Averaging principle and normal deviations for multi-scale stochastic hyperbolic–parabolic equations

Michael Röckner\textsuperscript{1,2} · Longjie Xie\textsuperscript{3} · Li Yang\textsuperscript{4}

Abstract
We study the asymptotic behavior of stochastic hyperbolic–parabolic equations with slow–fast time scales. Both the strong and weak convergence in the averaging principle are established. 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. An extra diffusion term appears in the limit which is given explicitly in terms of the solution of a Poisson equation. Furthermore, sharp rates for the above convergence are obtained, which are shown not to depend on the regularity of the coefficients in the equation for the fast variable.

Keywords Stochastic hyperbolic–parabolic equations · Averaging principle · Strong and weak convergence · Homogenization

Mathematics Subject Classification 60H15 · 60F05 · 70K70
1 Introduction

Let $T > 0$ and $D = (0, L) \subseteq \mathbb{R}$ be a bounded open interval. Consider the following system of stochastic hyperbolic–parabolic equations:

\[
\begin{aligned}
\frac{\partial^2 U^\varepsilon_t(\xi)}{\partial t^2} &= \Delta U^\varepsilon_t(\xi) + f(U^\varepsilon_t(\xi), Y^\varepsilon_t(\xi)) + \dot{W}^1_t(\xi), \\
\frac{\partial Y^\varepsilon_t(\xi)}{\partial t} &= \frac{1}{\varepsilon} \Delta Y^\varepsilon_t(\xi) + \frac{1}{\varepsilon} g(U^\varepsilon_t(\xi), Y^\varepsilon_t(\xi)) + \frac{1}{\sqrt{\varepsilon}} \dot{W}^2_t(\xi), \\
U^\varepsilon_0(\xi) &= Y^\varepsilon_0(\xi) = 0
\end{aligned}
\]

where $\Delta$ is the Laplacian operator, $f, g : \mathbb{R}^2 \to \mathbb{R}$ are measurable functions, $W^1_t$ and $W^2_t$ are two mutually independent $Q_1$- and $Q_2$-Wiener processes both defined on a complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$, and the small parameter $0 < \varepsilon \ll 1$ represents the separation of time scales between the ‘slow’ process $U^\varepsilon_t$ and the ‘fast’ motion $Y^\varepsilon_t$ (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, see e.g. [3, 4, 22]. 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), see e.g. [18, 36, 43, 49] and the references therein. In this respect, the question that how a thermal environment at large time scales may influence the dynamics of the whole system arises.

In the mathematical literature, powerful averaging and homogenization methods have been developed to study the asymptotic behavior of multi-scale systems as $\varepsilon \to 0$. The averaging principle can be viewed as a functional law of large numbers, which says 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 anymore and thus is much simpler. This theory was first studied by Bogoliubov [5] for deterministic ordinary differential equations, and extended to stochastic differential equations (SDEs for short) by Khasminskii [33], see also [1, 28, 29, 34, 47] and the references therein. As a rule, the averaging method requires certain smoothness on both the original and the averaged coefficients of the systems. Various assumptions have been studied in order to guarantee the above convergence. Recently, the averaging principle for two time scale stochastic partial differential equations (SPDEs for short) has attracted considerable attention. In [13], Cerrai and Freidlin proved the averaging principle for slow–fast stochastic reaction–diffusion equations with noise only in the fast motion. Later, Cerrai [10, 12] generalized this result to more general reaction–diffusion equations, see also [2, 15, 39, 46, 48] and the references therein for further developments. We also mention that Bréhier [6, 7] 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. [8, 24].

Concerning the stochastic hyperbolic–parabolic equation (1), Fu et al. [26] established the strong convergence in the averaging principle for system (1) when \( d = 1 \) by the classical Khasminskii time discretization method, and obtained the \( 1/4 \)-order rate of strong convergence. In [25], by using asymptotic expansion arguments, the authors studied the weak order convergence for system (1), but only in not fully coupled case, where \( g(u, y) = g(y) \), 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 system (1) with singular coefficients, see Theorem 1. Compared with [25, 26], we allow the system to be fully coupled. Moreover, we assume that the coefficients are only \( \eta \)-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. In addition, 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) 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 [6, 10, 12, 13, 25, 26]. 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) 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 \( \varepsilon \), the slow variable \( U_\varepsilon t \) 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

\[
Z_\varepsilon t := \frac{U_\varepsilon t - \bar{U}_t}{\sqrt{\varepsilon}} 
\]

as \( \varepsilon \) 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. [23, 30, 31]. 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 [33], see also [1, 16, 27, 32, 41, 42, 44]. The infinite dimensional situation is more open and papers on this subject are very few. In [11], 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 [48]. In both papers, the methods of proof are based on Khasminskii’s time discretization argument. Recently, we [46] 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 [46] to study the functional central limit theorem for the stochastic hyperbolic–parabolic system (1) with Hölder continuous coefficients. More precisely, we show that the normalized difference \( Z_{\varepsilon} t \), defined by (2), converges weakly as \( \varepsilon \to 0 \) to the solution of a linear stochastic wave equation, see Theorem 2. An additional diffusion term appears in the limit which is given explicitly in terms of the solution of a Poisson equation. Furthermore, 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 new for this kind of systems.

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 [46]. To be more precise, due to the hyperbolic structure, the corresponding semigroup generated by the linear operator is not analytic, which makes the method based on the regularity of the analytic semigroup to derive moment estimates in [46] no longer applicable. On the other hand, to prove both the weak convergence in the averaging principle and the normal deviation, we need to study the regularity of the solutions to the Kolmogorov equations associated with the limit stochastic wave equations, which is different from [46]. The arguments here mainly rely on the structure of the hyperbolic equation and are not applicable to coupled system of parabolic equations, see Lemma 6. Furthermore, we derive the higher order spatial-temporal convergence in the averaging principle and in the functional central limit theorem. These involve more complicated analysis although it is natural for the solutions of wave equations which possess high order regularity with respect to time and space.

The rest of this paper is organized as follows. In Sect. 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 1, and the normal deviation result, Theorem 2, in Sects. 4 and 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 \( \mathcal{L}(H_1, H_2) \) to denote the space of all linear and bounded operators from \( H_1 \) to \( H_2 \). If \( H_1 = H_2 \), we write \( \mathcal{L}(H_1) = \mathcal{L}(H_1, H_1) \) for simplicity. Recall that an operator \( Q \in \mathcal{L}(\hat{H}) \) is called Hilbert-Schmidt if

\[
\|Q\|_{\mathcal{L}_2(\hat{H})}^2 := Tr(Q Q^*) < +\infty.
\]

We shall denote the space of all Hilbert-Schmidt operators on \( \hat{H} \) by \( \mathcal{L}_2(\hat{H}) \). For \( k, m \in \mathbb{N} \) and \( \eta \in (0, 1) \), we also introduce the following spaces of functions.

- \( L^\infty_k(H_1 \times H_2, \hat{H}) \) denote the space of all measurable maps \( \phi : H_1 \times H_2 \to \hat{H} \) with at most linear growth, i.e.,

\[
\|\phi\|_{L^\infty_k(\hat{H})} := \sup_{(x, y) \in H_1 \times H_2} \frac{\|\phi(x, y)\|_{\hat{H}}}{\|x\|_{H_1} + \|y\|_{H_2}} < \infty.
\]
Let \( H \) be the usual space of square integrable functions on a bounded interval \( D = (0, L) \) in \( \mathbb{R} \) 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

\[
Ae_n = -\lambda_n e_n, \tag{4}
\]

\[ - C^{k,0}_\ell(H_1 \times H_2, \hat{H}) \text{ is the space of all } \phi \in L^\infty_{\ell}(H_1 \times H_2, \hat{H}) \text{ such that } \phi \text{ has } k \text{ times Gâteaux derivatives with respect to the } x\text{-variable satisfying}
\]

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

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

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

\[ - C^{k,m}_\ell(H_1 \times H_2, \hat{H}) \text{ is the space of all maps satisfying}
\]

\[
\|\phi\|_{C^{k,m}_\ell(H)} := \|\phi\|_{L^\infty(\hat{H})} + \|\phi\|_{C^{k,0}_\ell(H)} + \|\phi\|_{C^{0,m}_\ell(H)} < \infty, \tag{3}
\]

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

\[
\|\phi(x, y_1) - \phi(x, y_2)\|_{\hat{H}} \leq C_0 \|y_1 - y_2\|_{H_2}^\eta.
\]

When \( \hat{H} = \mathbb{R} \), we will omit the letter \( \hat{H} \) in all the above notations for simplicity.

### 2 Assumptions and main results

Let \( H := L^2(D) \) be the usual space of square integrable functions on a bounded interval \( D = (0, L) \) in \( \mathbb{R} \) with scalar product and norm denoted by \( \langle \cdot, \cdot \rangle \) and \( \| \cdot \| \), respectively.
with \(0 < \lambda_1 \leq \lambda_2 \leq \cdots \lambda_n \leq \cdots\). For \(\alpha \in \mathbb{R}\), let \(H^\alpha := D((-A)^{\alpha/2})\) be the Hilbert space endowed with the scalar product
\[
\langle x, y \rangle_\alpha := \langle (-A)^{\alpha/2}x, (-A)^{\alpha/2}y \rangle = \sum_{n=1}^{\infty} \lambda_n^{\alpha} \langle x, e_n \rangle \langle y, e_n \rangle, \quad \forall x, y \in H^\alpha,
\]
and norm
\[
\|x\|_\alpha := \left( \sum_{n=1}^{\infty} \lambda_n^{\alpha} \langle x, e_n \rangle^2 \right)^{1/2}, \quad \forall x \in H^\alpha.
\]
Then \(A\) can be regarded as an operator from \(H^\alpha\) to \(H^{\alpha-2}\). For the drift coefficients \(f\) and \(g\) given in system (1), we introduce two Nemytskii operators \(F, G : H \times H \to H\) by
\[
F(u, y)(\xi) := f(u(\xi), y(\xi)), \quad G(u, y)(\xi) := g(u(\xi), y(\xi)), \quad \xi \in D.
\]
We assume that for some \(\eta > 0\),
\[
f \in C_b^{2, \eta}(\mathbb{R} \times \mathbb{R}, \mathbb{R}) \quad \text{and} \quad g \in C_B^{2, \eta}(\mathbb{R} \times \mathbb{R}, \mathbb{R}). \tag{5}
\]
Then it can be checked that (see e.g. [7])
\[
F \in C_b^{1, \eta}(H \times H, H) \quad \text{and} \quad G \in C_B^{1, \eta}(H \times H, H). \tag{6}
\]
Moreover, for any \(u, y \in H\) and \(p, r_1, r_2 \in [1, \infty]\) satisfying \(\frac{1}{p} = \frac{1}{r_1} + \frac{1}{r_2}\), there exists a constant \(C_0 > 0\) such that
\[
\|D^2_u F(u, y)(h_1, h_2)\|_{L^p} \leq C_0 \|h_1\|_{L^{r_1}} \|h_2\|_{L^{r_2}} \tag{7}
\]
and
\[
\|D^2_u G(u, y)(h_1, h_2)\|_{L^p} \leq C_0 \|h_1\|_{L^{r_1}} \|h_2\|_{L^{r_2}}. \tag{8}
\]
To give precise results, we write the system (1) in the following abstract formulation in \(H\):
\[
\begin{cases}
\frac{\partial^2 U_t^\varepsilon}{\partial t^2} = AU_t^\varepsilon + F(U_t^\varepsilon, Y_t^\varepsilon) + \dot{W}_t^1, & t \in (0, T], \\
\frac{\partial Y_t^\varepsilon}{\partial t} = \frac{1}{\varepsilon} AY_t^\varepsilon + \frac{1}{\varepsilon} G(U_t^\varepsilon, Y_t^\varepsilon) + \frac{1}{\sqrt{\varepsilon}} \dot{W}_t^2, & t \in (0, T], \\
U_0^\varepsilon = u, \quad \frac{\partial U_t^\varepsilon}{\partial t} \big|_{t=0} = v, \quad Y_0^\varepsilon = y.
\end{cases} \tag{9}
\]
Throughout this paper, we assume that SPDE (9) is well-posed (see Remark 1 (ii) for an explanation). Recall that for \( i = 1, 2 \), \( W_i \) are \( Q_i \)-Wiener processes. We assume that \( Q_i \) are nonnegative, symmetric operators, which have the same eigenfunctions \( \{ e_n \}_{n \in \mathbb{N}} \) as the operator \( A \) (which implies that \( Q_i \) commute with \( A \) on \( D(A) \), hence commute with \( e^{\alpha A} \), \( t > 0 \)), i.e.,

\[
Q_i e_n = \beta_{i,n} e_n, \quad \beta_{i,n} > 0, n \in \mathbb{N}.
\]  

(10)

Note that this means that \( W_i \) (\( i = 1, 2 \)) are non-degenerate, which is necessary to obtain the well-posedness of the system with only Hölder continuous coefficients. In addition, we assume that

\[
\text{Tr}(Q_i) = \sum_{n \in \mathbb{N}} \beta_{i,n} < +\infty, \quad i = 1, 2,
\]  

(11)

and for any \( T > 0 \) and some \( \vartheta \geq \max(\eta, 1 - \eta) \) with \( \eta \) being the Hölder regularity of the coefficients in the assumption (6), we have

\[
\int_0^T \Lambda_t^{\frac{1+\vartheta}{2}} \, dt < \infty,
\]  

(12)

where

\[
\Lambda_t := \sup_{n \geq 1} \frac{2\lambda_n}{\beta_{2,n}(e^{2\lambda_n t} - 1)} < \infty,
\]

and \( \lambda_n \) is given by (4). We note that condition (12) comes from [20, Lemma 9] and is crucial to derive Theorem 3 below.

