end{pmatrix} = \begin{pmatrix} \frac{V_t^\varepsilon - \bar{V}_t}{\sqrt{\varepsilon}} \\ \frac{A(U_t^\varepsilon - \bar{U}_t)}{\sqrt{\varepsilon}} \end{pmatrix},$$

hence we have

$$\begin{aligned}\mathbb{E}\,\|\!|\, \mathcal{A}\mathcal{Z}_t^\varepsilon \|\!|_0^q &= \mathbb{E}\left(\left\|\frac{V_t^\varepsilon - \bar{V}_t}{\sqrt{\varepsilon}}\right\|^2 + \left\|\frac{A(U_t^\varepsilon - \bar{U}_t)}{\sqrt{\varepsilon}}\right\|_{-1}^2\right)^{q/2} \\ &= \mathbb{E}\left(\left\|\frac{V_t^\varepsilon - \bar{V}_t}{\sqrt{\varepsilon}}\right\|^2 + \left\|\frac{(U_t^\varepsilon - \bar{U}_t)}{\sqrt{\varepsilon}}\right\|_{1}^2\right)^{q/2} \leqslant C_T < \infty. \qquad (5.3)\end{aligned}$$

Similarly, we rewrite (2.11) as

$$\mathrm{d}\bar{\mathcal{Z}}_t = \mathcal{A}\bar{\mathcal{Z}}_t\mathrm{d}t + D_x\bar{\mathcal{F}}(\bar{X}_t).\bar{\mathcal{Z}}_t\mathrm{d}t + \Sigma(\bar{X}_t)\mathrm{d}W_t, \qquad (5.4)$$

where $\bar{\mathcal{Z}}_t = (\bar{Z}_t, \dot{\bar{Z}}_t)^T$, and Σ is a Hilbert-Schmidt operator satisfying

$$\frac{1}{2}\Sigma(x)\Sigma^*(x) = \overline{\delta\mathcal{F}\otimes\tilde{\Psi}}(x) := \int_H \big[\delta\mathcal{F}(x,y)\otimes\tilde{\Psi}(x,y)\big]\mu^x(\mathrm{d}y), \qquad (5.5)$$

(see e.g. [10, (1.6)] and [38, (11)]), and $\tilde{\Psi}$ is the solution of the following Poisson equation:

$$\mathcal{L}_2(x,y)\tilde{\Psi}(x,y) = -\delta\mathcal{F}(x,y). \qquad (5.6)$$

Recall that $\mathcal{L}_2(x,y) = \mathcal{L}_2(u,y)$ and $\Psi(u,y)$ solves the Poisson equation (2.9). Thus, we have $\tilde{\Psi}(x,y) = \Psi(\Pi_1(x),y) = \Psi(u,y)$. Combining (3.2) and (5.4), the process $(\bar{X}_t, \bar{\mathcal{Z}}_t)$ solves the system

$$\begin{cases} \mathrm{d}\bar{X}_t = \mathcal{A}\bar{X}_t\mathrm{d}t + \bar{\mathcal{F}}(\bar{X}_t)\mathrm{d}t + \mathrm{d}W_t^1, & \bar{X}_0 = x, \\ \mathrm{d}\bar{\mathcal{Z}}_t = \mathcal{A}\bar{\mathcal{Z}}_t\mathrm{d}t + D_x\bar{\mathcal{F}}(\bar{X}_t).\bar{\mathcal{Z}}_t\mathrm{d}t + \Sigma(\bar{X}_t)\mathrm{d}W_t, & \bar{\mathcal{Z}}_0 = 0. \end{cases}$$

Note that the processes $\bar{X}_t$ and $\bar{\mathcal{Z}}_t$ depend on the initial value x. Below, we shall write $\bar{X}_t(x)$ when we want to stress its dependence on the initial value, and use $\bar{\mathcal{Z}}_t(x,z)$ to denote the process $\bar{\mathcal{Z}}_t$ with initial point $\bar{\mathcal{Z}}_0 = z \in \mathcal{H}$.

Given $T > 0$, consider the following Cauchy problem on $[0,T]\times\mathcal{H}\times\mathcal{H}$:

$$\begin{cases} \partial_t\bar{u}(t,x,z) = \bar{\mathcal{L}}\bar{u}(t,x,z), & t\in(0,T], \\ \bar{u}(0,x,z) = \varphi(z), \end{cases} \qquad (5.7)$$

where $\varphi : \mathcal{H} \to \mathbb{R}$ is measurable and $\bar{\mathcal{L}}$ is formally the infinitesimal generator of the Markov process $(\bar{X}_t, \bar{\mathcal{Z}}_t)$, i.e.,

$$\bar{\mathcal{L}} := \bar{\mathcal{L}}_1 + \bar{\mathcal{L}}_3,$$

with $\bar{\mathcal{L}}_1$ given by (4.14) and $\bar{\mathcal{L}}_3$ defined by

$$\begin{aligned}\bar{\mathcal{L}}_3\varphi(z) := \bar{\mathcal{L}}_3(x,z)\varphi(z) &:= \langle \mathcal{A}z + D_x\bar{\mathcal{F}}(x).z, D_z\varphi(z)\rangle_{\mathcal{H}} \\ &\quad + \frac{1}{2}Tr\big(D_z^2\varphi(z)\Sigma(x)\Sigma^*(x)\big), \quad \forall\varphi\in C_\ell^2(\mathcal{H}).\end{aligned}$$

We have the following result.

