Bernoulli **23**(1), 2017, 645–669
DOI: 10.3150/14-BEJ677

Two-time-scale stochastic partial differential equations driven by α-stable noises: Averaging principles

JIANHAI BAO[1], GEORGE YIN[2] and CHENGGUI YUAN[3]

[1]*Department of Mathematics, Central South University, Changsha, Hunan, 410083, P.R. China.*
E-mail: jianhaibao13@gmail.com
[2]*Department of Mathematics, Wayne State University, Detroit, MI 48202, USA.*
E-mail: gyin@math.wayne.edu
[3]*Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, UK.*
E-mail: C.Yuan@swansea.ac.uk

This paper focuses on stochastic partial differential equations (SPDEs) under two-time-scale formulation. Distinct from the work in the existing literature, the systems are driven by α-stable processes with $\alpha \in (1, 2)$. In addition, the SPDEs are either modulated by a continuous-time Markov chain with a finite state space or have an addition fast jump component. The inclusion of the Markov chain is for the needs of treating random environment, whereas the addition of the fast jump process enables the consideration of discontinuity in the sample paths of the fast processes. Assuming either a fast changing Markov switching or an additional fast-varying jump process, this work aims to obtain the averaging principles for such systems. There are several distinct difficulties. First, the noise is not square integrable. Second, in our setup, for the underlying SPDE, there is only a unique mild solution and as a result, there is only mild Itô's formula that can be used. Moreover, another new aspect is the addition of the fast regime switching and the addition of the fast varying jump processes in the formulation, which enlarges the applicability of the underlying systems. To overcome these difficulties, a semigroup approach is taken. Under suitable conditions, it is proved that the pth moment convergence takes place with $p \in (1, \alpha)$, which is stronger than the usual weak convergence approaches.

Keywords: α-stable process; averaging principle; invariant measure; stochastic partial differential equation; strong convergence

1. Introduction

Averaging principles for stochastic differential equations (SDEs) have been studied extensively, for example, in Liu and Vanden-Eijnden [10], Freidlin and Wentzell [11], Khasminskii [20], Yin and Zhang [34]. Recently, averaging principles for stochastic partial differential equations (SPDEs) have also drawn much attention; see, for example, Kuksin and Piatnitski [23] and Maslowski *et al.* [27]. In particular, Blömker *et al.* [4] derived averaging results with explicit error bounds for SPDEs with quadratic nonlinearities, where the limiting system is an SDE; Cerrai and Freidlin [7] investigated the weak convergence for two-time-scale stochastic reaction–diffusion equations with additive noise by using an approach based on Kolmogorov equations and martingale solutions of stochastic equations; Cerrai [6] generalized Cerrai and Freidlin [7] to the case of slow–fast reaction–diffusion equations driven by multiplicative noise, where the reaction terms appear in both equations; Bréhier [5] gave the strong and weak orders in averaging for stochastic evolution equation of parabolic type with slow and fast time scales. For the finite-dimensional jump–diffusion case, we refer to Givon [14].

In view of the development on the aforementioned singularly perturbed SPDEs, the noise processes considered to date are mainly square integrable processes. However, such requirement rules out the interesting α-stable processes. It is well known that both Wiener processes and Poisson-jump processes have finite moments of any order, whereas an α-stable process only has finite pth moment for $p \in (0, \alpha)$. Stochastic equations driven by α-stable processes have proven to have numerous applications in physics because such processes can be used to model systems with heavy tails. As a result, such processes have received increasing attentions recently. For example, Priola and Zabczyk [30] gave a proper starting point on the investigation of structural properties of SPDEs driven by an additive cylindrical stable noise; Dong *et al.* [9] studied ergodicity of stochastic Burgers equations driven by $\alpha/2$-subordinated cylindrical Brownian motions with $\alpha \in (1, 2)$. For finite-dimensional SDEs driven by α-stable noises, Wang [33] derived gradient estimate for linear SDEs, Zhang [36] established the Bismut–Elworthy–Li derivative formula for nonlinear SDEs, and Ouyang [28] established Harnack inequalities for Ornstein–Uhlenbeck processes by the sharp estimates of density function for rotationally invariant symmetric α-stable Lévy processes. Nevertheless, two-time-scale formulation for stochastic processes driven by α-stable processes have not yet been considered to date to the best of our knowledge.