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

\[
dY_t^u = A Y_t^u \, dt + G(u, Y_t^u) \, dt + dW_t^2, \quad Y_0^u = y \in H^1.
\]  

(13)

Under conditions (6), (10), (11) and (12), there exists a unique strong solution \( Y_t^u \) to Eq. (13) (see [20]), and the process \( Y_t^u \) admits a unique invariant measure \( \mu^u(dy) \) (see e.g. [17, Theorem 4 and Proposition 4]). Then, the averaged equation for system (9) is

\[
\begin{aligned}
\frac{\partial^2 \bar{U}_t}{\partial t^2} &= A \bar{U}_t + \bar{F}(\bar{U}_t) + \dot{W}_t^1, \quad t \in (0, T], \\
\bar{U}_0 &= u, \quad \frac{\partial \bar{U}_t}{\partial t} \big|_{t=0} = v,
\end{aligned}
\]  

(14)

where

\[
\bar{F}(u) := \int_H F(u, y) \mu^u(dy).
\]  

(15)
By Theorem 3 below, we have $\bar{F} \in C_b^1(H, H)$. Thus Eq. (14) admits a unique solution $\bar{U}_t$. Let $\dot{U}_t^\varepsilon := \partial U_t^\varepsilon / \partial t$ and $\ddot{U}_t := \partial \dot{U}_t / \partial t$. The following is the first main result of this paper.

**Theorem 1** Let $T > 0$, $u, y \in H^1$, $v \in H$ and $(U_t^\varepsilon, Y_t^\varepsilon)$ satisfy the SPDE (9). Assume that (5) and (10)–(12) hold. Then we have:

(i) (strong convergence) for any $q \geq 1$,

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

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

$$
\sup_{t \in [0, T]} \left( \|\mathbb{E}[\phi(U_t^\varepsilon)] - \mathbb{E}[\tilde{\phi}(\bar{U}_t)]\|^2 + \|\mathbb{E}[\tilde{\phi}(\dot{U}_t^\varepsilon)] - \mathbb{E}[\tilde{\phi}(\ddot{U}_t)]\|^2 \right) \leq C_2 \varepsilon,
$$

where $C_1 = C(T, u, v, y)$, $C_2 = C(T, u, v, y, \phi, \tilde{\phi})$ are positive constants independent of $\varepsilon$ and $\eta$, and $\bar{U}_t$ is the unique solution of the SPDE (14).

**Remark 1** (i) The above result generalizes those of [25, 26] in that we allow the system (1) to be fully coupled, and the coefficients are assumed to be only $\eta$-Hölder continuous with respect to the fast variable. Moreover, the 1/2-order rate of strong convergence in (16) and the 1-order rate of weak convergence in (17) should be optimal, which coincides with the SDE case, see e.g. [37, 45].

Note that both the strong and weak convergence rates do not depend on the $\eta$-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) We mention that the strong well-posedness of the parabolic type SPDEs and stochastic wave equations with Hölder continuous coefficients are proven in [20] and [40] by using the Zvonkin’s transformation, respectively. Note that system (1) is a combination of stochastic wave equation and parabolic SPDE, thus under our assumptions the well-posedness of SPDE (1) may be proved by adopting their arguments. However, this is not the main theme of the current paper and thus we do not plan to carry it out.

In a particular case that the system is not fully coupled, i.e., the fast motion does not depend on the slow variable (when $g(u, y) = g(y)$ in (1)), the well-posedness for the fast equation with Hölder continuous coefficients is given by [20, Theorem 7]. Thus under our assumptions we can obtain the strong well-posedness of the whole system (1) directly in this case.

Recall that $Z_t^\varepsilon$ is defined by (2). In view of (9) and (14), we have

$$
\frac{\partial^2 Z_t^\varepsilon}{\partial t^2} = AZ_t^\varepsilon + \frac{1}{\sqrt{\varepsilon}} \left[ F(U_t^\varepsilon, Y_t^\varepsilon) - \bar{F} (\bar{U}_t) \right]
$$

$$
= AZ_t^\varepsilon + \frac{1}{\sqrt{\varepsilon}} \left[ \bar{F} (U_t^\varepsilon) - \bar{F} (\bar{U}_t) \right] + \frac{1}{\sqrt{\varepsilon}} \delta F(U_t^\varepsilon, Y_t^\varepsilon),
$$
where
\[ \delta F(u, y) := F(u, y) - \bar{F}(u). \]

To study the homogenization behavior of \( Z_t^\varepsilon \), we consider the following Poisson equation:
\[ \mathcal{L}_2(u, y)\Psi(u, y) = -\delta F(u, y), \]  \hspace{1cm} (18)

where \( \mathcal{L}_2 \) is the generator of the frozen equation (13) given by
\[ \mathcal{L}_2(u, y)\varphi(y) := \langle Ay + G(u, y), D_y\varphi(y) \rangle + \frac{1}{2} Tr \left[ D^2_y\varphi(y)Q_2 \right], \quad \forall \varphi \in C^2(\mathcal{L}), \]  \hspace{1cm} (19)

and \( u \in H \) is regarded as a parameter. According to Theorem 3 below, there exists a unique solution \( \Psi \) to equation (18). Then, the limit process \( \bar{Z}_t \) of \( Z_t^\varepsilon \) turns out to satisfy the following linear stochastic wave equation:
\[ \begin{cases} \frac{\partial^2 \tilde{Z}_t}{\partial t^2} = A\tilde{Z}_t + D_u\bar{F}(\bar{U}_t).\tilde{Z}_t + \sigma(\bar{U}_t)\tilde{W}_t, & t \in (0, T], \\
\tilde{Z}_0 = 0, \quad \frac{\partial \tilde{Z}_t}{\partial t} \bigg|_{t=0} = 0, \end{cases} \]  \hspace{1cm} (20)

where \( \tilde{W}_t \) is another cylindrical Wiener process independent of \( \tilde{W}_1^1 \), and \( \sigma \) is defined as
\[ \frac{1}{2} \sigma(u)\sigma^*(u) = \delta F \otimes \Psi(u) := \int_H \left[ \delta F(u, y) \otimes \Psi(u, y) \right] \mu^u(dy). \]

It can be checked that \( \sigma(u) \) is a Hilbert-Schmidt operator, for all \( u \in H \), see [46, Remark 2.5]. Moreover, by (6), (7) and Theorem 3 below, we have that
\[ \sigma(u) \in C^1_b(H, \mathcal{L}(H)), \]  \hspace{1cm} (21)

and for \( h \in H, k \in H^1 \),
\[ \|D^2_u\sigma(u).(h, k)\|_{\mathcal{L}(H)} \leq C_0 \|h\|\|k\|_1. \]  \hspace{1cm} (22)

Let \( \dot{Z}_t^\varepsilon := \partial Z_t^\varepsilon / \partial t \) and \( \dot{\tilde{Z}}_t := \partial \tilde{Z}_t / \partial t \). Itô’s formulation of Eq. (20) is
\[ \begin{cases} d\tilde{Z}_t = \dot{\tilde{Z}}_t dt, \\
d\tilde{Z}_t = A\tilde{Z}_t dt + D_u\bar{F}(\bar{U}_t).\tilde{Z}_t dt + \sigma(\bar{U}_t)dW_t, \\
\tilde{Z}_0 = 0, \quad \dot{\tilde{Z}}_0 = 0. \end{cases} \]
We have the following result.

**Theorem 2** (Normal deviations) Let $T > 0$, $u, y \in H^1$, $v \in H$, $(U^\varepsilon_t, Y^\varepsilon_t)$ satisfy the SPDE (9) and $Z^\varepsilon_t$ given by (2). Assume that (5) and (10)–(12) hold. Then for any $\phi \in C^3_b(H)$ and $\tilde{\phi} \in C^3_b(H^{-1})$, we have

$$
\sup_{t \in [0,T]} \left( |\mathbb{E}[\phi(Z^\varepsilon_t)] - \mathbb{E}[\phi(\bar{Z}_t)]| + |\mathbb{E}[\tilde{\phi}(\bar{Z}^\varepsilon_t)] - \mathbb{E}[\tilde{\phi}(\bar{Z}_t)]| \right) \leq C_3 \varepsilon^{\frac{1}{2}},
$$

(23)

where $C_3 = C(T, u, \dot{u}, y, \phi, \tilde{\phi}) > 0$ is a constant independent of $\varepsilon$ and $\eta$, and $\bar{Z}_t$ is the unique solution of the SPDE (20).

**Remark 2** The $1/2$ order rate of convergence in (23) is known to be optimal in the finite dimensional situation in view of the asymptotic expansion in [35]. Moreover, the convergence rate does not depend on the $\eta$-regularity of the coefficients in the equation for the fast variable.

3 Preliminaries

3.1 Poisson equation

We will rewrite the system (9) as an abstract evolution equation. To this end, we first introduce some notations. For $\alpha \in \mathbb{R}$, by $H^\alpha := H^\alpha \times H^\alpha_{-1}$ we denote the Hilbert space endowed with the scalar product

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

and norm

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

For simplicity, we write $H := H \times H^{-1}$. Let $\Pi_1$ be the canonical projection from $H$ to $H$, and define

$$
V^\varepsilon_t := \frac{d}{dt} U^\varepsilon_t \quad \text{and} \quad X^\varepsilon_t := (U^\varepsilon_t, V^\varepsilon_t)^T.
$$

Then, the system (9) can be rewritten as

$$
\begin{align*}
\frac{dX^\varepsilon_t}{dt} &= AX^\varepsilon_t dt + \mathcal{F}(X^\varepsilon_t, Y^\varepsilon_t) dt + BdW^1_t, \\
\frac{dY^\varepsilon_t}{dt} &= \varepsilon^{-1} AY^\varepsilon_t dt + \varepsilon^{-1} \mathcal{G}(X^\varepsilon_t, Y^\varepsilon_t) dt + \varepsilon^{-1/2} dW^2_t,
\end{align*}
$$

(24)

$$
X^\varepsilon_0 = x, \ Y^\varepsilon_0 = y,
$$
where \( x := (u, v)^T \), \( \mathcal{G}(x, y) := G(\Pi_1(x), y) \) and
\[
\mathcal{A} := \begin{pmatrix} 0 & I \\ A & 0 \end{pmatrix}, \quad \mathcal{F}(x, y) := \begin{pmatrix} 0 \\ F(\Pi_1(x), y) \end{pmatrix}, \quad \text{BdW}_t^1 := \begin{pmatrix} 0 \\ dW_t^1 \end{pmatrix}.
\]

Thus according to (6)–(8) and the Sobolev embedding \( H^1(0, L) \hookrightarrow L^\infty(0, L) \), we have
\[
\mathcal{F} \in C_b^{1, \eta}(\mathcal{H} \times \mathcal{H}, \mathcal{H}^1) \quad \text{and} \quad \mathcal{G} \in C_B^{1, \eta}(\mathcal{H} \times \mathcal{H}, \mathcal{H}), \quad (25)
\]
and for any \( x, k \in \mathcal{H}^1, h \in \mathcal{H}, y \in \mathcal{H} \), there exists a constant \( C_0 > 0 \) such that
\[
\|D_x^2 \mathcal{F}(x, y). (h, k)\|_1 \leq C_0 \|h\|_0 \|k\|_1 \quad (26)
\]
and
\[
\|D_x^2 \mathcal{G}(x, y). (h, k)\| \leq C_0 \|h\|_0 \|k\|_1. \quad (27)
\]
Similarly, concerning the averaged equation (14), let
\[
\bar{V}_t := \frac{d}{dt} \bar{U}_t \quad \text{and} \quad \bar{X}_t := (\bar{U}_t, \bar{V}_t)^T.
\]
Then we can transform (14) into a stochastic evolution equation:
\[
d\bar{X}_t = \mathcal{A} \bar{X}_t dt + \mathcal{F}(\bar{X}_t) dt + \text{BdW}_t^1, \quad \bar{X}_0 = x = (u, v)^T, \quad (28)
\]
where
\[
\mathcal{F}(x) := \begin{pmatrix} 0 \\ \bar{F}(\Pi_1(x)) \end{pmatrix}, \quad (29)
\]
and \( \bar{F} \) is defined by (15). It is known (see e.g. [4]) that \( \mathcal{A} \) generates a strongly continuous group \( \{e^{t\mathcal{A}}\}_{t \geq 0} \) which is given by
\[
e^{t\mathcal{A}} = \begin{pmatrix} C_t & (\mathcal{A})^{-\frac{1}{2}} S_t \\ -(\mathcal{A})^{\frac{1}{2}} S_t & C_t \end{pmatrix}, \quad (30)
\]
where \( C_t := \cos((-\mathcal{A})^{\frac{1}{2}} t) \) and \( S_t := \sin((-\mathcal{A})^{\frac{1}{2}} t) \). For any \( x \in \mathcal{H} \), we have
\[
\|e^{t\mathcal{A}}x\|_0 \leq \|x\|_0, \quad \forall x \in \mathcal{H}^1.
\]
(31)
The Poisson equation will be the crucial tool in our paper. Recall that \( L_2(u, y) \) is defined by (19). With a slight abuse of notation, we shall also write

\[
L_2\varphi(y) := L_2(x, y)\varphi(y) := L_2(\Pi_1(x), y)\varphi(y), \quad \forall \varphi \in C_\ell^2(H). \tag{32}
\]

For an \( H \)-valued function \( \bar{\varphi} \), we let \( \varphi_n(y) := \langle \bar{\varphi}(y), e_n \rangle \) and write

\[
L_2\bar{\varphi}(y) := \sum_{n \in \mathbb{N}} L_2\varphi_n(y) e_n.
\]

Consider the following Poisson equation:

\[
L_2(x, y)\psi(x, y) = -\phi(x, y), \tag{33}
\]

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

\[
\int_{\mathcal{H}} \phi(x, y)\mu^x(dy) = 0, \quad \forall x \in \mathcal{H}, \tag{34}
\]

where \( \mu^x(dy) = \mu^u(dy) \) is the unique invariant measure of (13). Furthermore, we assume that

\( A_\phi): \phi(x, y) \in C_\ell^{1,\eta}(\mathcal{H} \times H, \hat{\mathcal{H}}), \) and for any \( x, k \in \mathcal{H}^1, h \in \mathcal{H}, y \in H, \)

\[
\|D_x^2\phi(x, y).(h, k)\|_{\hat{\mathcal{H}}} \leq C_0(1 + \|x\|_1 + \|y\|)\|h\|_0\|k\|_1.
\]

The following results come from [46, Theorem 3.2 and Lemma 3.7], we provide a sketch of the proof for the reader’s convenience.