Lemma 5.1. *For every $\varphi \in \mathbb{C}_b^3(\mathcal{H})$, there exists a solution $\bar{u} \in C_b^{1,3,3}([0,T]\times\mathcal{H}\times\mathcal{H})$ to equation (5.7) which is given by*

$$\bar{u}(t,x,z) = \mathbb{E}\big[\varphi(\bar{\mathcal{Z}}_t(x,z))\big]. \tag{5.8}$$

Moreover, for any $t\in[0,T]$ and $x,z,h\in\mathcal{H}^1$, we have

$$\begin{aligned}&|\partial_t D_z\bar{u}(t,x,z).h| + |\partial_t D_x\bar{u}(t,x,z).h| \\ &\qquad \leqslant C_0\big(1+|||x|||_1^2 + |||\, z\,|||_1 + |||\,\mathcal{A}x\,|||_0 + |||\,\mathcal{A}z\,|||_0\,\big)\big(\,|||\,h\,|||_1 + |||\,\mathcal{A}h\,|||_0\,\big),\end{aligned} \tag{5.9}$$

where $C_0>0$ is a positive constant.

Proof. By using the same arguments as in [6, Section 7], we can prove that equation (5.7) admits a solution $\bar{u}\in C_b^{1,3,3}([0,T]\times\mathcal{H}\times\mathcal{H})$ which is given by (5.8), see also [8, Section 4]. Moreover, for $x,z,h\in\mathcal{H}^1$,

$$\partial_t D_z\bar{u}(t,x,z).h = D_z\partial_t\bar{u}(t,x,z).h = D_z(\bar{\mathcal{L}}_1+\bar{\mathcal{L}}_3)\bar{u}(t,x,z).h, \tag{5.10}$$

On the one hand, we have

$$\begin{aligned}&D_z\bar{\mathcal{L}}_1\bar{u}(t,x,z).h\\ &= D_zD_x\bar{u}(t,x,z).(\mathcal{A}x+\bar{\mathcal{F}}(x),h) + \frac{1}{2}\sum_{n=1}^{\infty}\beta_{1,n}D_zD_x^2\bar{u}(t,x,z).(Be_n,Be_n,h),\end{aligned}$$

which together with $\bar{u}\in C_b^{1,3,3}([0,T]\times\mathcal{H}\times\mathcal{H})$ yields that

$$|D_z\bar{\mathcal{L}}_1\bar{u}(t,x,z).h| \leqslant C_1(1+|||\mathcal{A}x\,|||_0 + |||\,x|||_1)\,|||\,h\,|||_1\,. \tag{5.11}$$

On the other hand, we have

$$\begin{aligned}&D_z\bar{\mathcal{L}}_3\bar{u}(t,x,z).h\\ &= \langle\mathcal{A}h, D_z\bar{u}(t,x,z)\rangle_{\mathcal{H}} + \langle D_x\bar{\mathcal{F}}(x).h, D_z\bar{u}(t,x,z))\rangle_{\mathcal{H}}\\ &+ D_z^2\bar{u}(t,x,z).(\mathcal{A}z+D_x\bar{\mathcal{F}}(x).z,h) + \frac{1}{2}\sum_{n=1}^{\infty}D_z^3\bar{u}(t,x,z).(\Sigma(x)e_n,\Sigma(x)e_n,h).\end{aligned}$$

Thus,

$$|D_z\bar{\mathcal{L}}_3\bar{u}(t,x,z).h| \leqslant C_2(1+|||\mathcal{A}z\,|||_0 + |||\,z\,|||_1 + |||\,x|||_1^2)(|||h\,|||_1 + |||\,\mathcal{A}h|||_0). \tag{5.12}$$

Combining (5.10), (5.11) and (5.12), we arrive at

$$|\partial_t D_z\bar{u}(t,x,z).h| \leqslant C_3\big(1+|||x\,|||_1^2 + |||\,z\,|||_1 + |||\,\mathcal{A}x\,|||_0 + |||\,\mathcal{A}z\,|||_0\big)\big(|||\,h\,|||_1 + |||\,\mathcal{A}h\,|||_0\big).$$

Similarly, we have

$$\begin{aligned}\partial_t D_x\bar{u}(t,x,z).h &= D_x^2\bar{u}(t,x,z).(\mathcal{A}x+\bar{\mathcal{F}}(x),h) + \langle\mathcal{A}h + D_x\bar{\mathcal{F}}(x).h, D_x\bar{u}(t,x,z)\rangle_{\mathcal{H}}\\ &\quad + \langle D_x^2\bar{\mathcal{F}}(x).(z,h), D_z\bar{u}(t,x,z))\rangle_{\mathcal{H}} + D_xD_z\bar{u}(t,x,z).(\mathcal{A}z+D_x\bar{\mathcal{F}}(x).z,h)\end{aligned}$$

$$+\frac{1}{2}\sum_{n=1}^{\infty}\beta_{1,n}D_x^3\bar{u}(t,x,z).(Be_n,Be_n,h)$$
$$+\frac{1}{2}\sum_{n=1}^{\infty}D_xD_z^2\bar{u}(t,x,z).(\Sigma(x)e_n,\Sigma(x)e_n,h)$$
$$+\sum_{n=1}^{\infty}D_z^2\bar{u}(t,x,z).(D_x(\Sigma(x))e_n,\Sigma(x)e_n,h).$$