Motivated by the previous works, in this paper we develop averaging principles for two-time-scale SPDEs driven by α-stable noises that admit unique mild solutions. The time-scale separation is given by introducing a small parameter $\varepsilon > 0$. For the case of mean-square integrable noise, the Itô formula plays an important role in the error analysis between the slow component and the averaging systems; see, for example, Givon [14], Fu and Duan [12] and Fu and Liu [13]. It has been noted that when the diffusion operators in Fu and Duan [12] and Fu and Liu [13] are Hilbert–Schmidt, the mild solution is indeed a strong solution. Nevertheless, in our case, only *mild Itô's formula* (see, e.g., Da Prato *et al.* [8], Theorem 1) is available since the stochastic systems considered only admit mild solutions, not strong solutions. Moreover, the technique adopted in Bréhier [5], Lemma 3.1, which is a key ingredient in discussing averaging principle, does not work for the case of SPDEs driven by α-stable noises either, although the mild solution is treated there. In our study, in addition to the SPDEs, we assume that the systems are modulated by a continuous-time Markov chain. This Markov chain has a finite state space resulting in a system of stochastic differential equations switching back and forth according to the state of the Markov chain. The Markov chain can be used to model discrete events that are not representable otherwise. It is by now widely recognized that such regime-switching formulation is an effective way of modeling many practical situations in which random environment and other random factors have to be taken into consideration. Perhaps, one of the first efforts in modeling random environment using a finite-state Markov chain can be traced back to Griego and Hersh [15] (see also the extended survey in Hersh [17], where multiple time scale was also used). Much of the recent modeling and analysis effort stems from the work of Hamilton and Susmel [16], who revealed the feature of the so-called regime-switching systems under which the dynamics of the systems can be quite different under different regimes. Their idea stimulated much of the subsequent study. For example, in the simplest setting, the successfully used regime-switching models in financial market portraits the random environment with two states bull and bear markets, whose volatilities are drastically different.

Our study is divided into two parts. In the first part, we assume that the switching process is subject to fast variation, either within a weakly irreducible class or within a number of nearly decomposable weakly irreducible classes (see Yin and Zhang [34], Chapter 4). The idea is that the original system subject to fast switching is more complex, but the limit system is much simpler. For many applications, it will be desirable to find the structure of the limit system leading substantial reduction of computational complexity for such tasks as control and optimization etc. We show that under suitable conditions, a limit process that is a solution of either an SPDE or an SPDE with switching is obtained. The key is that in the limit, the coefficients are averaged out with respect to the stationary measure of the switching processes. In the second part, we assume that there is an additional fast-varying random process. Although the process is fast varying, it does not blow up, but rather has an invariant measure. The ergodicity of the fast process helps us to get a limit process that is a solution of the SPDEs with the coefficients being averaged out with respect to the stationary distribution of the fast-varying process.

To summarize, there are several distinct difficulties in our problems. First, the noise is not square integrable. Second, the underlying SPDE admits only a unique mild solution and as a result, there is only mild Itô's formula that can be used. Moreover, another new aspect is the addition of the fast regime switching and the addition of the fast varying jump processes in the formulation, which enlarges the applicability of the underlying systems. To overcome these difficulties, using the mild solutions, a semigroup approach is taken. Under suitable conditions, it is proved that the pth moment convergence takes place with $p \in (1, \alpha)$, which is stronger than the usual weak convergence approaches. We thus term such a convergence as strong convergence.

The rest of the paper is organized as follows. In Section 2, we obtain not only averaging principles for SPDEs with two-time-scale Markov switching with a single weakly recurrent class but also for the case of two-time-scale Markov switching with multiple weakly irreducible classes. In Section 3, we demonstrate the strong convergence for SPDEs with an additional fast-varying random process driven by cylindrical stable processes.

2. SPDEs with two-time-scale Markov switching

We first recall some basics on stable processes. A real-valued random variable η is said to have a stable distribution with stability index $\alpha \in (0, 2)$, scale parameter $\sigma \in (0, \infty)$, skewness parameter $\beta \in [-1, 1]$, and location parameter $\mu \in (-\infty, \infty)$, if its characteristic function has the form:

$$\phi_\eta(u) = \mathbb{E}\exp(iu\eta) = \exp\bigl\{-\sigma^\alpha |u|^\alpha \bigl(1 - i\beta \operatorname{sgn}(u)\Phi\bigr) + i\mu u\bigr\}, \qquad u \in \mathbb{R},$$