**Theorem 3** Assume that (10)–(12), (25) and (27) hold. Then for every \( \phi : \mathcal{H} \times H \to \hat{\mathcal{H}} \) satisfying the centering condition (34) and the assumption \( (A_\phi) \), there exists a unique solution \( \psi \in C_\ell^{0,2}(\mathcal{H} \times H, \hat{\mathcal{H}}) \cap C_\ell^{1,0}(\mathcal{H} \times H, \hat{\mathcal{H}}) \) to Eq. (33) which is given by

\[
\psi(x, y) = \int_0^\infty \mathbb{E}[\phi(x, Y^x_t(y))] dt,
\]

where \( Y^x_t(y) = Y^u_t(y) \) satisfies the frozen equation (13). Moreover, for any \( x, k \in \mathcal{H}^1, h \in \mathcal{H} \) and \( y \in H, \)

\[
\|D_x^2\psi(x, y).(h, k)\|_{\hat{\mathcal{H}}} \leq C_0(1 + \|x\|_1 + \|y\|)\|h\|_0\|k\|_1. \tag{35}
\]

Furthermore, assume that \( F \) satisfies (6) and (7), and let \( \tilde{F}, \bar{F} \) be defined by (15) and (29), respectively. Then we have \( \tilde{F} \in C_b^1(H, H) \) and \( \bar{F} \in C_b^1(\mathcal{H}, \mathcal{H}^1) \). Moreover, for any \( u, k_1 \in H^1, y, h_1 \in H, \)

\[
\|D_u^2\tilde{F}(u).(h_1, k_1)\| \leq C_0\|h_1\|\|k_1\|_1.
\]
and for any $x, k \in \mathcal{H}^1, h \in \mathcal{H}, y \in H$,

$$\|D_x^2 \tilde{F}(x).(h, k)\|_1 \leq C_0 \|h\|_0 \|k\|_1,$$

where $C_0 > 0$ is a constant.

**Proof** For any $x \in \mathcal{H}^1$ and $y \in H$, let

$$T_t \phi(x, y) := \mathbb{E}\left[\phi(x, Y^x_t(y))\right].$$

Since $\phi$ satisfies the centering condition, we have (see [46, Lemma 3.4]) for any $t > 0$ that

$$|T_t \phi(x, y)| \leq C_0 \|\phi\|_{L^\infty} (1 + \|x\|_1 + \|y\|) e^{-\lambda t}, \quad (36)$$

where $\lambda > 0$ is a constant. Furthermore, by Duhamel’s formula we have

$$T_t \phi(x, y) = P_t \phi(x, y) + \int_0^t P_{t-s} \langle G, D_y T_s \phi \rangle(x, y) ds, \quad (37)$$

where

$$P_t \phi(x, y) := \mathbb{E}\left[\phi(x, R_t(y))\right],$$

and $R_t(y)$ is the Ornstein-Uhlenbeck process satisfying

$$dR_t = AR_t dt + dW_t^2, \quad R_0 = y \in H.$$

Under (10)–(12) and by [20, Theorem 4], for any $T > 0, \phi \in L^\infty_{\mathcal{L}}(\mathcal{H} \times H)$ and $t \in (0, T)$, we have $P_t \phi(x, y) \in C^{0,2}_\mathcal{L}(\mathcal{H} \times H)$ and

$$\|D_y P_t \phi(x, y)\| \leq C_T \frac{1}{\sqrt{t}} \|\phi\|_{L^\infty} (1 + \|x\|_1 + \|y\|), \quad (38)$$

and for any $\eta \in [0, 1]$,

$$\|D_y^2 P_t \phi(x, y)\|_{\mathcal{L}(H)} \leq C_T \frac{1}{t^{1-\eta/2}} \|\phi\|_{C^{0,\eta}_\mathcal{L}} (1 + \|x\|_1 + \|y\|), \quad (39)$$

where $C_T > 0$ is a constant. Estimates (38) and (39) together with (36) and (37) yield that (see [46, Lemma 3.6]) $T_t \phi(x, y) \in C^{0,2}_\mathcal{L}(\mathcal{H} \times H)$ and

$$\|D_y^2 T_t \phi(x, y)\|_{\mathcal{L}(H)} \leq C_0 \frac{1}{t^{1-\eta/2}} \wedge 1 \|\phi\|_{C^{0,\eta}_\mathcal{L}} (1 + \|x\|_1 + \|y\|) e^{-\lambda t}.$$
This in turn implies that \( \psi \in C^{0,2}_c(\mathcal{H} \times H) \), and the assertion that \( \psi \) is the unique solution for Eq. (33) follows by Itô’s formula, see e.g. [6, Lemma 4.3]. The assertion \( \psi \in C^{1,0}_c(\mathcal{H} \times H) \) and (35) can be proved in the same way as in [46, Theorem 3.2].

Below, we prove the regularity properties of the averaged coefficients. For \( x = (u, v)^T \in \mathcal{H}^1, y \in H \) and \( h = (h_1, h_2)^T \in \mathcal{H} \), we have by the transformation formula in [46, Lemma 3.7] that

\[
D_u \tilde{F}(u).h_1 = D_x \tilde{F}(\Pi_1(x)).h_1
\]

\[
= \int_{\mathcal{H}} \left[ D_x F(\Pi_1(x), y).h_1 + \langle D_x G(\Pi_1(x), y).h_1, D_y \psi(\Pi_1(x), y) \rangle \right] \mu^x(\text{d}y).
\]

Thus we have

\[
\| D_x \tilde{F}(\Pi_1(x)).h_1 \| \leq C_0 \| h_1 \|,
\]

and this further implies that

\[
\| D_x \tilde{F}(x).h \|_1 = \| D_u \tilde{F}(u).h_1 \| \leq C_0 \| h_1 \| \leq C_0 \| h \|_0.
\]

Similarly, for \( x = (u, v)^T \in \mathcal{H}^1, k = (k_1, k_2)^T \in \mathcal{H}^1, y \in H, h = (h_1, h_2)^T \in \mathcal{H} \), we have

\[
D_u^2 \tilde{F}(u).(h_1, k_1) = D_x^2 \tilde{F}(\Pi_1(x)).(h_1, k_1)
\]

\[
= \int_{\mathcal{H}} \left[ D_x^2 F(\Pi_1(x), y).(h_1, k_1) + 2 D_y D_x \psi(\Pi_1(x), y).(k_1, D_x G(\Pi_1(x), y).h_1)
\right.
\]

\[
+ \left. \langle D_x^2 G(\Pi_1(x), y).(h_1, k_1), D_y \psi(\Pi_1(x), y) \rangle \right] \mu^x(\text{d}y),
\]

which yields

\[
\| D_u^2 \tilde{F}(u).(h_1, k_1) \| \leq C_0 \| h_1 \| \| k_1 \|_1,
\]

and

\[
\| D_x^2 \tilde{F}(x).(h, k) \|_1 = \| D_u^2 \tilde{F}(u).(h_1, k_1) \| \leq C_0 \| h_1 \| \| k_1 \|_1 \leq C_0 \| h \|_0 \| k \|_1.
\]

The proof is finished. \( \square \)

### 3.2 Galerkin approximation

Itô’s formula will be used frequently below in the proof of the main result. However, due to the presence of unbounded operators in the equation, we can not apply Itô’s formula for SPDE (24) directly. For this reason, we use the standard Galerkin approximation scheme to approximate the infinite dimensional setting to a finite dimensional one.
We provide some details below for the readers’ convenience, see e.g., [6, Section 4] or [13, Section 6].

For every \( n \in \mathbb{N} \), let \( H_n = \text{span}\{e_1, e_2, \cdots, e_n\} \). Denote the projection of \( H \) onto \( H_n \) by \( P_n \), and set

\[
\mathcal{F}_n(x, y) := \begin{pmatrix} 0 \\ P_n F(\Pi_1(x), y) \end{pmatrix}, \quad \mathcal{G}_n(x, y) := P_n G(\Pi_1(x), y).
\]

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:

\[
\begin{aligned}
\begin{cases}
\text{d}X_t^{n,\varepsilon} &= \mathcal{A}X_t^{n,\varepsilon} \text{d}t + \mathcal{F}_n(X_t^{n,\varepsilon}, Y_t^{n,\varepsilon}) \text{d}t + P_n \text{d}W_t^1, \\
\text{d}Y_t^{n,\varepsilon} &= \varepsilon^{-1} AY_t^{n,\varepsilon} \text{d}t + \varepsilon^{-1} \mathcal{G}_n(X_t^{n,\varepsilon}, Y_t^{n,\varepsilon}) \text{d}t + \varepsilon^{-1/2} P_n \text{d}W_t^2,
\end{cases}
\end{aligned}
\]

with initial values \( X_0^{n,\varepsilon} = x^n \in H_n \times H_n \) and \( Y_0^{n,\varepsilon} = y^n \in H_n \). Since system (40) is a finite dimensional one, there exists a unique strong solution, see e.g. [45]. The corresponding averaged equation for system (40) is given by

\[
\text{d}\tilde{X}_t^n = \mathcal{A}\tilde{X}_t^n \text{d}t + \mathcal{F}_n(\tilde{X}_t^n) \text{d}t + P_n \text{d}W_t^1, \quad \tilde{X}_0^n = x^n \in H_n \times H_n,
\]

where

\[
\tilde{\mathcal{F}}_n(x) := \int_{H_n} \mathcal{F}_n(x, y) \mu_n^x(\text{d}y),
\]

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

\[
\text{d}Y_t^{x,n} = AY_t^{x,n} \text{d}t + \mathcal{G}_n(x, Y_t^{x,n}) \text{d}t + P_n \text{d}W_t^2, \quad Y_0^{x,n} = y^n \in H_n.
\]

Recall that \( Y_t^n(y) \) satisfies (13) and note that \( \mathcal{G}(x, y) = G(u, y) \). We know that \( Y_t^{x,n}(y^n) \) converges strongly to \( Y_t^n(y) := Y_t^n(y) \) (see e.g. [21, Theorem 7, Step 3].) Let \( T > 0, x \in \mathcal{H}_1 \) and \( y \in H^1 \). Then as shown in the proof of [19, Lemma 3.1], for any \( q \geq 1 \) 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 (45), (50) and (30) we deduce that

\[
\mathbb{E}\|\tilde{X}_t^n - \tilde{X}_t\|_1^q \leq \mathbb{E}\left\|\int_0^t e^{(t-s)A}(I - P_n)BdW_s^1\right\|_1^q
\]

\[
+ \mathbb{E}\left(\int_0^t \left\|(-A)^{-\frac{1}{2}}S_{t-s}(\tilde{F}(\tilde{U}_s) - \tilde{F}_n(\tilde{U}_s))\right\|_1 + \|C_{t-s}(\tilde{F}(\tilde{U}_s) - \tilde{F}_n(\tilde{U}_s))\|_1\right)^q ds\right)^q.
\]
Since \( \| \bar{F}_n - \bar{F} \| \to 0 \) as \( n \to \infty \) (see e.g. [6, (4.4)]), the first two terms go to 0 as \( n \to \infty \) by the dominated convergence theorem. For the last term, by Theorem 3 we know that \( \bar{F} \) is Lipschitz continuous, hence \( \bar{F}_n \) is also Lipschitz continuous with bounds which are uniform with respect to \( n \). Thus, we have

\[
\mathbb{E} \left( \int_0^t \left( \| -A \|^{-1/2} S_{t-s} (\bar{F}_n (\bar{U}_s) - \bar{F}_n (\bar{U}_s^n)) \|_1 \right. \\
+ \left. \| C_{t-s} (\bar{F}_n (\bar{U}_s) - \bar{F}_n (\bar{U}_s^n)) \| ds \right\}^q \right) \leq C_1 \mathbb{E} \left( \int_0^t \| \bar{U}_s - \bar{U}_s^n \|_1 ds \right\}^q \leq C_1 \mathbb{E} \left( \int_0^t \| \bar{X}_s - \bar{X}_s^n \|_1 ds \right\}^q,
\]

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 1, we only need to show that for any \( q \geq 1 \),

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

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

\[
\sup_{t \in [0,T]} \| \mathbb{E} [\varphi (X_t^{n, \varepsilon})] - \mathbb{E} [\varphi (\bar{X}_t^n)] \| \leq C_T \varepsilon,
\]

where \( C_T > 0 \) is a constant independent of \( n \). In the rest of this paper, we shall only work with the approximating system (40), 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 \( H \).

### 3.3 Moment estimates

We prove the following estimates for the mild solution of the system (24).

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

\[
\begin{align*}
X_t^\varepsilon &= e^{tA}x + \int_0^t e^{(t-s)A} \mathcal{F} (X_s^\varepsilon, Y_s^\varepsilon) ds + \int_0^t e^{(t-s)A} B dW^1_s,
Y_t^\varepsilon &= e^{tA}y + \varepsilon^{-1} \int_0^t e^{(t-s)A} \mathcal{G} (X_s^\varepsilon, Y_s^\varepsilon) ds + \varepsilon^{-1/2} \int_0^t e^{(t-s)A} dW^2_s.
\end{align*}
\]

\( \square \) Springer
Then for any $q \geq 1$, we have
\[
\sup_{\varepsilon \in (0,1)} \mathbb{E} \left( \sup_{t \in [0,T]} \| X_t^\varepsilon \|_1^{2q} \right) \leq C_{T,q} \left( 1 + \| x \|_1^{2q} + \| y \|^{2q} \right)
\]
and
\[
\sup_{\varepsilon \in (0,1)} \sup_{t \in [0,T]} \mathbb{E} \| Y_t^\varepsilon \|^{2q} + \sup_{\varepsilon \in (0,1)} \mathbb{E} \left( \int_0^T \| Y_t^\varepsilon \|_1^q \, dt \right) \leq C_{T,q} \left( 1 + \| y \|^{2q} \right),
\]
where $C_{T,q} > 0$ is a constant.