By the same argument as above, we can obtain

$$|\partial_t D_x\bar{u}(t,x,z).h|\leqslant C_4\big(1+|||x|||_1^2+|||\, z\,|||_1+|||\,\mathcal{A}x\,|||_0+|||\,\mathcal{A}z\,|||_0\big)\big(|||\,h\,|||_1+|||\,\mathcal{A}h\,|||_0\big),$$

which completes the proof. □

5.2. **Proof of Theorem 2.3.** As before, we reduce the infinite dimensional problem to a finite dimensional one by the Galerkin approximation. Recall that $X_t^{n,\varepsilon}$ and $\bar{X}_t^n$ are defined by (4.1) and (4.2), respectively. Define

$$\mathcal{Z}_t^{n,\varepsilon}:=\frac{X_t^{n,\varepsilon}-\bar{X}_t^n}{\sqrt{\varepsilon}}.$$

Then we have

$$\mathrm{d}\mathcal{Z}_t^{n,\varepsilon}=\mathcal{A}\mathcal{Z}_t^{n,\varepsilon}\mathrm{d}t+\varepsilon^{-1/2}[\bar{\mathcal{F}}_n(X_t^{n,\varepsilon})-\bar{\mathcal{F}}_n(\bar{X}_t^n)]\mathrm{d}t+\varepsilon^{-1/2}\delta\mathcal{F}_n(X_t^{n,\varepsilon},Y_t^{n,\varepsilon})\mathrm{d}t,$$

where $\bar{\mathcal{F}}_n$ is given by (4.3), and $\delta\mathcal{F}_n(x,y):=\mathcal{F}_n(x,y)-\bar{\mathcal{F}}_n(x)$. Let $\bar{\mathcal{Z}}_t^n$ satisfy the following linear equation:

$$\mathrm{d}\bar{\mathcal{Z}}_t^n=\mathcal{A}\bar{\mathcal{Z}}_t^n\mathrm{d}t+D_x\bar{\mathcal{F}}_n(\bar{X}_t^n).\bar{\mathcal{Z}}_t^n\mathrm{d}t+P_n\Sigma(\bar{X}_t^n)\mathrm{d}W_t,$$

where W_t is a cylindrical Wiener process in H, and $\Sigma(x)$ is defined by (5.5). As in [36, Lemma 5.4], one can check that

$$\lim_{n\to\infty}\mathbb{E}\Big(|||\,\mathcal{Z}_t^\varepsilon-\mathcal{Z}_t^{n,\varepsilon}\,|||_1+|||\,\bar{\mathcal{Z}}_t-\bar{\mathcal{Z}}_t^n\,|||_1\Big)=0. \tag{5.13}$$

For any $T>0$ and $\varphi\in\mathbb{C}_b^3(\mathcal{H})$, we have for $t\in[0,T]$,

$$\begin{aligned}\big|\mathbb{E}[\varphi(\mathcal{Z}_t^\varepsilon)]-\mathbb{E}[\varphi(\bar{\mathcal{Z}}_t)]\big|&\leqslant|\mathbb{E}[\varphi(\mathcal{Z}_t^\varepsilon)]-\mathbb{E}[\varphi(\mathcal{Z}_t^{n,\varepsilon})]|\\&\quad+\big|\mathbb{E}[\varphi(\mathcal{Z}_t^{n,\varepsilon})]-\mathbb{E}[\varphi(\bar{\mathcal{Z}}_t^n)]\big|+\big|\mathbb{E}[\varphi(\bar{\mathcal{Z}}_t^n)]-\mathbb{E}[\varphi(\bar{\mathcal{Z}}_t)]\big|.\end{aligned} \tag{5.14}$$

According to (5.13), the first and the last terms on the right-hand of (5.14) converge to 0 as $n\to\infty$. Therefore, in order to prove Theorem 2.3, we only need to show that

$$\sup_{t\in[0,T]}\big|\mathbb{E}[\varphi(\mathcal{Z}_t^{n,\varepsilon})]-\mathbb{E}[\varphi(\bar{\mathcal{Z}}_t^n)]\big|\leqslant C_T\,\varepsilon^{\frac{1}{2}}, \tag{5.15}$$

where $C_T>0$ **is a constant independent of** n. We shall only work with the approximating system in the following subsection, and proceed to prove bounds that are uniform with respect to n. To simplify the notations, we shall omit the index n as before.

Define

$$\mathcal{L}_3\varphi(z) := \mathcal{L}_3(x,y,\bar{x},z)\varphi(z) := \langle \mathcal{A}z, D_z\varphi(z)\rangle_{\mathcal{H}} + \frac{1}{\sqrt{\varepsilon}}\langle \bar{\mathcal{F}}(x)-\bar{\mathcal{F}}(\bar{x}), D_z\varphi(z)\rangle_{\mathcal{H}} + \frac{1}{\sqrt{\varepsilon}}\langle \delta\mathcal{F}(x,y), D_z\varphi(z)\rangle_{\mathcal{H}}, \quad \forall \varphi \in C_\ell^1(\mathcal{H}). \tag{5.16}$$

Given a function $\phi \in C_\ell^{1,2,\eta,2}([0,T]\times\mathcal{H}\times H\times\mathcal{H})$ satisfying the centering condition:

$$\int_H \phi(t,x,y,z)\mu^x(\mathrm{d}y) = 0, \quad \forall t>0, x,z\in\mathcal{H}, \tag{5.17}$$

let $\psi(t,x,y,z)$ solve the following Poisson equation

$$\mathcal{L}_2(x,y)\psi(t,x,y,z) = -\phi(t,x,y,z), \tag{5.18}$$

where t,x,z are regarded as parameters. Define

$$\overline{\delta\mathcal{F}\cdot\nabla_z\psi}(t,x,z) := \int_H \nabla_z\psi(t,x,y,z).\delta\mathcal{F}(x,y)\mu^x(\mathrm{d}y). \tag{5.19}$$

We first establish the following weak fluctuation estimates for an appropriate integral functional of $(X_s^\varepsilon, Y_s^\varepsilon, \mathcal{Z}_s^\varepsilon)$ over the time interval $[0,t]$, which will play an important role in the proof of (5.15).