where $\Phi = \tan(\pi\alpha/2)$ for $\alpha \neq 1$ and $\Phi = -(2/\pi)\log|u|$ for $\alpha = 1$. Note that the monograph Samorodnitsky and Taqqu [31], pages 2–10, also gives three other equivalent definitions of a stable distribution. We denote the family of stable distributions by $S_\alpha(\sigma, \beta, \mu)$ and write $X \sim S_\alpha(\sigma, \beta, \mu)$ to indicate that X has the stable distribution $S_\alpha(\sigma, \beta, \mu)$. A random variable $X \sim S_\alpha(\sigma, \beta, \mu)$ is said to be strictly stable if $\mu = 0$ for $\alpha \neq 1$ (if $\beta = 0$ for $\alpha = 1$), symmetric if $\beta = \mu = 0$, and standard (normalized) if $\beta = \mu = 0$ and $\sigma = 1$. Let $(H, \langle\cdot,\cdot\rangle, |\cdot|_H)$ be a real separable Hilbert space. Let $L=\{L(t)\}_{t\geq 0}$ and $Z=\{Z(t)\}_{t\geq 0}$ be a cylindrical α-stable process and β-stable process defined by the orthogonal expansion, respectively,

$$L(t):=\sum_{k=1}^{\infty}\beta_k L_k(t)e_k \quad \text{and} \quad Z(t):=\sum_{k=1}^{\infty}q_k Z_k(t)e_k, \tag{2.1}$$

where $\{e_k\}_{k\geq 1}$ is an orthonormal basis of H, $\{L_k(t)\}_{k\geq 1}$ and $\{Z_k(t)\}_{k\geq 1}$ are sequences of i.i.d. (independent and identically distributed) real-valued symmetric α-stable processes and β-stable processes defined on the stochastic basis $(\Omega,\mathscr{F},\{\mathscr{F}_t\}_{t\geq 0},\mathbb{P})$, respectively, and $\beta_k, q_k>0$ for each $k\geq 1$. $\|\cdot\|$ stands for the operator norm, and $\mathscr{L}(\xi)$ means the law of an H-valued random variable ξ. Throughout this paper, we assume that $\alpha,\beta\in(1,2]$. Generic constants will be denoted by c, and we use the shorthand notation $a\lesssim b$ to mean $a\leq cb$. If the constant c depends on a parameter p, we shall also write c_p and $a\lesssim_p b$.

2.1. Two-time-scale Markov switching with a single weakly irreducible class

Hybrid systems driven by continuous-time Markov chains have been used to model many practical scenarios in which abrupt changes may be experienced in the structure and parameters caused by phenomena such as component failures or repairs; see Sethi and Zhang [32], Remark 5.1, page 94, for discussions on the modeling of such a system and related optimal control problems. For finite-dimensional cases, there is extensive literature on such topic, for example, Mao and Yuan [25], Mariton [26], Yin and Zhu [35] and the references therein. As an infinite-dimensional example, we consider a one-dimensional rod of length π whose ends are maintained at 0° and whose sides are insulated. Assume that there is an exothermic reaction taking place inside the rod with heat being produced proportionally to the temperature. The temperature u in the rod may be modeled by

$$\begin{cases}\dfrac{\partial u}{\partial t}=\dfrac{\partial^2 u}{\partial x^2}+cu, & t>0, x\in(0,\pi),\\ u(t,0)=u(t,\pi)=0, & u(0,x)=u_0(x),\end{cases} \tag{2.2}$$

where $u=u(t,x)$ and c is a constant dependent on the rate of reaction. In lieu of assuming the system to be in a fixed configuration, let system (2.2) switch from one mode to another in a random way when it experiences abrupt changes in its structure and parameters caused by phenomena such as component failures or repairs, changing subsystem interconnections, or abrupt environmental disturbances. The system under regime switching could be described by the following random model

$$\begin{cases}\dfrac{\partial u}{\partial t}=\dfrac{\partial^2 u}{\partial x^2}+c\big(r(t)\big)u, & t>0, x\in(0,\pi),\\ u(t,0)=u(t,\pi)=0, & u(0,x)=u_0(x), r(0)=r_0,\end{cases}$$

where $r(t)$ is a continuous-time Markov chain with a finite state space $\mathbb{S}$ and $c:\mathbb{S}\to\mathbb{R}$. For further details, we refer to, for example, Anabtawi [1] and Bao *et al.* [3].

With the motivation above, assuming that $r^\varepsilon(t)$ is a continuous-time Markov chain with a finite state space $\mathbb{S} := \{1, \dots, n\}$, we consider the following SPDE

$$\mathrm{d}X^\varepsilon(t) = \{AX^\varepsilon(t) + b(X^\varepsilon(t), r^\varepsilon(t))\}\,\mathrm{d}t + \mathrm{d}L(t), \qquad t > 0 \tag{2.3}$$

with the initial values $X^\varepsilon(0) = x \in H$ and $r^\varepsilon(0) = r_0 \in \mathbb{S}$.