**Proof** Applying Itô’s formula (see e.g. [38, Section 4.2]) to $\| Y_t^\varepsilon \|^{2q}$ and taking expectation, we have
\[
\frac{d}{dt} \mathbb{E} \| Y_t^\varepsilon \|^{2q} = \frac{2q}{\varepsilon} \mathbb{E} \left[ \| Y_t^\varepsilon \|^{2q-2} (AY_t^\varepsilon, Y_t^\varepsilon) \right] + \frac{2q}{\varepsilon} \mathbb{E} \left[ \| Y_t^\varepsilon \|^{2q-2} (G(X_t^\varepsilon, Y_t^\varepsilon), Y_t^\varepsilon) \right] + \left( \frac{q}{\varepsilon} + \frac{2q(q-1)}{\varepsilon} \right) T r (Q) \mathbb{E} \| Y_t^\varepsilon \|^{2q-2}.
\]
It follows from Poincaré inequality, Young’s inequality and (11) that
\[
\frac{d}{dt} \mathbb{E} \| Y_t^\varepsilon \|^{2q} \leq -\frac{2q \lambda_1}{\varepsilon} \mathbb{E} \| Y_t^\varepsilon \|^{2q} + \frac{2qC_0}{\varepsilon} \mathbb{E} \| Y_t^\varepsilon \|^{2q-1} + \left( \frac{q}{\varepsilon} + \frac{2q(q-1)}{\varepsilon} \right) T r (Q) \mathbb{E} \| Y_t^\varepsilon \|^{2q-2} \leq -\frac{qC_1}{\varepsilon} \mathbb{E} \| Y_t^\varepsilon \|^{2q} + \frac{C_2}{\varepsilon}.
\]
Using Gronwall’s inequality, we obtain
\[
\mathbb{E} \| Y_t^\varepsilon \|^{2q} \leq e^{-\frac{qC_1}{\varepsilon} t} \| y \|^{2q} + \frac{C_2}{\varepsilon} \int_0^t e^{-\frac{qC_1}{\varepsilon} (t-s)} \, ds \leq C_3 (1 + \| y \|^{2q}).
\]
Furthermore, in view of [26, Theorem 3.1], the process $X_t^\varepsilon = (U_t^\varepsilon, V_t^\varepsilon)^T$ enjoys the following energy equality:
\[
\| X_t^\varepsilon \|_1^2 = \| x \|_1^2 + 2 \int_0^t \langle V_s^\varepsilon, F(U_s^\varepsilon, Y_s^\varepsilon) \rangle \, ds + 2 \int_0^t \langle V_s^\varepsilon, dW_s^1 \rangle + \int_0^t T r Q_1 \, ds.
\]
On the one hand, note that
\[
\mathbb{E} \sup_{0 \leq t \leq T} \left| \int_0^t \langle V_s^\varepsilon, F(U_s^\varepsilon, Y_s^\varepsilon) \rangle \, ds \right|^q \leq C_1 \mathbb{E} \left( \int_0^T \| V_s^\varepsilon \|_1^2 \, ds \right)^q + C_1 \mathbb{E} \left( \int_0^T (1 + \| U_s^\varepsilon \|^2 + \| Y_s^\varepsilon \|^2) \, ds \right)^q \leq C_1 \mathbb{E} \left( \int_0^T \| X_s^\varepsilon \|_1^2 \, ds \right)^q + C_1 \mathbb{E} \left( \int_0^T (1 + \| Y_s^\varepsilon \|^2) \, ds \right)^q.
\]
On the other hand, in view of Burkholder–Davis–Gundy’s inequality and Young’s inequality, we have

\[
\mathbb{E} \sup_{0 \leq t \leq T} \left| \int_0^t \langle V_s^\varepsilon, \, dW_s^1 \rangle \right|^q \leq C_2 T r Q_1 \mathbb{E} \left( \int_0^T \| V_s^\varepsilon \|^2 \, ds \right)^{\frac{q}{2}} + C_2 \\
\leq C_2 \mathbb{E} \left( \int_0^T \| X_s^\varepsilon \|_1^{2q} \, ds \right) + C_2. \tag{49}
\]

Combining (48) and (49), we get

\[
\mathbb{E} \left( \sup_{0 \leq t \leq T} \| X_t^\varepsilon \|_1^{2q} \right) \leq C_3 \left( 1 + \| x \|_1^{2q} + \int_0^T \mathbb{E} \| Y_s^\varepsilon \|_1^{2q} \, ds \right).
\]

Thus, it follows from Gronwall’s inequality that

\[
\mathbb{E} \left( \sup_{0 \leq t \leq T} \| X_t^\varepsilon \|_1^{2q} \right) \leq C_4 \left( 1 + \| x \|_1^{2q} + \int_0^T \mathbb{E} \| Y_s^\varepsilon \|_1^{2q} \, ds \right),
\]

which together with (47) yields

\[
\mathbb{E} \left( \sup_{0 \leq t \leq T} \| X_t^\varepsilon \|_1^{2q} \right) \leq C_5 \left( 1 + \| x \|_1^{2q} + \| y \|^{2q} \right).
\]

To prove the second estimate of (46), we deduce that

\[
\mathbb{E} \left( \int_0^T \| Y_t^\varepsilon \|_1^{2q} \, dt \right)^{\frac{q}{2}} \leq C_q \left( \int_0^T \| e^{\frac{t}{2}} A y \|_1^2 \, dt \right)^{\frac{q}{2}} \\
+ C_q \mathbb{E} \left( \int_0^T \| e^{-1} \int_0^t e^{\frac{t-s}{2}} A G(X_s^\varepsilon, Y_s^\varepsilon) \, ds \|_1^2 \, dt \right)^{\frac{q}{2}} \\
+ C_q \mathbb{E} \left( \int_0^T \| e^{-1/2} \int_0^t e^{\frac{t-s}{2}} A dW_s^2 \|_1^2 \, dt \right)^{\frac{q}{2}} \\
= : \sum_{i=1}^3 \mathcal{G}_i(T, \varepsilon).
\]

For the first term, we have

\[
\mathcal{G}_1(T, \varepsilon) \leq C_6 \varepsilon \left( \int_0^{T/\varepsilon} \sum_{k=1}^{\infty} \lambda_k e^{-2\lambda_k t} \langle y, e_k \rangle^2 \, dt \right)^{q} \\
\leq C_6 \left( \sum_{k=1}^{\infty} (1 - e^{-2\lambda_k T/\varepsilon}) \langle y, e_k \rangle^2 \right)^{q} \leq C_6 \| y \|^{2q}.
\]
Note that
\[
\| \varepsilon^{-1} \int_0^t e^{\frac{t-s}{\varepsilon}} A G(s, X_s^\varepsilon, Y_s^\varepsilon) ds \| \leq C_7 \varepsilon^{-1} \int_0^t \left( \frac{t-s}{\varepsilon} \right)^{-1/2} e^{-\frac{\lambda_1 (t-s)}{2\varepsilon}} ds
\]
\[
\leq C_7 \int_0^{t/\varepsilon} e^{-\frac{\lambda_1 s}{2 s^{1/2}}} ds \leq C_7,
\]
which implies that
\[
\mathcal{B}_2(T, \varepsilon) \leq C_8.
\]
For the last term, by Minkowski’s inequality, Burkholder–Davis–Gundy’s inequality and (11), we deduce that
\[
\mathcal{B}_3(T, \varepsilon) \leq C_9 \left\{ \int_0^T \left( \mathbb{E} \left( \int_0^t e^{-\frac{t-s}{\varepsilon}} A dW_s^2 \right)^{2q} \right)^{1/q} ds \right\}^{1/q} \leq C_9.
\]
Combining the above computations, we get the desired result.

The following estimates for the solution of the averaged equation (28) can be proved in a similar way as Lemma 1, hence we omit the details here.

**Lemma 2** Let \( T > 0 \) and \( x \in H^1 \). The averaged equation (28) admits a unique mild solution \( \tilde{X}_t \) such that for all \( t \geq 0 \),
\[
\tilde{X}_t = e^{tA} x + \int_0^t e^{(t-s)A} \tilde{F}(\tilde{X}_s) ds + \int_0^t e^{(t-s)A} B dW_1^s.
\]
Moreover, for any \( q \geq 1 \) we have
\[
\mathbb{E} \left( \sup_{t \in [0, T]} \| \tilde{X}_t \|_{1}^{2q} \right) \leq C_{T,q} (1 + \|x\|_1^{2q}),
\]
where \( C_{T,q} > 0 \) is a constant.
4 Strong and weak convergence in the averaging principle

4.1 Proof of Theorem 1 (strong convergence)

For simplicity, let
\[
\mathcal{L}_1 \phi(x) := \mathcal{L}_1(x, y) \phi(x) := \langle Ax + \mathcal{F}(x, y), D_x \phi(x) \rangle_H + \frac{1}{2} \text{Tr} \left( D_x^2 \phi(x) (BQ_1)^{1/2} ((BQ_1)^{1/2})^* \right), \quad \forall \phi \in C^2_c(H). \tag{51}
\]

As shown in Subsection 4.1, to prove the strong convergence result (16), we only need to prove (43). To this end, we first establish the following fluctuation estimate for an integral functional of \((X^\varepsilon_s, Y^\varepsilon_s)\) over time interval \([0, t]\), which will play an important role in proving (43).

**Lemma 3** (Strong fluctuations estimate) Let \(T, \eta > 0, x = (u, v)^T \in \mathcal{H}^1\) and \(y \in H^1\). Assume that (25)–(27) hold. Then for any \(t \in [0, T], q \geq 1\) and every \(\tilde{\phi}(x, y) := \left( \begin{array}{c} 0 \\ \phi(u, y) \end{array} \right)\) with \(\phi \in C^{1,\eta}_b(\mathcal{H} \times \mathcal{H}, H)\) satisfying the centering condition (34) and

\[
\|D_u^2 \phi(u, y). (h, k)\| \leq C_0 \|h\| \|k\|_1, \quad \forall h \in H, k \in H^1,
\]

we have

\[
E \left\| \int_0^t e^{(t-s)A} \tilde{\phi}(X^\varepsilon_s, Y^\varepsilon_s) ds \right\|_1^q \leq C_{T, q} \varepsilon^{q/2},
\]

where \(C_{T, q} > 0\) is a constant independent of \(\varepsilon, \eta\) and \(n\).

**Proof** Since \(\phi\) satisfies the centering condition (34) and assumption \((A, \phi)\), according to Theorem 3, \(\psi \in C^{0,2}_c(\mathcal{H} \times \mathcal{H}, H) \cap C^{1,0}_b(\mathcal{H} \times \mathcal{H}, H)\) solves the Poisson equation

\[
\mathcal{L}_2(u, y) \psi(u, y) = -\phi(u, y).
\]

Moreover, for any \(h \in H, k \in H^1\),

\[
\|D_u^2 \psi(u, y). (h, k)\| \leq C_0 \|h\| \|k\|_1.
\]

Define

\[
\tilde{\psi}_t(s, x, y) := e^{(t-s)A} \tilde{\psi}(x, y) := e^{(t-s)A} \left( \begin{array}{c} 0 \\ \psi(u, y) \end{array} \right).
\]

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

\[
\mathcal{L}_2 \tilde{\psi}_t(s, x, y) = -e^{(t-s)A} \tilde{\phi}(x, y).
\]
Applying Itô’s formula to \( \tilde{\psi}_t(t, X_t^\varepsilon, Y_t^\varepsilon) \), we get

\[
\tilde{\psi}_t(t, X_t^\varepsilon, Y_t^\varepsilon) = \tilde{\psi}_t(0, x, y) + \int_0^t (\partial_s + \mathcal{L}_1) \tilde{\psi}_t(s, X_s^\varepsilon, Y_s^\varepsilon) \, ds \\
+ \frac{1}{\varepsilon} \int_0^t \mathcal{L}_2 \tilde{\psi}_t(s, X_s^\varepsilon, Y_s^\varepsilon) \, ds + M_t^1 + \frac{1}{\sqrt{\varepsilon}} M_t^2,
\]

(53)

where \( M_t^1 \) and \( M_t^2 \) are defined by

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

Multiplying both sides of (53) by \( \varepsilon \) and using (52), we obtain

\[
\int_0^t e^{(t-s)} A \phi(X_s^\varepsilon, Y_s^\varepsilon) \, ds = - \int_0^t \mathcal{L}_2 \tilde{\psi}_t(s, X_s^\varepsilon, Y_s^\varepsilon) \, ds \\
= \varepsilon \left[ \tilde{\psi}_t(0, x, y) - \tilde{\psi}_t(t, X_t^\varepsilon, Y_t^\varepsilon) \right] + \varepsilon \int_0^t \partial_s \tilde{\psi}_t(s, X_s^\varepsilon, Y_s^\varepsilon) \, ds \\
+ \varepsilon \int_0^t \mathcal{L}_1 \tilde{\psi}_t(s, X_s^\varepsilon, Y_s^\varepsilon) \, ds + \varepsilon M_t^1 + \sqrt{\varepsilon} M_t^2 =: \sum_{i=1}^5 \mathcal{J}_i(t, \varepsilon).
\]

(54)