Lemma 5.2 (Weak fluctuation estimates)**.** *Let $T,\eta>0$, $x\in\mathcal{H}^1$ and $y\in H$. Assume that $\mathcal{F}\in C_b^{2,\eta}(\mathcal{H}\times H,\mathcal{H}^1)$ and $\mathcal{G}\in C_B^{2,\eta}(\mathcal{H}\times H,H)$. Then for any $t\in[0,T]$, $\phi\in C_\ell^{1,2,\eta,2}([0,T]\times\mathcal{H}\times H\times\mathcal{H})$ satisfying (5.17) and*

$$|\partial_t\phi(t,x,y,z)| \leqslant C_0\big(1+|||x|||_1^2+|||z|||_1 + |||\mathcal{A}x|||_0+|||\mathcal{A}z|||_0\big)\big(1+|||x|||_1+\|y\|\big), \tag{5.20}$$

we have

$$\mathbb{E}\left(\int_0^t \phi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right) \leqslant C_T\,\varepsilon^{\frac{1}{2}}, \tag{5.21}$$

and

$$\mathbb{E}\left(\frac{1}{\sqrt{\varepsilon}}\int_0^t \phi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right) - \mathbb{E}\left(\int_0^t \overline{\delta\mathcal{F}\cdot\nabla_z\psi}(s,X_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right) \leqslant C_T\,\varepsilon^{\frac{1}{2}}, \tag{5.22}$$

where $C_T>0$ is a constant independent of ε,η and n.

Proof. The proof will be divided into two steps.

Step 1. We first prove estimate (5.21). Applying Itô's formula to $\psi(t,X_t^\varepsilon,Y_t^\varepsilon,\mathcal{Z}_t^\varepsilon)$ and taking expectation, we have

$$\mathbb{E}[\psi(t,X_t^\varepsilon,Y_t^\varepsilon,\mathcal{Z}_t^\varepsilon)] = \psi(0,x,y,0) + \mathbb{E}\left(\int_0^t(\partial_s+\mathcal{L}_1+\mathcal{L}_3)\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right)$$

$$+\frac{1}{\varepsilon}\mathbb{E}\left(\int_0^t \mathcal{L}_2\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right),$$

where $\mathcal{L}_1$ and $\mathcal{L}_3$ are defined by (4.6) and (5.16), respectively. Combining this with (5.18), we obtain

$$\begin{aligned}\mathbb{E}\left(\int_0^t \phi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right) &= \varepsilon\mathbb{E}\big[\psi(0,x,y,0)-\psi(t,X_t^\varepsilon,Y_t^\varepsilon,\mathcal{Z}_t^\varepsilon)\big]\\ &\quad+\varepsilon\mathbb{E}\left(\int_0^t \partial_s\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right)+\varepsilon\mathbb{E}\left(\int_0^t \mathcal{L}_1\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right)\\ &\quad+\varepsilon\mathbb{E}\left(\int_0^t \mathcal{L}_3\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right)=:\sum_{i=1}^4\mathscr{Q}_i(t,\varepsilon). \qquad (5.23)\end{aligned}$$

By Theorem 3.1 and Lemma 3.2, we have

$$\mathscr{Q}_1(t,\varepsilon)\leqslant C_1\varepsilon\mathbb{E}\big(1+|||X_t^\varepsilon|||_1+\|Y_t^\varepsilon\|\big)\leqslant C_1\varepsilon.$$

For the second term, by using Theorem 3.1, condition (5.20), Lemma 3.2, (5.2) and (5.3), we get

$$\begin{aligned}\mathscr{Q}_2(t,\varepsilon)&\leqslant C_2\int_0^t\mathbb{E}\big(1+|||\mathcal{A}X_s^\varepsilon|||_0^2+|||\mathcal{A}\mathcal{Z}_s^\varepsilon|||_0^2+|||X_s^\varepsilon|||_1^4++\|Y_s^\varepsilon\|^2+|||\mathcal{Z}_s^\varepsilon|||_1^2\big)\mathrm{d}s\\ &\leqslant C_2\,\varepsilon.\end{aligned}$$