In (2.3), for any $\varepsilon \in (0, 1)$, $r^\varepsilon(t)$ is a Markov chain with a finite state space $\mathbb{S}$ and generator

$$\mathbb{Q}^\varepsilon := \frac{\widetilde{\mathbb{Q}}}{\varepsilon} + \widehat{\mathbb{Q}},$$

where $\widetilde{\mathbb{Q}}$ and $\widehat{\mathbb{Q}}$ are suitable generators of some Markov chains such that $\widetilde{\mathbb{Q}}/\varepsilon$ and $\widehat{\mathbb{Q}}$ represent the fast-varying and the slow-changing parts, respectively. In what follows, we further assume that $\widetilde{\mathbb{Q}}$ is weakly irreducible. That is, the system of equations

$$\begin{cases} \nu\widetilde{\mathbb{Q}} = 0, \\ \sum_{i=1}^{n} \nu_i = 1, \end{cases} \tag{2.4}$$

has a unique solution satisfying $\nu_i \geq 0$ for all $i \in \mathbb{S}$. Throughout this subsection, we assume that the following conditions fulfill.

(A1) $A : \mathcal{D}(A) \subset H \mapsto H$ is a self-adjoint compact operator on H such that $-A$ has discrete spectrum $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_k < \cdots$ and $\lim_{k\to\infty} \lambda_k = \infty$. In this case, A generates an analytic contraction semigroup $\{\mathrm{e}^{tA}\}_{t\geq 0}$, such that $\|\mathrm{e}^{tA}\| \leq \mathrm{e}^{-\lambda_1 t}$.

(A2) For each $i \in \mathbb{S}$, there exists $K_i > 0$ such that

$$|b(x, i) - b(y, i)|_H \leq K_i |x - y|_H, \qquad x, y \in H.$$

(A3) There exists $\theta \in (0, 1)$ such that $\alpha\theta \in (0, 1)$ and

$$\delta := \sum_{k=1}^{\infty} \frac{\beta_k^\alpha}{\lambda_k^{1-\alpha\theta}} < \infty.$$

Under assumption (A1)–(A3), according to Mao and Yuan [25], Theorem 3.13, page 89, and Priola and Zabczyk [30], Theorem 5.3, (2.3) admits a unique mild solution, that is, there exists a predictable H-valued stochastic process $\{X^\varepsilon(t)\}_{t\geq 0}$ such that

$$X^\varepsilon(t) = \mathrm{e}^{At}x + \int_0^t \mathrm{e}^{A(t-s)} b(X^\varepsilon(s), r^\varepsilon(s))\,\mathrm{d}s + \int_0^t \mathrm{e}^{A(t-s)}\,\mathrm{d}L(s), \qquad \mathbb{P}\text{-a.s.} \tag{2.5}$$

As can be seen, compared with the fast varying $r^\varepsilon(\cdot)$, $X^\varepsilon(\cdot)$ changes relatively slowly. The intuitive idea can be explained as follows. Using the methods of stochastic averaging initiated in Khasminskii [19] (see also Khasminskii [20], Khasminskii and Yin [21,22]) and subsequently developed by Kushner [24], $r^\varepsilon(t)$ can be treated essentially as a "noise" process. With the slow variable "fixed" or "frozen," a law of large numbers holds so the noise is averaged out. Moreover, the slow component $X^\varepsilon(t)$ converges to $\overline{X}(t)$ in an appropriate sense. We will show that the limit $\{\overline{X}(t)\}_{t\geq 0}$ satisfies in the mild sense an SPDE

$$\mathrm{d}\overline{X}(t) = \big\{A\overline{X}(t) + \overline{b}\big(\overline{X}(t)\big)\big\}\,\mathrm{d}t + \mathrm{d}L(t), \qquad t > 0, \tag{2.6}$$

with initial value $\overline{X}(0) = x \in H$, where $\overline{b}(x) := \sum_{i=1}^n b(x,i)\nu_i$, $x \in H$, an average with respect to the invariant measure $\nu := (\nu_1, \dots, \nu_n)$ given in (2.4). Our main result of this section is given as follows.