Note that

\[
\left\| e^{(t-s)} A \tilde{\psi}(x, y) \right\|_1^2 = \left\| \begin{pmatrix} (-A)^{-\frac{1}{2}} S_{t-s} \psi(u, y) \\ C_{t-s} \tilde{\psi}(u, y) \end{pmatrix} \right\|_1^2 \\
= \left\| (-A)^{-\frac{1}{2}} S_{t-s} \psi(u, y) \right\|_1^2 + \left\| C_{t-s} \tilde{\psi}(u, y) \right\|_1^2 \\
\leq 2 \left\| \psi(u, y) \right\|_1^2 \leq C_1 (1 + \|u\|_2^2 + \|y\|_2^2).
\]

As a result, by Lemma 1 we get

\[
\mathbb{E} \left\| \mathcal{J}_1(t, \varepsilon) \right\|_1^q \leq C_1 \varepsilon^q (1 + \mathbb{E} \|U_t^\varepsilon\|_1^q + \mathbb{E} \|Y_t^\varepsilon\|_1^q) \leq C_1 \varepsilon^q.
\]

Note that

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

and by the fact that \( \psi(\cdot, y) \in C_b^1(H, H), \psi(u, \cdot) \in C^1_\ell(H, H) \) and [14, (2.17)], we have

\[
\left\| A e^{(t-s)} A \tilde{\psi}(x, y) \right\|_1 = \left\| \begin{pmatrix} C_{t-s} \psi(u, y) \\ -(-A)^{\frac{1}{2}} S_{t-s} \psi(u, y) \end{pmatrix} \right\|_1
\]
\[
\|C_{t-s} \psi(u, y)\|_1 + \| - (A)^{1/2} S_{t-s} \psi(u, y)\|_1 \\
\leq 2\|\psi(u, y)\|_1 = 2\|\psi(u, y)\| + 2\|\nabla_u \psi(u, y)\| \nabla u + \nabla_y \psi(u, y)\|y \\
\leq C_2 (1 + \|u\| + \|y\|) + C_2 (1 + \|y\|) (\|u\|_1 + \|y\|_1) \\
\leq C_2 (1 + \|y\|^2 + \|y\|^2).
\]

Thus, using Minkowski’s inequality and Lemma 1 again, we have
\[
E\|J_2(t, \varepsilon)\|_1^q \leq C_2 \varepsilon^q \left( \int_0^T \left( 1 + E\|U^\varepsilon_s\|_1^{2q} \right)^{1/q} \, dr \right)^q \\
+ C_2 \varepsilon^q E\left( \int_0^T \|Y^\varepsilon_s\|_1^q \, ds \right)^q \leq C_2 \varepsilon^q.
\]

For the third term, we have
\[
|L_1 \tilde{\psi}_t(s, X^\varepsilon_s, Y^\varepsilon_s)| \leq C_3 (1 + \|X^\varepsilon_s\|^2 + \|Y^\varepsilon_s\|^2),
\]
which together with Minkowski’s inequality and Lemmas 1 yields that
\[
E\|J_3(t, \varepsilon)\|_1^q \leq C_3 \varepsilon^q \left( \int_0^t \left( E\|X^\varepsilon_s\|_1^2 + \|Y^\varepsilon_s\|^2 \right)^{1/q} \, ds \right)^q \leq C_3 \varepsilon^q.
\]

Finally, by Burkholder–Davis–Gundy’s inequality, Theorem 3, Lemma 1 and (11), we have
\[
E\|J_4(t, \varepsilon)\|_1^q \leq C_4 \varepsilon^q \left( \int_0^T \left( E\|X^\varepsilon_s\|^2 + \|Y^\varepsilon_s\|^2 \right)^{1/2} \, ds \right)^{q/2} \\
\leq C_4 \varepsilon^q \left( \int_0^T \left( 1 + E\|X^\varepsilon_s\|^2 + \|Y^\varepsilon_s\|^2 \right) \, ds \right)^{q/2} \leq C_4 \varepsilon^q,
\]
and similarly,
\[
E\|J_5(t, \varepsilon)\|_1^q \leq C_5 \varepsilon^{q/2}.
\]

Combining the above inequalities with (54), we get the desired estimate. \(\square\)

We are now in the position to give:

**Proof of estimate (43)** Fix \(T > 0\) below. In view of (45) and (50), for every \(t \in [0, T]\) we have
\[
X^\varepsilon_t - \tilde{X}_t = \int_0^t e^{(t-s)A} [\mathcal{F}(X^\varepsilon_s) - \mathcal{F}(\tilde{X}_s)] \, ds + \int_0^t e^{(t-s)A} \delta \mathcal{F}(X^\varepsilon_s, Y^\varepsilon_s) \, ds,
\]
where $\delta F$ is defined by

$$
\delta F(x, y) := F(x, y) - \bar{F}(x) = \left( \begin{array}{c} 0 \\ \delta F(P_1(x), y) \end{array} \right).
$$

(55)

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

$$
\mathbb{E}\|X_t^\varepsilon - \bar{X}_t\|_1^q \leq C_0 \mathbb{E}\left( \int_0^t e^{(t-s)} A[\bar{F}(X_s^\varepsilon) - \bar{F}(\bar{X}_s)]ds \right)^q
+ C_0 \mathbb{E}\left( \int_0^t e^{(t-s)} A \delta F(X_s^\varepsilon, Y_s^\varepsilon)ds \right)^q
=: \mathcal{J}_1(t, \varepsilon) + \mathcal{J}_2(t, \varepsilon).
$$

For the first term, by Minkowski’s inequality and Theorem 3 we deduce that

$$
\mathcal{J}_1(t, \varepsilon) \leq C_1 \mathbb{E}\left( \int_0^t \|\bar{F}(X_s^\varepsilon) - \bar{F}(\bar{X}_s)\|_1 ds \right)^q
\leq C_1 \int_0^t \mathbb{E}\|X_s^\varepsilon - \bar{X}_s\|_1^q ds.
$$

For the second term, noting that $\delta F(x, y)$ satisfies the centering condition (34), it follows by Lemma 3 directly that

$$
\mathcal{J}_2(t, \varepsilon) \leq C_2 \varepsilon^{q/2}.
$$

Thus, we arrive at

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

which together with Gronwall’s inequality yields the desired result. $\square$

### 4.2 Proof of Theorem 1 (weak convergence)

As in the previous subsection, to prove the weak convergence result in Theorem 1, we only need to show (44). 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** (Weak fluctuations estimate) *Let $T, \eta > 0, x = (u, v)^T \in \mathcal{H}^1, y \in H^1$ and (25)–(27) hold. Assume that for every $t > 0$, $\phi(t, \cdot, \cdot)$ satisfies (34), and $\phi(\cdot, x, y) \in C_b^1([0, T]), \phi(t, \cdot, \cdot) \in C^1_{\ell}(\mathcal{H} \times H)$, and for any $h \in \mathcal{H}, k \in H^1$,

$$
|D_x^2 \phi(t, x, y). (h, k)| \leq C_0 (1 + \|x\|_1 + \|y\|) \|h\|_0 \|k\|_1,
$$

and

$$
|\partial_t \phi(t, x, y)| \leq C_0 (1 + \|x\|_1^2 + \|y\|^2).
$$

(56)*
Then we have

\[ \left| \mathbb{E} \left( \int_0^t \phi(s, X_s^\varepsilon, Y_s^\varepsilon) ds \right) \right| \leq C_T \varepsilon, \]

where \( C_T > 0 \) is a constant independent of \( \varepsilon, \eta \) and \( n \).

Proof Since \( \phi \) satisfies the centering condition and \((A_\phi)\), by Theorem 3 there exists a unique \( \psi \) solving the Poisson equation

\[ L_2 \psi(t, x, y) = -\phi(t, x, y), \quad (57) \]

where \( L_2 \) is given by (32). Moreover, \( \psi(t, \cdot, \cdot) \in C^{0,2}_\ell (H \times H) \cap C^{1,0}_\ell (H \times H) \) and for any \( h \in \mathcal{H}, k \in \mathcal{H}^1 \),

\[ |D_x^2 \psi(t, x, y). (h, k)| \leq C_0(1 + \|x\|_1 + \|y\|)\|h\|_0\|k\|_1. \]

Applying Itô’s formula to \( \psi(t, X_t^\varepsilon, Y_t^\varepsilon) \) to get that

\[ \mathbb{E}[\psi(t, X_t^\varepsilon, Y_t^\varepsilon)] = \psi(0, x, y) + \mathbb{E}\left( \int_0^t (\partial_s + L_1) \psi(s, X_s^\varepsilon, Y_s^\varepsilon) ds \right) + \frac{1}{\varepsilon} \mathbb{E}\left( \int_0^t L_2 \psi(s, X_s^\varepsilon, Y_s^\varepsilon) ds \right). \]

Combining this with (57), we obtain

\[ \mathbb{E} \left( \int_0^t \phi(s, X_s^\varepsilon, Y_s^\varepsilon) ds \right) = \varepsilon \mathbb{E}[\psi(0, x, y) - \psi(t, X_t^\varepsilon, Y_t^\varepsilon)] + \varepsilon \mathbb{E}\left( \int_0^t L_1 \psi(s, X_s^\varepsilon, Y_s^\varepsilon) ds \right) + \varepsilon \mathbb{E}\left( \int_0^t \partial_t \psi(s, X_s^\varepsilon, Y_s^\varepsilon) ds \right) =: \sum_{i=1}^3 \mathcal{W}_i(t, \varepsilon). \]

By using exactly the same arguments as in the proof of Lemma 3, we can get that

\[ |\mathcal{W}_1(t, \varepsilon)| + |\mathcal{W}_2(t, \varepsilon)| \leq C_1 \varepsilon. \]

To control the third term, note that

\[ L_2 \partial_t \psi(t, x, y) = -\partial_t \phi(t, x, y). \]

In view of condition (56) and using Theorem 3, we have

\[ |\partial_t \psi(t, x, y)| \leq C_0(1 + \|x\|_1^2 + \|y\|^2), \]
which together with Lemma 1 implies that

\[ |\mathcal{W}_3(t, \varepsilon)| \leq C_2 \varepsilon \mathbb{E} \left( \int_0^t (1 + ||X^\varepsilon_s||_1^2 + ||Y^\varepsilon_s||_1^2) ds \right) \leq C_2 \varepsilon. \]

Combining the above estimates, we get the desired result. \qed

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

\[
\begin{aligned}
\partial_t \bar{u}(t, x) &= \tilde{L}_1 \bar{u}(t, x), \quad t \in (0, T], \\
\bar{u}(0, x) &= \varphi(x),
\end{aligned}
\tag{58}
\]

where \( \varphi: \mathcal{H} \to \mathbb{R} \) is measurable and \( \tilde{L}_1 \) is formally the infinitesimal generator of the process \( \tilde{X} \), given by

\[
\tilde{L}_1 \varphi(x) = \langle Ax + \tilde{F}(x), D_x \varphi(x) \rangle_{\mathcal{H}} + \frac{1}{2} Tr \left( D^2 \varphi(x) (B Q_1 1^2 ((B Q_1 1^2)^*) \right), \quad \forall \varphi \in C^2_b(\mathcal{H}).
\tag{59}
\]

The following result can be proved similarly in [25, Lemmas A.3-A.5 and 4.3].

**Lemma 5** For every \( \varphi \in C^3_b(\mathcal{H}) \), there exists a solution to Eq. (58) which is given by

\[
\bar{u}(t, x) = \mathbb{E} \left[ \varphi(\tilde{X}_t(x)) \right],
\]

and satisfy

(i) \( \bar{u} \in C^{1,1}_b([0, T] \times \mathcal{H}) \);

(ii) for any \( t \in [0, T], h \in \mathcal{H}, x, k \in \mathcal{H}^1 \),

\[
|D^2_x \bar{u}(t, x). (h, k)| \leq C_T ||h||_0 ||k||_1;
\]

(iii) for any \( t \in [0, T], h \in \mathcal{H}, x, k, l \in \mathcal{H}^1 \),

\[
|D^3_x \bar{u}(t, x). (h, k, l)| \leq C_T ||h||_0 ||k||_1 ||l||_1;
\]

(iv) for any \( t \in [0, T] \) and \( x, h \in \mathcal{H}^1 \),

\[
|\partial_t D_x \bar{u}(t, x). h| \leq C_T ||h||_1 (1 + ||x||_1),
\]

where \( C_T > 0 \) is a constant.