To treat the third term, since for each $t\in[0,T]$, $\phi(t,\cdot,\cdot,\cdot)\in C_\ell^{2,\eta,2}(\mathcal{H}\times H\times\mathcal{H})$, by Theorem 3.1, we have $\psi(t,\cdot,\cdot,\cdot)\in C_\ell^{2,2,2}(\mathcal{H}\times H\times\mathcal{H})$, hence

$$\begin{aligned}|\mathcal{L}_1\psi(t,X_t^\varepsilon,Y_t^\varepsilon,\mathcal{Z}_t^\varepsilon)|&\leqslant|\langle\mathcal{A}X_t^\varepsilon+\mathcal{F}(X_t^\varepsilon,Y_t^\varepsilon),D_x\psi(t,X_t^\varepsilon,Y_t^\varepsilon,\mathcal{Z}_t^\varepsilon)\rangle_{\mathcal{H}}|\\ &\quad+\frac{1}{2}Tr((BQ_1^{\frac{1}{2}})(BQ_1^{\frac{1}{2}})^*)\|D_x^2\psi(t,X_t^\varepsilon,Y_t^\varepsilon,\mathcal{Z}_t^\varepsilon)\|_{\mathscr{L}(\mathcal{H}\times\mathcal{H})}\\ &\leqslant C_3\big(1+|||\mathcal{A}X_t^\varepsilon|||_0^2+|||X_t^\varepsilon|||_1^2+\|Y_t^\varepsilon\|^2\big).\end{aligned}$$

As a result of Lemmas 3.2 and 3.3, we deduce that

$$\mathscr{Q}_3(t,\varepsilon)\leqslant C_3\,\varepsilon\mathbb{E}\left(\int_0^t\Big(|||\mathcal{A}X_s^\varepsilon|||_0^2+|||X_s^\varepsilon|||_1^2+\|Y_s^\varepsilon\|^2\Big)\mathrm{d}s\right)\leqslant C_3\,\varepsilon.$$

For the last term, we have

$$\begin{aligned}\mathscr{Q}_4(t,\varepsilon)&=\varepsilon\mathbb{E}\left(\int_0^t\langle\mathcal{A}\mathcal{Z}_s^\varepsilon,D_z\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\rangle_{\mathcal{H}}\mathrm{d}s\right)\\ &\quad+\sqrt{\varepsilon}\mathbb{E}\left(\int_0^t\langle\bar{\mathcal{F}}(X_s^\varepsilon)-\bar{\mathcal{F}}(\bar{X}_s),D_z\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\rangle_{\mathcal{H}}\mathrm{d}s\right)\\ &\quad+\sqrt{\varepsilon}\mathbb{E}\left(\int_0^t\langle\delta\mathcal{F}(X_s^\varepsilon,Y_s^\varepsilon),D_z\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\rangle_{\mathcal{H}}\mathrm{d}s\right).\end{aligned}$$

It follows from (5.3) and Lemma 3.2 again that

$$\mathscr{Q}_4(t,\varepsilon) \leqslant C_4\sqrt{\varepsilon}\mathbb{E}\left(\int_0^t (1+|||\mathcal{A}\mathcal{Z}_s^\varepsilon|||_0^2 + |||\, X_s^\varepsilon\,|||_1^2 + \|Y_s^\varepsilon\|^2)\mathrm{d}s\right) \leqslant C_4\sqrt{\varepsilon}.$$

Combining the above inequalities with (5.23), we get the desired result.

Step 2. We proceed to prove estimate (5.22). By following exactly the same arguments as in the proof of Step 1, we get that

$$\mathbb{E}\left(\frac{1}{\sqrt{\varepsilon}}\int_0^t \phi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right) \leqslant C_0\sqrt{\varepsilon} + \sqrt{\varepsilon}\mathbb{E}\left(\int_0^t \mathcal{L}_3\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right).$$

For the last term, by definition (5.16) we have

$$\begin{aligned}
&\sqrt{\varepsilon}\mathbb{E}\left(\int_0^t \mathcal{L}_3\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right)\\
&= \sqrt{\varepsilon}\mathbb{E}\left(\int_0^t \langle \mathcal{A}\mathcal{Z}_s^\varepsilon, D_z\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\rangle_{\mathcal{H}}\mathrm{d}s\right)\\
&+\mathbb{E}\left(\int_0^t \langle \bar{\mathcal{F}}(X_s^\varepsilon)-\bar{\mathcal{F}}(\bar{X}_s), D_z\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\rangle_{\mathcal{H}}\mathrm{d}s\right)\\
&+\mathbb{E}\left(\int_0^t \langle \delta\mathcal{F}(X_s^\varepsilon,Y_s^\varepsilon), D_z\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\rangle_{\mathcal{H}}\mathrm{d}s\right) =: \sum_{i=1}^3 \mathscr{T}_i(t,\varepsilon).
\end{aligned}$$

Using Lemma 3.2 and (5.3), we get

$$\mathscr{T}_1(t,\varepsilon) \leqslant C_1\sqrt{\varepsilon}\mathbb{E}\left(\int_0^t |||\mathcal{A}\mathcal{Z}_s^\varepsilon\,|||_0\,(1+|||X_s^\varepsilon\,|||_1 + \|Y_s^\varepsilon\|)\mathrm{d}s\right) \leqslant C_1\sqrt{\varepsilon}.$$

According to Hölder's inequality, Lemma 3.2 and Theorem 2.1, we have

$$\mathscr{T}_2(t,\varepsilon) \leqslant C_2\int_0^t \left(\mathbb{E}\,|||\,X_s^\varepsilon-\bar{X}_s\,|||_1^2\right)^{1/2}\left(1+\mathbb{E}|||X_s^\varepsilon|||_1^2+\mathbb{E}\|Y_s^\varepsilon\|^2\right)^{1/2}\mathrm{d}s \leqslant C_2\sqrt{\varepsilon}.$$