Theorem 2.1. *Let* (A1)–(A3) *hold and assume further that the initial value* $X^\varepsilon(0) = x \in \mathcal{D}((-A)^\theta)$. *Then, for any sufficiently small* $\varepsilon \in (0,1)$,

$$\sup_{0\leq t\leq T}\big(\mathbb{E}\big|X^\varepsilon(t) - \overline{X}(t)\big|_H^p\big)^{1/p} \lesssim_T \varepsilon^{\rho\theta}, \qquad p \in (1,\alpha),$$

where $\theta \in (0,1)$ *is the constant such that* (A3) *holds and* $\rho < (\alpha - p)/(\alpha - p + p\theta\alpha)$.

To prove Theorem 2.1, we need the following lemma.

Lemma 2.2. *Let the assumptions of Theorem* 2.1 *hold. Then, for any* $h \in (0,1)$ *and* $p \in (1,\alpha)$,

$$\sup_{0\leq t\leq T}\big(\mathbb{E}\big|\overline{X}(t+h) - \overline{X}(t)\big|_H^p\big)^{1/p} \lesssim_T h^\theta.$$

Proof. Noting that $(\mathbb{E}|\cdot|_H^p)^{1/p}$, $p \in (1,\alpha)$, is a norm, we get from (A1), (A2), and Priola and Zabczyk [30], (4.12), that

$$\begin{aligned}
&\big(\mathbb{E}\big|\overline{X}(t)\big|_H^p\big)^{1/p} \\
&\quad\leq |x|_H + \int_0^t \big\|\mathrm{e}^{(t-s)A}\big\|\big(\mathbb{E}\big|\overline{b}\big(\overline{X}(s)\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s + \mathbb{E}\bigg(\bigg|\int_0^t \mathrm{e}^{(t-s)A}\,\mathrm{d}L(s)\bigg|_H^p\bigg)^{1/p} \\
&\quad\leq |x|_H + \sum_{i=1}^n \nu_i \int_0^t \mathrm{e}^{-\lambda_1(t-s)}\big(\mathbb{E}\big|b\big(\overline{X}(s),i\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s + c\bigg(\sum_{k=1}^\infty \frac{\beta_k^\alpha(1-\mathrm{e}^{-\alpha\lambda_k t})}{\lambda_k}\bigg)^{1/\alpha} \\
&\quad\leq |x|_H + \sum_{i=1}^n \nu_i \int_0^t \mathrm{e}^{-\lambda_1(t-s)}\big\{K_i\big(\mathbb{E}\big|\overline{X}(s)\big|_H^p\big)^{1/p} + \big(\mathbb{E}\big|b(0,i)\big|_H^p\big)^{1/p}\big\}\,\mathrm{d}s + c\tau,
\end{aligned} \tag{2.7}$$

where $\tau := (\sum \frac{\beta_k^\alpha}{\lambda_k})^{1/\alpha} < \infty$ according to (A3). Multiplying $\mathrm{e}^{\lambda_1 t}$ on both sides of (2.7) gives

$$\mathrm{e}^{\lambda_1 t}\big(\mathbb{E}\big|\overline{X}(t)\big|_H^p\big)^{1/p} \leq c\big(1+\mathrm{e}^{\lambda_1 t}\big) + \sum_{i=1}^n \nu_i K_i \int_0^t \mathrm{e}^{\lambda_1 s}\big(\mathbb{E}\big|\overline{X}(s)\big|_H^p\big)^{1/p}\,\mathrm{d}s.$$

This, together with the Gronwall inequality, yields that

$$\sup_{0\leq t\leq T}\mathbb{E}\big|\overline{X}(t)\big|_H^p < \infty. \tag{2.8}$$

From (2.6), one has

$$
\begin{aligned}
&\big(\mathbb{E}\big|\overline{X}(t+h)-\overline{X}(t)\big|_H^p\big)^{1/p}\\
&\quad\le \big|\big(\mathrm{e}^{hA}-\mathbf{1}\big)\mathrm{e}^{tA}x\big|_H+\sum_{i=1}^n \nu_i\int_0^t\big(\mathbb{E}\big|\big(\mathrm{e}^{hA}-\mathbf{1}\big)\mathrm{e}^{(t-s)A}b\big(\overline{X}(s),i\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\
&\qquad+\sum_{i=1}^n \nu_i\int_t^{t+h}\big(\mathbb{E}\big|\mathrm{e}^{(t+h-s)A}b\big(\overline{X}(s),i\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\
&\qquad+\left(\mathbb{E}\left|\int_0^t\big(\mathrm{e}^{hA}-\mathbf{1}\big)\mathrm{e}^{(t-s)A}\,\mathrm{d}L(s)\right|_H^p\right)^{1/p}\\
&\qquad+\left(\mathbb{E}\left|\int_t^{t+h}\mathrm{e}^{(t+h-s)A}\,\mathrm{d}L(s)\right|_H^p\right)^{1/p}\\
&\quad=:\sum_{j=1}^5\Lambda_j(t),
\end{aligned}
\tag{2.9}
$$