Now, we are in the position to give:

**Proof of estimate (44)** Given \( T > 0 \) and \( \varphi \in C^3_b(\mathcal{H}) \), let \( \bar{u} \) solve the Cauchy problem (58). 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
\[ \tilde{u}(T, x) = \bar{u}(0, x) = \varphi(x) \quad \text{and} \quad \tilde{u}(0, x) = \bar{u}(T, x) = \mathbb{E}[\varphi(X_T(x))]. \]

Using Itô’s formula and taking expectation, we deduce that
\[ \mathbb{E}[\varphi(X_T^\varepsilon)] - \mathbb{E}[\varphi(X_T^\varepsilon)] = \mathbb{E}[\tilde{u}(T, X_T^\varepsilon) - \tilde{u}(0, x)] \]
\[ = \mathbb{E}\left( \int_0^T (\partial_t + L_1)\tilde{u}(t, X_t^\varepsilon)dt \right) \]
\[ = \mathbb{E}\left( \int_0^T [\mathcal{L}_1\tilde{u}(t, X_t^\varepsilon) - \bar{\mathcal{L}}_1\tilde{u}(t, X_t^\varepsilon)]dt \right) \]
\[ = \mathbb{E}\left( \int_0^T \langle \delta\mathcal{F}(X_t, Y_t), D_x\tilde{u}(t, X_t^\varepsilon) \rangle_H dt \right). \]

Note that the function
\[ \phi(t, x, y) := \langle \delta\mathcal{F}(x, y), D_x\tilde{u}(t, x) \rangle_H \]
satisfies the centering condition (34). Moreover, by Lemma 5, it is easy to check \( \phi(\cdot, x, y) \in C^1_b([0, T]), \phi(t, \cdot, \cdot) \in C^{1,\eta}_\varepsilon(H \times H), \) and for any \( h \in H, k \in H^1, \)
\[ |D^2_x\phi(t, x, y). (h, k)| \leq C_0 (1 + \|x\|_1 + \|y\|_1)\|h\|_0\|k\|_1, \]
and
\[ |\partial_t\phi(t, x, y)| = |\langle \delta\mathcal{F}(x, y), \partial_t D_x\tilde{u}(T - t, x) \rangle_H | \leq C_0 (1 + \|x\|_1^2 + \|y\|_2^2). \]

As a result of Lemma 4, we have
\[ |\mathbb{E}[\varphi(X_T^\varepsilon)] - \mathbb{E}[\varphi(X_T^\varepsilon)]| \leq C_1 \varepsilon, \]
which completes the proof. \( \square \)

5 Normal deviations

5.1 Cauchy problem

Define
\[ Z_t^\varepsilon := \frac{X_t^\varepsilon - \bar{X}_t}{\sqrt{\varepsilon}}. \]
In view of (24) and (28), we consider the process \((X^\varepsilon_t, Y^\varepsilon_t, \bar{X}_t, Z^\varepsilon_t)\) as the solution to the following system of equations:

\[
\begin{align*}
\text{d}X^\varepsilon_t &= AX^\varepsilon_t \text{d}t + \mathcal{F}(X^\varepsilon_t, Y^\varepsilon_t)\text{d}t + BdW^1_t, \\
\text{d}Y^\varepsilon_t &= \varepsilon^{-1}AY^\varepsilon_t \text{d}t + \varepsilon^{-1}G(X^\varepsilon_t, Y^\varepsilon_t)\text{d}t + \varepsilon^{-1/2}\text{d}W^2_t, \\
\text{d}\bar{X}_t &= A\bar{X}_t \text{d}t + \bar{\mathcal{F}}(\bar{X}_t)dt + BdW^1_t, \\
\text{d}Z^\varepsilon_t &= A\bar{Z}^\varepsilon_t \text{d}t + \varepsilon^{-1/2}[\bar{\mathcal{F}}(X^\varepsilon_t) - \bar{\mathcal{F}}(\bar{X}_t)]\text{d}t + \varepsilon^{-1/2}\delta\mathcal{F}(X^\varepsilon_t, Y^\varepsilon_t)\text{d}t, \quad Z^\varepsilon_0 = 0,
\end{align*}
\]

where \(\delta\mathcal{F}\) is defined by (55). By the definition and as a result of Theorem 1, we have that for any \(q \geq 1\),

\[
\sup_{0 \leq t \leq T} \mathbb{E}\|Z^\varepsilon_t\|_1^q = \sup_{0 \leq t \leq T} \mathbb{E}\left(\frac{\|U^\varepsilon_t - \tilde{U}_t\|_1^2}{\sqrt{\varepsilon}} + \frac{\|\dot{U}^\varepsilon_t - \dot{\tilde{U}}_t\|_1^2}{\sqrt{\varepsilon}}\right)^{q/2} \leq C_T < \infty. \tag{60}
\]

In view of (31), we also have

\[
\sup_{0 \leq t \leq T} \mathbb{E}\|AZ^\varepsilon_t\|_0^q \leq C_T < \infty. \tag{61}
\]

Similarly, we rewrite (20) as

\[
\text{d}\tilde{Z}_t = A\tilde{Z}_t \text{d}t + D_x\bar{\mathcal{F}}(\bar{X}_t).\tilde{Z}_t dt + \Sigma(\bar{X}_t) \text{d}\tilde{W}_t, \tag{62}
\]

where \(\tilde{Z}_t = (\tilde{Z}_t, \dot{\tilde{Z}}_t)^T\), and

\[
\Sigma(x) := \begin{pmatrix} 0 & 0 \\ 0 & \sigma(u) \end{pmatrix}, \quad \text{d}\tilde{W}_t := \begin{pmatrix} 0 \\ \text{d}W_t \end{pmatrix}.
\]

Then \(\Sigma\) is a Hilbert-Schmidt operator satisfying

\[
\frac{1}{2} \Sigma(x) \Sigma^*(x) = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{2}\sigma(u)\sigma^*(u) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \int_H [\delta\mathcal{F}(u, y) \otimes \Psi(u, y)]\mu^u(dy) \end{pmatrix} = \int_H [\delta\mathcal{F}(x, y) \otimes \tilde{\Psi}(x, y)]\mu^x(dy) := \delta\mathcal{F} \otimes \tilde{\Psi}(x), \tag{63}
\]

(see e.g. [11, (1.6)] and [48, (11)]), where \(\tilde{\Psi} := (0, \Psi)^T\) is the solution of the following Poisson equation:

\[
\mathcal{L}_2(x, y)\tilde{\Psi}(x, y) = -\delta\mathcal{F}(x, y). \tag{64}
\]
Moreover, by (21) and (22), we have that

\[ \Sigma(x) \in C^1_b(H, \mathcal{L}(H)), \]  

and for \( h \in \mathcal{H}, x, k \in \mathcal{H}^1 \),

\[ \|D_x^2 \Sigma(x)(h, k)\|_{\mathcal{L}(\mathcal{H})} \leq C_0 \|h\|_0 \|k\|_1. \]  

The unique solution of Eq. (62) can be rewritten in a mild form as follows:

\[ \bar{Z}_t = \int_0^t e^{(t-s)A} D_x \bar{F}(\bar{X}_s). \bar{Z}_s ds + \int_0^t e^{(t-s)A} \Sigma(\bar{X}_s)d\tilde{W}_s. \]

By the Lipschitz continuity of \( \bar{F} \) due to Theorem 3, (65), Gronwall’s inequality and Lemma 2, one can check that for \( q \geq 1 \),

\[ \mathbb{E}\|\bar{Z}_t\|_1^q \leq C_T < \infty. \]  

Combining (28) and (62), the process \((\bar{X}_t, \bar{Z}_t)\) solves the system

\[
\begin{align*}
\frac{d\bar{X}_t}{dt} &= AX_t dt + \bar{F}(\bar{X}_t)dt + dW^1_t, \\
\frac{d\bar{Z}_t}{dt} &= A\bar{Z}_t dt + D_x \bar{F}(\bar{X}_t). \bar{Z}_t dt + \Sigma(\bar{X}_t)d\tilde{W}_t, \\
\bar{X}_0 &= x, \\
\bar{Z}_0 &= 0. 
\end{align*}
\]

Note that the processes \( \bar{X}_t \) and \( \bar{Z}_t \) 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 \( \bar{Z}_t(x, z) \) to denote the process \( \bar{Z}_t \) with initial point \( \bar{Z}_0 = z \in \mathcal{H} \).

As before, approximate the infinite dimensional problem to a finite dimensional one by the Galerkin approximation. Recall that \( X^{n,\varepsilon}_t \) and \( \bar{X}^n_t \) are defined by (40) and (41), respectively. Define

\[ Z^{n,\varepsilon}_t := X^{n,\varepsilon}_t - \bar{X}^n_t. \]

Then we have

\[ dZ^{n,\varepsilon}_t = A Z^{n,\varepsilon}_t dt + \varepsilon^{-1/2} [\bar{F}_n(X^{n,\varepsilon}_t) - \bar{F}_n(\bar{X}^n_t)]dt + \varepsilon^{-1/2} \delta F_n(X^{n,\varepsilon}_t, Y^{n,\varepsilon}_t)dt, \]

where \( \bar{F}_n \) is given by (42), and \( \delta F_n(x, y) := F_n(x, y) - \bar{F}_n(x) \). Let \( \tilde{Z}^n_t \) satisfy the following linear equation:

\[ d\tilde{Z}^n_t = A \tilde{Z}^n_t dt + D_x \bar{F}_n(\bar{X}^n_t). \tilde{Z}^n_t dt + P_n \Sigma(\bar{X}^n_t)d\tilde{W}_t, \]
where \( W_t \) is a cylindrical Wiener process in \( H \), and \( \Sigma(x) \) is defined by (63). As in [46, Lemma 5.4] and [21, Theorem 7, Step 3], one can check that

\[
\lim_{n \to \infty} E\left( \left\| Z_t^\varepsilon - Z_t^{n,\varepsilon} \right\|_1 + \left\| \tilde{Z}_t - \tilde{Z}_t^n \right\|_1 \right) = 0.
\]  

(68)

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

\[
\left| E[\varphi(Z_t^\varepsilon)] - E[\varphi(\tilde{Z}_t)] \right| \leq \left| E[\varphi(Z_t^{n,\varepsilon})] - E[\varphi(Z_t^\varepsilon)] \right| + \left| E[\varphi(Z_t^{n,\varepsilon})] - E[\varphi(Z_t^n)] \right| + \left| E[\varphi(Z_t^n)] - E[\varphi(\tilde{Z}_t)] \right|.
\]  

(69)

According to (68), the first and the last terms on the right-hand of (69) converge to 0 as \( n \to \infty \). Therefore, in order to prove Theorem 2, we only need to show that

\[
\sup_{t \in [0, T]} \left| E[\varphi(Z_t^{n,\varepsilon})] - E[\varphi(Z_t^n)] \right| \leq C_T \varepsilon^{\frac{1}{2}},
\]  

(70)

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.

Given \( T > 0 \), consider the following Cauchy problem on \([0, T] \times H \times H\):

\[
\begin{aligned}
\partial_t \bar{u}(t, x, z) &= \tilde{L}\bar{u}(t, x, z), \quad t \in (0, T], \\
\bar{u}(0, x, z) &= \varphi(z),
\end{aligned}
\]  

(71)

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

\[
\tilde{L} := \tilde{L}_1 + \tilde{L}_3,
\]

with \( \tilde{L}_1 \) given by (59) and \( \tilde{L}_3 \) defined by

\[
\tilde{L}_3\varphi(z) := \tilde{L}_3(x, z)\varphi(z) := \langle Az + D_x \hat{F}(x), z \rangle_H + \frac{1}{2} Tr(D^2\varphi(z) \Sigma(x) \Sigma^*(x)), \quad \forall \varphi \in C^2_\ell(H).
\]

We have the following result.

**Lemma 6** For every \( \varphi \in C^3_b(H) \), there exists a solution to Eq. (71) which is given by

\[
\bar{u}(t, x, z) = E[\varphi(\tilde{Z}_t(x, z))].
\]  

(72)

Moreover, we have

(i) \( \bar{u}(:, x, z) \in C^1_\delta([0, T]) \) and \( \bar{u}(t, :, \cdot) \in C^{1,3}_\delta(H \times H) \);
(ii) for any \( t \in [0, T] \) and \( h \in \mathcal{H}, x, z, k \in \mathcal{H}^1 \),

\[
|D_x^2 \bar{u}(t, x, z)(h, k)| + |D_x D_z \bar{u}(t, x, z)(h, k)| \leq C_0 \|h\|_0 \|k\|_1;
\]

(iii) for any \( t \in [0, T] \) and \( h \in \mathcal{H}, x, z, k, l \in \mathcal{H}^1 \),

\[
|D_x^3 \bar{u}(t, x, z)(h, k, l)| + |D_x D_z^2 \bar{u}(t, x, z)(h, k, l)| \leq C_0 \|h\|_0 \|k\|_1 \|l\|_1;
\]

(iv) for any \( t \in [0, T] \) and \( x, z, h \in \mathcal{H}^1 \),

\[
|\partial_t D_z \bar{u}(t, x, z)| + |\partial_t D_x \bar{u}(t, x, z)| \leq C_0(1 + \|x\|_1^2 + \|z\|_1) \|h\|_1;
\]  

(73)

where \( C_0 > 0 \) is a positive constant.

**Proof** (i)–(iii) Let \( \bar{u}(t, x, z) \) be defined by (72). Since \( \varphi \in C^3_b(\mathcal{H}) \) and \( \bar{Z}_t(x, z) \) satisfy the second equation in (67) (which is a linear one), it is easy to prove that \( \tilde{u}(t, x, \cdot) \in C^3_b(\mathcal{H}) \). To prove the regularity of \( \bar{u} \) with respect to the \( x \)-variable, note that the process \( \zeta_t^h := \nabla_x \bar{X}_t.h \) satisfies

\[
\zeta_t^h = e^{tA}h + \int_0^t e^{(t-s)A} D_x \tilde{F}(\bar{X}_s) \zeta_s^h ds,
\]

and \( \tilde{F} \) is Lipschitz continuous, thus we have

\[
\mathbb{E} \|\zeta_t^h\|_0^2 \leq C_0 \|h\|_0^2.
\]  

(74)

In addition, for any \( h \in \mathcal{H} \) we have

\[
D_x \bar{u}(t, x, z).h = \mathbb{E}[(\varphi'(\bar{Z}_t(x, z)), \eta_t^h)],
\]

where the process \( \eta_t^h := \nabla_x \bar{Z}_t.h \) satisfies

\[
\eta_t^h = \int_0^t e^{(t-s)A} D_x \tilde{F}(\bar{X}_s) \eta_s^h ds + \int_0^t e^{(t-s)A} D_x^2 \tilde{F}(\bar{X}_s) (\bar{Z}_s, \zeta_s^h) ds
\]

\[
+ \int_0^t e^{(t-s)A} D_x \Sigma(\bar{X}_s) \zeta_s^h d\tilde{W}_s,
\]

By the regularity of \( \tilde{F} \), formulas (65), (66) and (74) we have

\[
\mathbb{E} \|\eta_t^h\|_0 \leq C_0 \left( \mathbb{E} \int_0^t \|\zeta_s^h\|_0^2 ds \right)^{1/2} \left( \mathbb{E} \int_0^t \|\bar{Z}_s\|_1^2 ds \right)^{1/2}
\]

\[
+ C_0 \left( \mathbb{E} \int_0^t \|\zeta_s^h\|_0^2 ds \right)^{1/2} \leq C_0 \|h\|_0.
\]
which implies
\[ |D_x \tilde{u}(t, x, z).h| \leq C_0 \|E\|_1 \|\eta_t^j\|_0 \leq C_0 \|h\|_0. \]

Similarly, we can prove the estimates of the higher order derivatives. We omit the details here. The fact that \( \tilde{u} \) solves (71) can be proved in the same way as in [9, Section 4].