Thus, we deduce that

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{\sqrt{\varepsilon}}\int_0^t \phi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right) - \mathbb{E}\left(\int_0^t \overline{\delta\mathcal{F}\cdot\nabla_z\psi}(s,X_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\mathrm{d}s\right)\\
&\leqslant C_3\sqrt{\varepsilon} + \mathbb{E}\left(\int_0^t \left(\langle \delta\mathcal{F}(X_s^\varepsilon,Y_s^\varepsilon), D_z\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\rangle_{\mathcal{H}} - \overline{\delta\mathcal{F}\cdot\nabla_z\psi}(s,X_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\right)\mathrm{d}s\right),
\end{aligned}$$

where $\overline{\delta\mathcal{F}\cdot\nabla_z\psi}$ is defined by (5.19). Note that the function

$$\tilde{\phi}(t,x,y,z) := \langle \delta\mathcal{F}(x,y), D_z\psi(t,x,y,z)\rangle_1 - \overline{\delta\mathcal{F}\cdot\nabla_z\psi}(t,x,z)$$

satisfies the centering condition (5.17) and condition (5.20). Thus, using (5.21) directly, we obtain

$$\mathbb{E}\left(\int_0^t \left(\langle\delta\mathcal{F}(X_s^\varepsilon,Y_s^\varepsilon), D_z\psi(s,X_s^\varepsilon,Y_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\rangle_{\mathcal{H}} - \overline{\delta\mathcal{F}\cdot\nabla_z\psi}(s,X_s^\varepsilon,\mathcal{Z}_s^\varepsilon)\right)\mathrm{d}s\right) \leqslant C_4\sqrt{\varepsilon},$$

which completes the proof. □

Now, we are in the position to give:

Proof of estimate (5.15). Fix $T>0$. For any $t\in[0,T]$ and $x,z\in\mathcal{H}^1$, let

$$\tilde{u}(t,x,z)=\bar{u}(T-t,x,z).$$

It is easy to check that

$$\tilde{u}(0,x,0)=\bar{u}(T,x,0)=\mathbb{E}[\varphi(\bar{\mathcal{Z}}_T)] \quad\text{and}\quad \tilde{u}(T,x,z)=\bar{u}(0,x,z)=\varphi(z).$$

Applying Itô's formula, by (5.7) we have

$$\begin{aligned}
&\mathbb{E}[\varphi(\mathcal{Z}_T^\varepsilon)]-\mathbb{E}[\varphi(\bar{\mathcal{Z}}_T)]=\mathbb{E}[\tilde{u}(T,\bar{X}_T,\mathcal{Z}_T^\varepsilon)-\tilde{u}(0,x,0)]\\
&=\mathbb{E}\left(\int_0^T \big(\partial_t+\mathcal{L}_1+\mathcal{L}_3\big)\tilde{u}(t,X_t^\varepsilon,\mathcal{Z}_t^\varepsilon)\mathrm{d}t\right)\\
&=\mathbb{E}\left(\int_0^T (\mathcal{L}_1-\bar{\mathcal{L}}_1)\tilde{u}(t,X_t^\varepsilon,\mathcal{Z}_t^\varepsilon)\mathrm{d}t\right)+\mathbb{E}\left(\int_0^T (\mathcal{L}_3-\bar{\mathcal{L}}_3)\tilde{u}(t,X_t^\varepsilon,\mathcal{Z}_t^\varepsilon)\mathrm{d}t\right)\\
&=\mathbb{E}\left(\int_0^T \langle\mathcal{F}(X_t^\varepsilon,Y_t^\varepsilon)-\bar{\mathcal{F}}(X_t^\varepsilon), D_x\tilde{u}(t,X_t^\varepsilon,\mathcal{Z}_t^\varepsilon))\rangle_{\mathcal{H}}\mathrm{d}t\right)\\
&\quad+\mathbb{E}\left(\int_0^T \left\langle\frac{\bar{\mathcal{F}}(X_t^\varepsilon)-\bar{\mathcal{F}}(\bar{X}_t)}{\sqrt{\varepsilon}}-D_x\bar{\mathcal{F}}(X_t^\varepsilon).\mathcal{Z}_t^\varepsilon, D_z\tilde{u}(t,X_t^\varepsilon,\mathcal{Z}_t^\varepsilon))\right\rangle_{\mathcal{H}}\mathrm{d}t\right)\\
&\quad+\left[\mathbb{E}\left(\frac{1}{\sqrt{\varepsilon}}\int_0^T \langle\mathcal{F}(X_t^\varepsilon,Y_t^\varepsilon)-\bar{\mathcal{F}}(X_t^\varepsilon), D_z\tilde{u}(t,X_t^\varepsilon,\mathcal{Z}_t^\varepsilon))\rangle_{\mathcal{H}}\mathrm{d}t\right)\right.\\
&\qquad\left.-\frac{1}{2}\mathbb{E}\left(\int_0^T Tr(D_z^2\tilde{u}(t,X_t^\varepsilon,\mathcal{Z}_t^\varepsilon)\Sigma(X_t^\varepsilon)\Sigma(X_t^\varepsilon)^*)\mathrm{d}t\right)\right]:=\sum_{i=1}^3\mathscr{N}_i(T,\varepsilon).
\end{aligned}$$

For the first term, recall that $\tilde{\Psi}$ solves the Poisson equation (5.6) and define

$$\psi(t,x,y,z):=\langle\tilde{\Psi}(x,y),D_x\tilde{u}(t,x,z)\rangle_{\mathcal{H}}.$$