where $\mathbf{1}$ is the identity operator on H. By the spectral properties of operator A, observe that

$$
\big\|(-A)^{\delta}\mathrm{e}^{tA}\big\|\le \mathrm{e}^{-\delta}\delta^{\delta}t^{-\delta},\qquad t>0,\delta\in(0,1) \tag{2.10}
$$

and that

$$
\big\|(-A)^{-\delta}\big(\mathbf{1}-\mathrm{e}^{tA}\big)\big\|\le c_\delta t^{\delta},\qquad t>0,\delta\in(0,1) \tag{2.11}
$$

for some $c_\delta>0$. Due to $x\in\mathcal{D}((-A)^{\theta})$, taking (A1), (A2), (2.8), (2.10), and (2.11) into account yields that

$$
\begin{aligned}
\Lambda_1(t)+\Lambda_2(t)\ &\le\ \big\|\big(\mathrm{e}^{hA}-\mathbf{1}\big)(-A)^{-\theta}\big\|\cdot\big\|\mathrm{e}^{tA}\big\|\cdot\big|(-A)^{\theta}x\big|_H\\
&\qquad+\sum_{i=1}^n\nu_i\int_0^t\big\|\big(\mathrm{e}^{hA}-\mathbf{1}\big)(-A)^{-\theta}\big\|\cdot\big\|\mathrm{e}^{(t-s)A/2}\big\|\\
&\qquad\qquad\times\big\|\mathrm{e}^{(t-s)A/2}(-A)^{\theta}\big\|\big(\mathbb{E}\big|b\big(\overline{X}(s),i\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\
&\lesssim_T\left(1+\int_0^t\mathrm{e}^{-\lambda_1(t-s)/2}\left(\frac{t-s}{2}\right)^{-\theta}\mathrm{d}s\right)h^{\theta}\\
&\lesssim_T\big(1+\Gamma(1-\theta)\big)h^{\theta},
\end{aligned}
\tag{2.12}
$$

where $\Gamma(\cdot)$ is the Gamma function. Also, by (A1), (A2), and (2.8), we arrive at

$$
\Lambda_3(t)\lesssim_T h. \tag{2.13}
$$

Note that

$$\int_0^t (-A)^\theta \mathrm{e}^{(t-s)A}\,\mathrm{d}L(s) = \sum_{k=1}^\infty \Big(\beta_k \lambda_k^\theta \int_0^t \mathrm{e}^{-(t-s)\lambda_k}\,\mathrm{d}L_k(s)\Big) e_k.$$

Upon using an argument similar to that of Priola and Zabczyk [30], Theorem 4.5, we obtain from (A3) that

$$\left(\mathbb{E}\left|\int_0^t (-A)^\theta \mathrm{e}^{(t-s)A}\,\mathrm{d}L(s)\right|_H^p\right)^{1/p} \lesssim \left(\sum_{k=1}^\infty \beta_k^\alpha \frac{1}{\alpha\lambda_k^{1-\alpha\theta}}\big(1-e^{-\alpha\lambda_k t}\big)\right)^{1/\alpha} \lesssim \delta^{1/\alpha}, \quad (2.14)$$

and that

$$\Lambda_5(t) \lesssim \left(\sum_{k=1}^\infty \frac{\beta_k^\alpha}{\lambda_k}\big(1-\mathrm{e}^{-\lambda_k h}\big)\right)^{1/\alpha} \lesssim \left(\sum_{k=1}^\infty \frac{\beta_k^\alpha}{\lambda_k}(\lambda_k h)^{\alpha\theta}\right)^{1/\alpha} \lesssim \delta^{1/\alpha} h^\theta, \qquad (2.15)$$

where we have used the fundamental inequality: for any $\gamma \in (0, 1]$, there exists $c_\gamma > 0$ such that

$$\big|\mathrm{e}^{-x} - \mathrm{e}^{-y}\big| \le c_\gamma |x-y|^\gamma, \qquad x, y \ge 0.$$

Thus we deduce from (2.11), (2.14), and (2.15) that

$$\Lambda_4(t) + \Lambda_5(t) \lesssim_T h^\theta. \qquad (2.16)$$

As a result, the desired assertion follows by substituting (2.12), (2.13), and (2.16) into (2.9). □

With the aid of Lemma 2.2, we complete the proof Theorem 2.1 in what follows.