To prove (iv), for \( t, z, h \in \mathcal{H}^1 \),
\[ \partial_t D_z \tilde{u}(t, x, z).h = D_{\tilde{z}} \partial_t \tilde{u}(t, x, z).h = D_{\tilde{z}}(\tilde{\mathcal{L}}_1 + \tilde{\mathcal{L}}_3)\tilde{u}(t, x, z).h, \tag{75} \]

On the one hand, we have
\[
D_{\tilde{z}} \tilde{\mathcal{L}}_1 \tilde{u}(t, x, z).h
= D_{\tilde{z}} D_x \tilde{u}(t, x, z). (Ax + \tilde{F}(x), h) + \frac{1}{2} \sum_{n=1}^{\infty} \beta_{1,n} D_{\tilde{z}} D_x^2 \tilde{u}(t, x, z). (Be_n, Be_n, h),
\]

which together with (i), (ii) and (31) yields that
\[ |D_{\tilde{z}} \tilde{\mathcal{L}}_1 \tilde{u}(t, x, z).h| \leq C_1 (1 + \|\|x\|_1\|) \|h\|_1. \tag{76} \]

On the other hand, we have
\[
D_{\tilde{z}} \tilde{\mathcal{L}}_3 \tilde{u}(t, x, z).h
= \langle Ah, D_{\tilde{z}} \tilde{u}(t, x, z) \rangle_{\mathcal{H}} + \langle D_{\tilde{x}} \tilde{F}(x).h, D_{\tilde{z}} \tilde{u}(t, x, z) \rangle_{\mathcal{H}}
+ D_{\tilde{z}}^3 \tilde{u}(t, x, z). (Az + D_x \tilde{F}(x).z).h + \frac{1}{2} \sum_{n=1}^{\infty} D_{\tilde{z}}^3 \tilde{u}(t, x, z). (\Sigma(x)e_n, \Sigma(x)e_n, h).
\]

Thus,
\[ |D_{\tilde{z}} \tilde{\mathcal{L}}_3 \tilde{u}(t, x, z).h| \leq C_2 (1 + \|\|z\|_1 + \|\|x\|_1^2\|) \|h\|_1. \tag{77} \]

Combining (75), (76) and (77), we arrive at
\[ |\partial_t D_{\tilde{z}} \tilde{u}(t, x, z).h| \leq C_3 (1 + \|\|x\|_1^2 + \|\|z\|_1\|) \|h\|_1. \]

Similarly, we have
\[
\partial_t D_x \tilde{u}(t, x, z).h
= D_{\tilde{x}} D_x \tilde{u}(t, x, z). (Ax + \tilde{F}(x), h) + \langle Ah, D_{\tilde{x}} \tilde{F}(x).h, D_x \tilde{u}(t, x, z) \rangle_{\mathcal{H}}
+ \langle D_{\tilde{x}}^2 \tilde{F}(x).(z, h), D_x \tilde{u}(t, x, z) \rangle_{\mathcal{H}} + D_x D_{\tilde{x}} \tilde{u}(t, x, z). (Az + D_x \tilde{F}(x).z, h)
+ \frac{1}{2} \sum_{n=1}^{\infty} \beta_{1,n} D_{\tilde{x}}^3 \tilde{u}(t, x, z). (Be_n, Be_n, h)
\]
\[ + \frac{1}{2} \sum_{n=1}^{\infty} D_x D^2_x \bar{u}(t, x, z). (\Sigma(x)e_n, \Sigma(x)e_n, h) \\
+ \sum_{n=1}^{\infty} D_z^2 \bar{u}(t, x, z). (D_x (\Sigma(x))e_n, \Sigma(x)e_n, h). \]

By the same argument as above, we can obtain

\[ |\partial_t D_x \bar{u}(t, x, z). h| \leq C_4 (1 + \|x\|_1^2 + \|z\|_1^2) \|h\|_1, \]

which completes the proof. \(\square\)

### 5.2 Proof of Theorem 2

Define

\[ \mathcal{L}_3 \varphi(z) := \mathcal{L}_3(x, y, \bar{x}, z) \varphi(z) := \langle \mathcal{A}z, D_z \varphi(z) \rangle_{\mathcal{H}} + \frac{1}{\sqrt{\varepsilon}} \langle \tilde{\mathcal{F}}(x) - \tilde{\mathcal{F}}(\bar{x}), D_z \varphi(z) \rangle_{\mathcal{H}} + \frac{1}{\sqrt{\varepsilon}} \langle \delta \mathcal{F}(x, y), D_z \varphi(z) \rangle_{\mathcal{H}}, \quad \forall \varphi \in C^1_{\ell}(\mathcal{H}). \tag{78} \]

Given a function satisfying the centering condition

\[ \int_{\mathcal{H}} \varphi(t, x, y, z) \mu^x(dy) = 0, \quad \forall t > 0, x, z \in \mathcal{H}, \tag{79} \]

and the following conditions hold:

\textbf{(H): } \varphi(\cdot, x, y, \cdot) \in C^{1,2}_{b}([0, T] \times \mathcal{H}), \varphi(t, \cdot, \cdot, z) \in C^{1,\eta}_{\ell}(\mathcal{H} \times H), \text{ and for any } t \in [0, T] \text{ and } x, z, k \in \mathcal{H}, h, y \in H,

\[ |D^2_x \varphi(t, x, y, z)(h, k)| \leq C_0 (1 + \|x\|_1 + \|y\|) \|h\|_0 \|k\|_1. \]

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

\[ \mathcal{L}_2(x, y) \psi(t, x, y, z) = -\varphi(t, x, y, z), \tag{80} \]

where \(t, x, z\) are regarded as parameters. Define

\[ \overline{\delta \mathcal{F}} \cdot \nabla_z \psi(t, x, z) := \int_{\mathcal{H}} \nabla_z \psi(t, x, y, z) \delta \mathcal{F}(x, y) \mu^y(dy). \]

We first establish the following weak fluctuation estimates for an appropriate integral functional of \((X^s, Y^s, Z^s)\) over the time interval \([0, t]\), which will play an important role in the proof of (70).
Lemma 7 (Weak fluctuations estimates) Let \( T, \eta > 0, x, z \in \mathcal{H}_1 \) and \( y \in H^1 \). Assume that (25)–(27) hold. Then for any \( t \in [0, T] \), \( \phi(t, x, y, z) \) satisfying the centering condition (79), the condition (H) and

\[
|\partial_t \phi(t, x, y, z)| \leq C_0 \left( 1 + \|x\|_1^2 + \|z\|_1 \right) \left( 1 + \|x\|_1 + \|y\| \right),
\]

we have

\[
\left| \mathbb{E} \left( \frac{1}{\sqrt{\varepsilon}} \int_0^t \phi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds - \int_0^t \delta \mathcal{F} \cdot \nabla \psi(s, X_s^\varepsilon, Z_s^\varepsilon) ds \right) \right| \leq C_T \varepsilon \frac{1}{2},
\]

where \( C_T > 0 \) is a constant independent of \( \varepsilon, \eta \) and \( n \). In particular, we have

\[
\left| \mathbb{E} \left( \int_0^t \phi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \right) \right| \leq C_T \varepsilon \frac{1}{2}.
\]

**Proof** The proof will be divided into two steps.

**Step 1** We first prove estimate (83). Applying Itô’s formula to \( \psi(t, X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon) \) and taking expectation, we have

\[
\mathbb{E}[\psi(t, X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon)] = \psi(0, x, y, 0) + \mathbb{E}\left( \int_0^t (\partial_s + \mathcal{L}_1 + \mathcal{L}_3^\varepsilon) \psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \right)
\]

\[
+ \frac{1}{\varepsilon} \mathbb{E}\left( \int_0^t \mathcal{L}_2 \psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \right),
\]

where \( \mathcal{L}_1 \) and \( \mathcal{L}_3^\varepsilon \) are defined by (51) and (78), respectively. Combining this with (80), we obtain

\[
\mathbb{E}\left( \int_0^t \phi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \right) = \varepsilon \mathbb{E}[\psi(0, x, y, 0) - \psi(t, X_t^\varepsilon, Y_t^\varepsilon, Z_t^\varepsilon)]
\]

\[
+ \varepsilon \mathbb{E}\left( \int_0^t \partial_s \psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \right) + \varepsilon \mathbb{E}\left( \int_0^t \mathcal{L}_1 \psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \right)
\]

\[
+ \varepsilon \mathbb{E}\left( \int_0^t \mathcal{L}_3^\varepsilon \psi(s, X_s^\varepsilon, Y_s^\varepsilon, Z_s^\varepsilon) ds \right) =: \sum_{i=1}^4 \mathcal{D}_i(t, \varepsilon).
\]

By Theorem 3 and Lemma 1, we have

\[
|\mathcal{D}_1(t, \varepsilon)| \leq C_1 \varepsilon \mathbb{E} \left( 1 + \|X_t^\varepsilon\|_1 + \|Y_t^\varepsilon\| \right) \leq C_1 \varepsilon.
\]

For the second term, by using Theorem 3, condition (81), Lemma 1 and (60), we get

\[
|\mathcal{D}_2(t, \varepsilon)| \leq C_2 \left( \int_0^t \mathbb{E} \left( 1 + \|X_s^\varepsilon\|_1^4 + \|Y_s^\varepsilon\|^2 + \|Z_s^\varepsilon\|^2 \right) ds \right) \leq C_2 \varepsilon.
\]
To treat the third term, since for each \( \phi(t, x, y, z) \) satisfying (79) and (H), by Theorem 3, we have \( \psi(t, \cdot, \cdot, \cdot) \in C_{0,2}^{1,0}(\mathcal{H} \times \mathcal{H}), \psi(\cdot, x, y, \cdot) \in C_{0,1}^{1,0}(0, T) \times \mathcal{H} \) and for any \( x, k \in \mathcal{H}, h \in \mathcal{H}, y \in \mathcal{H} \),

\[
|D^2_{x} \psi(t, x, y, z)(h, k)| \leq C_0(1 + ||x||_{1} + ||y||)||h||_{0}||k||_{1},
\]

and thus

\[
|\mathcal{L}_1 \psi(t, X^e_t, Y^e_t, Z^e_t)| \leq |\langle AX^e_t + \mathcal{F}(X^e_t, Y^e_t, D_x \psi(t, X^e_t, Y^e_t, Z^e_t)) \rangle_{\mathcal{H}}| + \frac{1}{2} Tr((B Q^1_1)(B Q^1_1)^{1}) \|D^2_{x} \psi(t, X^e_t, Y^e_t, Z^e_t)\|_{\mathcal{L}(\mathcal{H} \times \mathcal{H})} \leq C_3(1 + ||AX^e_t||_{0}^{2} + ||X^e_t||_{1}^{2} + ||Y^e_t||^{2}).
\]

As a result of Lemma 1 and (31), we deduce that

\[
|\mathcal{Q}_3(t, \varepsilon)| \leq C_3 \varepsilon \mathbb{E} \left( \int_{0}^{t} \left( 1 + ||AX^e_s||_{0}^{2} + ||X^e_s||_{1}^{2} + ||Y^e_s||^{2} \right) ds \right) \leq C_3 \varepsilon.
\]

For the last term, we have

\[
\mathcal{Q}_4(t, \varepsilon) = \varepsilon \mathbb{E} \left( \int_{0}^{t} \langle AZ^e_s, D_x \psi(s, X^e_s, Y^e_s, Z^e_s) \rangle_{\mathcal{H}} ds \right) + \sqrt{\varepsilon} \mathbb{E} \left( \int_{0}^{t} \langle \tilde{\mathcal{F}}(X^e_s) - \mathcal{F}(X^e_s), D_x \psi(s, X^e_s, Y^e_s, Z^e_s) \rangle_{\mathcal{H}} ds \right) + \sqrt{\varepsilon} \mathbb{E} \left( \int_{0}^{t} \langle \delta \mathcal{F}(X^e_s, Y^e_s), D_x \psi(s, X^e_s, Y^e_s, Z^e_s) \rangle_{\mathcal{H}} ds \right).
\]

It follows from Lemma 1 again and (61) that

\[
|\mathcal{Q}_4(t, \varepsilon)| \leq C_4 \sqrt{\varepsilon} \mathbb{E} \left( \int_{0}^{t} \left( 1 + ||AZ^e_s||_{0}^{2} + ||X^e_s||_{1}^{2} + ||Y^e_s||^{2} \right) ds \right) \leq C_4 \sqrt{\varepsilon}.
\]

Combining the above inequalities with (84), we get the desired result.