Since $\mathcal{L}_2$ is an operator with respect to the y-variable, one can check that ψ solves the following Poisson equation:

$$\mathcal{L}_2(x,y)\psi(t,x,y,z)=-\langle\delta\mathcal{F}(x,y),D_x\tilde{u}(t,x,z)\rangle_{\mathcal{H}}=:-\phi(t,x,y,z).$$

It is obvious that ϕ satisfies the centering condition (5.17). Furthermore, by (5.9) we get

$$|\partial_t\phi(t,x,y,z)|=|\langle\delta\mathcal{F}(x,y),\partial_t D_x\bar{u}(T-t,x,z)\rangle_{\mathcal{H}}|$$

$$\leqslant C_1\big(1+|||x|||_1^2+|||\, z\,|||_1+|||\,\mathcal{A}x\,|||_0+|||\,\mathcal{A}z\,|||_0\big)\big(\,|||\,\delta\mathcal{F}(x,y)\,|||_1+|||\,\mathcal{A}\delta\mathcal{F}(x,y)\,|||_0\,\big)$$
$$\leqslant C_1\big(1+|||x|||_1^2+|||\, z\,|||_1+|||\,\mathcal{A}x\,|||_0+|||\,\mathcal{A}z\,|||_0\big)\big(1+|||x\,|||_1+\|y\|\big).$$

Thus, it follows from (5.21) directly that

$$\mathcal{N}_1(T,\varepsilon)\leqslant C_1\sqrt{\varepsilon}.$$

To control the second term, by the mean value theorem, Hölder's inequality, Lemma 5.1, Theorem 2.1 and (5.2) we deduce that for $\vartheta\in(0,1)$,

$$\mathcal{N}_2(T,\varepsilon)\leqslant\mathbb{E}\left(\int_0^T\big|\big\langle[D_x\bar{\mathcal{F}}(X_t^\varepsilon+\vartheta(X_t^\varepsilon-\bar{X}_t))-D_x\bar{\mathcal{F}}(X_t^\varepsilon)].\mathcal{Z}_t^\varepsilon,D_z\tilde{u}(t,X_t^\varepsilon,\mathcal{Z}_t^\varepsilon)\big\rangle_{\mathcal{H}}\big|\,\mathrm{d}t\right)$$
$$\leqslant C_2\int_0^T\big(\mathbb{E}\,|||\,X_t^\varepsilon-\bar{X}_t\,|||_1^2\,\big)^{1/2}\big(\mathbb{E}\,|||\,\mathcal{Z}_t^\varepsilon\,|||_1^2\,\big)^{1/2}\mathrm{d}t\leqslant C_2\sqrt{\varepsilon}.$$

For the last term, define

$$\hat{\psi}(t,x,y,z):=\langle\tilde{\Psi}(x,y),D_z\tilde{u}(t,x,z)\rangle_{\mathcal{H}}.$$

Then $\hat{\psi}$ solves the Poisson equation

$$\mathcal{L}_2(x,y)\hat{\psi}(t,x,y,z)=-\langle\delta\mathcal{F}(x,y),D_z\tilde{u}(t,x,z)\rangle_{\mathcal{H}}=:-\hat{\phi}(t,x,y,z).$$

By exactly the same arguments as above, we have that $\hat{\phi}$ satisfies the centering condition (5.17) and condition (5.20). Furthermore, by the definition of Σ in (5.5), we have

$$\overline{\delta\mathcal{F}\cdot\nabla_z\hat{\psi}}(t,x,z)=\int_H D_z\hat{\psi}(t,x,y,z).\delta\mathcal{F}(x,y)\mu^x(\mathrm{d}y)$$
$$=\int_H D_z^2\tilde{u}(t,x,z).(\tilde{\Psi}(x,y),\delta\mathcal{F}(x,y))\mu^x(\mathrm{d}y)=\frac{1}{2}Tr(D_z^2\tilde{u}(t,x,z)\Sigma(x)\Sigma^*(x)).$$

Thus, it follows by (5.22) directly that

$$\mathcal{N}_3(T,\varepsilon)\leqslant C_3\,\sqrt{\varepsilon}.$$

Combining the above computations, we get the desired result. □

References

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

[2] Bao J., Yin G. and Yuan C.: Two-time-scale stochastic partial differential equations driven by α-stable noises: Averaging principles. *Bernoulli*, **23** (2017), 645–669.

[3] Barbu V., Da Prato G. and Tubaro L.: Stochastic wave equations with dissipative damping. *Stoch. Proc. Appl.*, **117** (2007), 1001–1013.

[4] Bogoliubov N. N. and Mitropolsky Y. A.: Asymptotic methods in the theory of non-linear oscillations. Gordon and Breach Science Publishers, New York, 1961.