Proof of Theorem 2.1. It follows from (2.3) and (2.6) that

$$\begin{aligned}\big(\mathbb{E}\big|X^\varepsilon(t) - \overline{X}(t)\big|_H^p\big)^{1/p} \le{}& \sum_{i=1}^n \int_0^t \big(\mathbb{E}\big|\mathrm{e}^{(t-s)A}\big\{b\big(X^\varepsilon(s), i\big) - b\big(\overline{X}(s), i\big)\big\}\big|_H^p\big)^{1/p}\,\mathrm{d}s \\ &+ \sum_{i=1}^n \left(\mathbb{E}\left|\int_0^t \mathrm{e}^{(t-s)A} b\big(\overline{X}(s), i\big)\{\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i\}\,\mathrm{d}s\right|_H^p\right)^{1/p} \\ =:{}& \Xi_1(t) + \sum_{i=1}^n \Xi_{2i}(t),\end{aligned}$$

where $\mathbb{1}_\Gamma$ is the indicator function of a set Γ. Taking (A1) and (A2) into account, we have

$$\Xi_1(t) \le \sum_{i=1}^n K_i \int_0^t \mathrm{e}^{-\lambda_1(t-s)}\big(\mathbb{E}\big|X^\varepsilon(s) - \overline{X}(s)\big|_H^p\big)^{1/p}\,\mathrm{d}s.$$

Next, note that from the boundedness of $|\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i|$,

$$\begin{aligned}
\Xi_{2i}(t) \le\ & \int_0^t \big\| \mathrm{e}^{(t-s)A}\big(\mathbf{1} - \mathrm{e}^{(s-\lfloor s\rfloor)A}\big)\big\| \big(\mathbb{E}\big|b\big(\overline{X}(s), i\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s \\
&+ \int_0^t \big\|\mathrm{e}^{(t-\lfloor s\rfloor)A}\big\| \big(\mathbb{E}\big|b\big(\overline{X}(s), i\big) - b\big(\overline{X}\big(\lfloor s\rfloor\big), i\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s \\
&+ \bigg(\mathbb{E}\bigg|\int_0^t \mathrm{e}^{(t-\lfloor s\rfloor)A} b\big(\overline{X}\big(\lfloor s\rfloor\big), i\big)\{\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i\}\,\mathrm{d}s\bigg|_H^p\bigg)^{1/p} \\
=:\ & \Upsilon_{1i}(t) + \Upsilon_{2i}(t) + \Upsilon_{3i}(t),
\end{aligned}$$

where $\lfloor s\rfloor := [s/\varepsilon^\rho]\varepsilon^\rho$ with $[s/\varepsilon^\rho]$ denoting the integer part of s/ε^ρ for $\rho < (\alpha - p)/(\alpha - p + p\theta\alpha)$. By a similar calculation as in (2.12), one has

$$\Upsilon_{1i}(t) \lesssim_T \varepsilon^{\rho\theta}. \tag{2.17}$$

By virtue of (A1), (A2), and Lemma 2.2, it follows that

$$\Upsilon_{2i}(t) \lesssim \int_0^t \mathrm{e}^{-\lambda_1(t-\lfloor s\rfloor)}\big(\mathbb{E}\big|\overline{X}(s) - \overline{X}\big(\lfloor s\rfloor\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s \lesssim_T \varepsilon^{\rho\theta}. \tag{2.18}$$

Let $t_j := j\varepsilon^\rho$, $j = 0, \dots, [t/\varepsilon^\rho]$, and $t_{[t/\varepsilon^\rho]+1} := t$. Then, an application of the Hölder inequality gives that

$$\begin{aligned}
\Upsilon_{3i}(t) &\le \sum_{j=0}^{\lfloor t/\varepsilon^\rho\rfloor} \bigg\{\mathbb{E}\big|\mathrm{e}^{(t-t_j)A} b\big(\overline{X}(t_j), i\big)\big|_H^p \bigg|\int_{t_j}^{t_{j+1}} \{\mathbf{1}_{\{r^\varepsilon(s)=i\}} - \nu_i\}\,\mathrm{d}s\bigg|^p\bigg\}^{1/p} \\
&\le \sum_{j=0}^{\lfloor t/\varepsilon^\rho\rfloor} \big(\mathbb{E}\big|\mathrm{e}^{(t-t_j)A} b\big(\overline{X}(t_j), i\big)\big|_H^{p(1+\delta)}\big)^{1/(p(1+\delta))} \\
&\qquad \times \bigg(\mathbb{E}\bigg|\int_{t_j}^{t_{j+1}} \{\mathbf{1}_{\{r^\varepsilon(s)=i\}} - \nu_i\}\,\mathrm{d}s\bigg|^{(p(1+\delta))/\delta}\bigg)^{\delta/(p(1+\delta))}
\end{aligned} \tag{2.19}$$