**Step 2** We proceed to prove estimate (82). Note that the three terms \( \mathcal{Q}_i(t, \varepsilon) (i = 1, 2, 3) \) on the right hand side of (84) are of order \( \varepsilon \) while \( \mathcal{Q}_4(t, \varepsilon) \) is of order \( \sqrt{\varepsilon} \). By following exactly the same arguments as in the proof of Step 1, we get that

\[
\left| \mathbb{E} \left( \frac{1}{\sqrt{\varepsilon}} \int_{0}^{t} \phi(s, X^e_s, Y^e_s, Z^e_s) ds - \int_{0}^{t} \delta \mathcal{F} \cdot \nabla_x \psi(s, X^e_s, Z^e_s) ds \right) \right| \leq C_0 \sqrt{\varepsilon} + \mathbb{E} \left( \sqrt{\varepsilon} \int_{0}^{t} \mathcal{L}^e_3 \psi(s, X^e_s, Y^e_s, Z^e_s) ds - \int_{0}^{t} \delta \mathcal{F} \cdot \nabla_x \psi(s, X^e_s, Z^e_s) ds \right).
\]
For the last term, by definition (78) we have
\[
\left| \mathbb{E} \left( \sqrt{\varepsilon} \int_0^t L_3^e \psi(s, X_s^e, Y_s^e, Z_s^e) ds - \int_0^t \delta F \cdot \nabla \psi(s, X_s^e, Z_s^e) ds \right) \right| \\
\leq \sqrt{\varepsilon} \left| \mathbb{E} \left( \int_0^t \left( A Z_s^e, D \psi(s, X_s^e, Y_s^e, Z_s^e) \right) \mathcal{H} ds \right) \right| \\
+ \left| \mathbb{E} \left( \int_0^t \left( \tilde{F}(X_s^e) - \tilde{F}(\tilde{X}_s), D \psi(s, X_s^e, Y_s^e, Z_s^e) \right) \mathcal{H} ds \right) \right| \\
+ \left| \mathbb{E} \left( \int_0^t \left( \langle \delta \mathcal{F}(X_s^e, Y_s^e), D \psi(s, X_s^e, Y_s^e, Z_s^e) \rangle - \delta \mathcal{F} \cdot \nabla \psi(s, X_s^e, Z_s^e) \right) ds \right) \right| \\
=: \sum_{i=1}^3 T_i(t, \varepsilon).
\]

Using Lemma 1 and (61), we get
\[
T_1(t, \varepsilon) \leq C_1 \sqrt{\varepsilon} \mathbb{E} \left( \int_0^t \| A Z_s^e \|_0 (1 + \| X_s^e \|_1 + \| Y_s^e \|) ds \right) \leq C_1 \sqrt{\varepsilon}.
\]

According to Hölder’s inequality, Lemma 1 and Theorem 1, we have
\[
T_2(t, \varepsilon) \leq C_2 \int_0^t \left( \mathbb{E} \| X_s^e - \tilde{X}_s \|_1^2 \right)^{1/2} \left( 1 + \mathbb{E} \| X_s^e \|_1^2 + \mathbb{E} \| Y_s^e \|_1^2 \right)^{1/2} ds \\
\leq C_2 \sqrt{\varepsilon}.
\]

In view of $T_3(t, \varepsilon)$, note that the function
\[
\tilde{\phi}(t, x, y, z) := \langle \delta \mathcal{F}(x, y), D \psi(t, x, y, z) \rangle_1 - \delta \mathcal{F} \cdot \nabla \psi(t, x, z)
\]
satisfies the centering condition (79) and condition (81). Thus, using (83) directly, we obtain
\[
T_3(t, \varepsilon) \leq C_4 \sqrt{\varepsilon},
\]
which completes the proof.

Now, we are in the position to give:

**Proof of estimate (70)** Fix $T > 0$. For any $t \in [0, T]$ and $x, z \in \mathcal{H}^1$, let
\[
\tilde{u}(t, x, z) = \tilde{u}(T - t, x, z).
\]

It is easy to check that
\[
\tilde{u}(0, x, 0) = \tilde{u}(T, x, 0) = \mathbb{E} [\varphi(\tilde{Z}_T)] \quad \text{and} \quad \tilde{u}(T, x, z) = \tilde{u}(0, x, z) = \varphi(z).
\]
Applying Itô’s formula, by (71) we have
\[
\mathbb{E}[\varphi(Z^e_{T})] - \mathbb{E}[\varphi(\tilde{Z}_{T})] = \mathbb{E}[\tilde{u}(T, X^e_T, Z^e_T) - \tilde{u}(0, x, 0)]
\]
\[
= \mathbb{E}\left( \int_0^T (\partial_t + L_1 + L_3^e) \tilde{u}(t, X^e_t, Z^e_t) dt \right)
\]
\[
= \mathbb{E}\left( \int_0^T (L_1 - \tilde{L}) \tilde{u}(t, X^e_t, Z^e_t) dt \right) + \mathbb{E}\left( \int_0^T (L_3^e - \tilde{L}_3) \tilde{u}(t, X^e_t, Z^e_t) dt \right)
\]
\[
+ \mathbb{E}\left( \int_0^T \left( \frac{\tilde{F}(X^e_t) - \tilde{F}(X^e_t)}{\sqrt{\varepsilon}} - D_x \tilde{F}(X^e_t, Z^e_t) D_z \tilde{u}(t, X^e_t, Z^e_t) \right) \right) dt
\]
\[
+ \mathbb{E}\left( \frac{1}{\sqrt{\varepsilon}} \int_0^T \left( \tilde{F}(X^e_t, Y^e_t) - \tilde{F}(X^e_t) - D_y \tilde{F}(X^e_t, Z^e_t) \right) dt \right)
\]
\[
- \frac{1}{2} \mathbb{E}\left( \int_0^T tr(D^2_x \tilde{u}(t, X^e_t, Z^e_t) \Sigma(X^e_t) \Sigma(X^e_t)^*) dt \right) := \sum_{i=1}^3 M_i(T, \varepsilon).
\]

For the first term, recall that \( \tilde{\Psi} \) solves the Poisson equation (64) and define
\[
\psi(t, x, y, z) := \langle \tilde{\Psi}(x, y), D_x \tilde{u}(t, x, z) \rangle_{\mathcal{H}}.
\]

Since \( L_2 \) is an operator with respect to the \( y \)-variable, one can check that \( \psi \) solves the following Poisson equation:
\[
L_2(x, y) \psi(t, x, y, z) = -\langle \delta \tilde{F}(x, y), D_x \tilde{u}(t, x, z) \rangle_{\mathcal{H}} = -\phi(t, x, y, z).
\]

It is obvious that \( \phi \) satisfies the centering condition (79). Furthermore, by the regularity of \( \tilde{F} \) due to Theorem 3, Lemma 6 (ii), (iii) and (73), it is easy to check that \( \phi \) satisfies conditions (H) and (81). Thus, it follows from (83) directly that
\[
|M_1(T, \varepsilon)| \leq C_1 \sqrt{\varepsilon}.
\]

To control the second term, by the mean value theorem, Hölder’s inequality, Lemma 6, Theorem 1 and (60) we deduce that for \( \vartheta \in (0, 1) \),
\[
|M_2(T, \varepsilon)| \leq \mathbb{E}\left( \int_0^T \left| \left[ D_x \tilde{F}(X^e_t) + \vartheta (X^e_t - \tilde{X}_t) \right] \right| dt \right)
\]
\[
\leq C_2 \int_0^T (\mathbb{E}[||X^e_t - \tilde{X}_t||^2_1])^{1/2} (\mathbb{E}[||Z^e_t||^2_1])^{1/2} dt \leq C_2 \sqrt{\varepsilon}.
\]
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 (79) and conditions (H) and (81). Furthermore, by the definition of \( \Sigma \) in (63), we have
\[
\delta \mathcal{F} \cdot \nabla_z \hat{\psi}(t, x, z) = \int_{\mathcal{H}} D_z \hat{\psi}(t, x, y, z) \delta \mathcal{F}(x, y) \mu^x(dy)
= \int_{\mathcal{H}} D_z^2 \tilde{u}(t, x, z) (\tilde{\Psi}(x, y), \delta \mathcal{F}(x, y)) \mu^x(dy) = \frac{1}{2} \text{Tr} (D_z^2 \tilde{u}(t, x, z) \Sigma(x) \Sigma^*(x)).
\]

Thus, it follows by (82) directly that
\[ |\mathcal{N}_3(T, \varepsilon)| \leq C_3 \sqrt{\varepsilon}. \]

Combining the above computations, we get the desired result. \( \square \)

**Acknowledgements** The authors would like to thank the referees for a careful reading of the first submitted version and for a number of useful comments and suggestions for improving the paper.

**Data availability** Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

**References**

1. Bakhtin, V., Kifer, Y.: Diffusion approximation for slow motion in fully coupled averaging. Probab. Theory Rel. Fields 129, 157–181 (2004)
2. Bao, J., Yin, G., Yuan, C.: Two-time-scale stochastic partial differential equations driven by \( \alpha \)-stable noises: averaging principles. Bernoulli 23, 645–669 (2017)
3. Barbu, V., Da Prato, G.: The stochastic nonlinear damped wave equation. Appl. Math. Optim. 46, 125–141 (2002)
4. Barbu, V., Da Prato, G., Tubaro, L.: Stochastic wave equations with dissipative damping. Stoch. Proc. Appl. 117, 1001–1013 (2007)
5. Bogoliubov, N.N., Mitropolsky, Y.A.: Asymptotic Methods in the Theory of Non-linear Oscillations. Gordon and Breach Science Publishers, New York (1961)
6. Bréhier, C.E.: Strong and weak orders in averaging for SPDEs. Stoch. Process. Appl. 122, 2553–2593 (2012)
7. Bréhier, C.E.: Orders of convergence in the averaging principle for SPDEs: the case of a stochastically forced slow component. Stoch. Proc. Appl. 130, 3325–3368 (2020)
8. Bréhier, C.E.: Analysis of an HMM time-discretization scheme for a system of stochastic PDEs. SIAM J. Numer. Anal. 51, 1185–1210 (2013)
9. Bréhier, C.E., Debussche, A.: Kolmogorov equations and weak order analysis for SPDEs with nonlinear diffusion coefficient. J. Math. Pures Appl. 119, 193–254 (2018)
10. Cerrai, S.: A Khasminskii type averaging principle for stochastic reaction–diffusion equations. Ann. Appl. Probab. 19, 899–948 (2009)
11. Cerrai, S.: Normal deviations from the averaged motion for some reaction–diffusion equations with fast oscillating perturbation. J. Math. Pures Appl. 91, 614–647 (2009)
12. Cerrai, S.: Averaging principle for systems of reaction–diffusion equations with polynomial nonlinearities perturbed by multiplicative noise. SIAM J. Math. Anal 43, 2482–2518 (2011)
13. Cerrai, S., Freidlin, M.: Averaging principle for stochastic reaction–diffusion equations. Probab. Theory Related Fields 144, 137–177 (2009)
14. Cerrai, S., Glatt-Holtz, N.: On the convergence of stationary solutions in the Smoluchowski–Kramers approximation of infinite dimensional systems. J. Funct. Anal. 278, 108421 (2020)
15. Cerrai, S., Lunardi, A.: Averaging principle for non-autonomous slow–fast systems of stochastic reaction–diffusion equations in Hilbert spaces perturbed by a bounded measurable drift. Ann. Probab. 41(5), 3306–3344 (2013)
16. Dalang, R., Khoshnevisan, D., Mueller, C., Nualart, D., Xiao, Y.: A Minicourse on Stochastic Partial Differential Equations. Lecture Notes in Mathematics, Springer, Berlin (2009)
17. Debussche, A., Vovelle, J.: Diffusion limit for a stochastic kinetic problem. Commun. Pure Appl. Anal. 11, 2305–2326 (2012)
18. Einen, W., Liu, D., Vanden-Eijnden, E.: Analysis of multiscale methods for stochastic differential equations. Comm. Pure Appl. Math. 58, 1544–1585 (2005)
19. Guérin, Y.: Fluctuations around a homogenised semilinear random PDE. Arch. Ration Mech. Anal. 239(1), 151–217 (2021)
20. Kelly, D., Melbourne, I.: Homogenization for deterministic fast–slow systems with multidimensional multiplicative noise. J. Funct. Anal. 272, 4063–4102 (2017)
21. Khasminskii, R.Z.: On stochastic processes defined by differential equations with a small parameter. Theory Probab. Appl. 11, 211–228 (1966)
22. Khasminskii, R.Z., Yin, G.: On averaging principles: an asymptotic expansion approach. SIAM J. Math. Anal. 35, 1534–1560 (2004)
23. Leung, A.W.: Asymptotically stable invariant manifold for coupled nonlinear parabolic–hyperbolic partial differential equations. J. Differ. Equ. 212(1), 85–113 (2005)
24. Liu, D.: Strong convergence of principle of averaging for multiscale stochastic dynamical systems. Commun. Math. Sci. 8(4), 999–1020 (2010)
25. Liu, W., Röckner, M.: Stochastic Partial Differential Equations: An Introduction. Universitext, Springer (2015)
26. Liu, W., Röckner, M.: Averaging principle for slow–fast stochastic differential equations with time dependent locally Lipschitz coefficients. J. Differ. Equ. 268(6), 2910–2948 (2020)
40. Masiero, F., Priola, E.: Well-posedness of semilinear stochastic wave equations with Hölder continuous coefficients. J. Differ. Equ. 263, 1773–1812 (2017)
41. Pardoux, E., Veretennikov, AYu.: On the Poisson equation and diffusion approximation. I. Ann. Probab. 29, 1061–1085 (2001)
42. Pardoux, E., Veretennikov, AYu.: On the Poisson equation and diffusion approximation 2. Ann. Probab. 31, 1166–1192 (2003)
43. Rivera, J.E.M., Racke, R.: Smoothing properties, decay and global existence of solution to nonlinear coupled systems of thermoelasticity type. SIAM J. Math. Anal. 26, 1547–1563 (1995)
44. Röckner, M., Xie, L.: Diffusion approximation for fully coupled stochastic differential equations. Ann. Probab. 49(3), 1205–1236 (2021)
45. Röckner, M., Xie, L.: Averaging principle and normal deviations for multiscale stochastic systems. Commun. Math. Phys. 383, 1889–1937 (2021)
46. Röckner, M., Xie, L., Yang, L.: Asymptotic behavior of multiscale stochastic partial differential equations. arXiv:2010.14897.pdf
47. Veretennikov, AYu.: On the averaging principle for systems of stochastic differential equations. Math. USSR Sborn. 69, 271–284 (1991)
48. Wang, W., Roberts, A.J.: Average and deviation for slow–fast stochastic partial differential equations. J. Differ. Equ. 253, 1265–1286 (2012)
49. Zhang, X., Zuazua, E.: Long-time behavior of a coupled heat-wave system arising in fluid–structure interaction. Arch. Ration. Mech. Anal. 184, 49–120 (2007)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.