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

[6] Bréhier C. E.: Orders of convergence in the averaging principle for SPDEs: the case of a stochastically forced slow component. *Stoch. Proc. Appl.*, **130** (2020), 3325–3368.
[7] Bréhier C. E.: Analysis of an HMM time-discretization scheme for a system of stochastic PDEs. *SIAM J. Numer. Anal.*, **51** (2013), 1185–1210.
[8] Bréhier C. E. and Debussche A.: Kolmogorov equations and weak order analysis for SPDEs with nonlinear diffusion coefficient. *J. Math. Pures Appl.*, **119** (2018), 193–254.
[9] Cerrai S.: A Khasminskii type averaging principle for stochastic reaction-diffusion equations. *Ann. Appl. Probab.*, **19** (2009), 899–948.
[10] Cerrai S.: Normal deviations from the averaged motion for some reaction-diffusion equations with fast oscillating perturbation. *J. Math. Pures Appl.*, **91** (2009), 614–647.
[11] Cerrai S.: Averaging principle for systems of reaction-diffusion equations with polynomial nonlinearities perturbed by multiplicative noise. *SIAM J. Math. Anal*, **43** (2011), 2482–2518.
[12] Cerrai S. and Freidlin M.: Averaging principle for stochastic reaction-diffusion equations. *Probab. Theory Related Fields*, **144** (2009), 137–177.
[13] Cerrai S. and Glatt-Holtz N.: On the convergence of stationary solutions in the Smoluchowski-Kramers approximation of infinite dimensional systems. *J. Funct. Anal.*, **278** (2020), 108421.
[14] Cerrai S. and Lunardi A.: Averaging principle for non-autonomous slow-fast systems of stochastic reaction-diffusion equations: the almost periodic case. *SIAM J. Math. Anal.*, **49** (2017), 2843–2884.
[15] Chevyrev I., Friz P., Korepanov A. and Melbourne I.: Superdiffusive limits for deterministic fast-slow dynamical systems. arXiv:1907.04825v1.
[16] Chow P. L.: Thermoelastic wave propagation in a random medium and some related problems. *Inter. J. Engineering Sci.*, **11** (1973), 953–971.
[17] Chow P. L.: Asymptotics of solutions to semilinear stochastic wave equations. *Anna. Appl. Probab.*, **16** (2006), 757–789.
[18] Chow P. L.: Stochastic Partial Differential Equations. CRC Press, 2014.
[19] Da Prato G. and Flandoli F.: Pathwise uniqueness for a class of SDEs in Hilbert spaces and applications. *J. Funct. Anal.*, **259** (2010), 243–267.
[20] E W., Liu D. and Vanden-Eijnden E.: Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.*, **58** (2005), 1544–1585.
[21] Fu H., Wan L., Liu J. and Liu X.: Weak order in averaging principle for stochastic wave equation with a fast oscillation. *Stoch. Process. Appl.*, **128** (2018), 2557–2580.
[22] Fu H., Wan L. and Liu J.: Strong convergence in averaging principle for stochastic hyperbolic-parabolic equations with two time-scales. *Stoch. Process. Appl.*, **125** (2015), 3255–3279.
[23] Gonzales-Gargate I. I. and Ruffino P. R.: An averaging principle for diffusions in foliated spaces. *Ann. Probab.*, **44** (2016) 567–588.
[24] Hairer M. and Li X. M.: Averaging dynamics driven by fractional Brownian motion. https://arxiv.org/pdf/1902.11251.pdf.
[25] Hairer M. and Pardoux E.: Homogenization of periodic linear degenerate PDEs. *J. Func. Anal.*, **255** (2008), 2462–2487.
[26] Hairer M. and Pardoux E.: Fluctuations around a homogenised semilinear random PDE. https://arxiv.org/abs/1911.02865.
[27] Kelly D. and Melbourne I.: Homogenization for deterministic fast-slow systems with multidimensional multiplicative noise. *J. Funct. Anal.*, **272** (2017), 4063–4102.
[28] Khasminskii R. Z.: On stochastic processes defined by differential equations with a small parameter. *Theory Probab. Appl.*, **11** (1966), 211–228.
[29] Khasminskii R. Z. and Yin G.: On averaging principles: An asymptotic expansion approach. *SIAM J. Math. Anal.*, **35** (2004), 1534–1560.

[30] Leung A. W.: Asymptotically stable invariant manifold for coupled nonlinear parabolic-hyperbolic partial differential equations. *J. Diff. Equ.*, **187** (2003), 184–200.
[31] Liu W. and Röckner M.: Stochastic partial differential equations: an introduction. Universitext, Springer, 2015.
[32] Pardoux E. and Veretennikov A.Yu.: On the Poisson equation and diffusion approximation. I. *Ann. Probab.*, **29** (2001), 1061–1085.
[33] Pardoux E. and Veretennikov A. Yu.: On the Poisson equation and diffusion approximation 2. *Ann. Probab.*, **31** (2003), 1166–1192.
[34] Rivera J. E. M. and Racke R.: Smoothing properties, decay and global existence of solution to nonlinear coupled systems of thermoelasticity type. *SIAM J. Math. Anal.*, **26** (1995) 1547–1563.
[35] Röckner M. and Xie L.: Diffusion approximation for fully coupled stochastic differential equations. Accepted by *Ann. Prob.*, https://bibos.math.uni-bielefeld.de/preprints/19-09-560.pdf.
[36] Röckner M., Xie L. and Yang L.: Asymptotic behavior of multiscale stochastic partial differential equations. https://bibos.math.uni-bielefeld.de//preprints/20-09-573.pdf.
[37] Veretennikov A.Yu.: On the averaging principle for systems of stochastic differential equations, *Math. USSR Sborn.* **69** (1991) 271–284.
[38] Wang W. and Roberts A.J.: Average and deviation for slow-fast stochastic partial differential equations. *J. Differential Equ.*, **253** (2012), 1265–1286.
[39] Zhang X. and Zuazua E.: Long-time behavior of a coupled heat-wave system arising in fluid-structure interaction. *Arch. Ration. Mech. Anal.*, **184** (2007), 49–120.

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

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

Li Yang: Department of Mathematics, Shandong University, Jinan, Shandong 250100, P.R.China, Email: llyang@mail.sdu.edu.cn