for arbitrary $0 < \delta < (\alpha - p)/p$. Thanks to $\alpha \in (1, 2)$ and $p \in (1, \alpha)$, one has $p > (2\alpha)/(2 + \alpha)$, which further gives that $(\alpha - p)/p < p/(2 - p)$. Then, for $0 < \delta < (\alpha - p)/p$, we have

$$p(1 + \delta) < \alpha \quad \text{and} \quad \big(p(1 + \delta)\big)/\delta > 2. \tag{2.20}$$

Hence, (A1), (2.8), and (2.20) yield that

$$\big(\mathbb{E}\big|\mathrm{e}^{(t-t_j)A} b\big(\overline{X}(t_j), i\big)\big|_H^{p(1+\delta)}\big)^{1/(p(1+\delta))} \lesssim_T \mathrm{e}^{-\lambda_1(t-t_j)}. \tag{2.21}$$

We claim that

$$\bigg(\mathbb{E}\bigg|\int_{t_j}^{t_{j+1}} \{\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i\}\,\mathrm{d}s\bigg|^{(p(1+\delta))/\delta}\bigg)^{\delta/(p(1+\delta))} \lesssim \varepsilon^{\rho+((\beta-\rho)\delta)/(p(1+\delta))}, \tag{2.22}$$

for sufficiently small $\varepsilon \in (0,1)$, where $\beta \in (\rho, 1)$ is some constant. To show (2.22), we adopt an argument similar to that of Yin and Zhang [34], Theorem 7.2, page 170. Let

$$\eta^\varepsilon(u) := \frac{1}{2}\mathbb{E}\left|\int_{t_j}^u \{\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i\}\,\mathrm{d}s\right|^2, \qquad u \in [t_j, t_{j+1}].$$

Then, it is easy to see from the chain rule that

$$\frac{\mathrm{d}\eta^\varepsilon(u)}{\mathrm{d}u} = \mathbb{E}\int_{t_j}^u \big\{(\mathbb{1}_{\{r^\varepsilon(u)=i\}} - \nu_i)(\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i)\big\}\,\mathrm{d}s, \qquad u \in [t_j, t_{j+1}].$$

Let $\bar{t}_k := k\varepsilon^\beta, k = 0, 1, \ldots, [(u - t_j)/\varepsilon^\beta] + 1$, where $\bar{t}_0 := t_j$ and $\bar{t}_{\lfloor (u-t_j)/\varepsilon^\beta \rfloor + 1} := u$. Thus, by the boundedness of $|\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i|$, we obtain that

$$\begin{aligned}
\frac{\mathrm{d}\eta^\varepsilon(u)}{\mathrm{d}u} &= \mathbb{E}\int_{\bar{t}_0}^{\widetilde{t}_j} \big\{(\mathbb{1}_{\{r^\varepsilon(u)=i\}} - \nu_i)(\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i)\big\}\,\mathrm{d}s \\
&\quad + \mathbb{E}\int_{\widetilde{t}_j}^{t} \big\{(\mathbb{1}_{\{r^\varepsilon(u)=i\}} - \nu_i)(\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i)\big\}\,\mathrm{d}s \\
&\lesssim \varepsilon^\beta + \mathbb{E}\int_{\bar{t}_0}^{\widetilde{t}_j} \big\{(\mathbb{1}_{\{r^\varepsilon(u)=i\}} - \nu_i)(\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i)\big\}\,\mathrm{d}s,
\end{aligned}$$

where $\widetilde{t}_j := \bar{t}_{[(t-t_j)/\varepsilon^\beta]-1}$. Recall from Yin and Zhang [34], Lemma 7.1, page 169, that

$$\big|\mathbb{P}^\varepsilon(u,s) - \nu\big| \lesssim \left(\varepsilon + \exp\left(-\frac{\kappa(u-s)}{\varepsilon}\right)\right), \qquad u \geq s \geq 0, \tag{2.23}$$

where $\mathbb{P}^\varepsilon(t,s) := (p_{ij}^\varepsilon(u,s))_{1\leq i,j\leq n} = (\mathbb{P}(r^\varepsilon(u) = j)|r^\varepsilon(s) = i)_{1\leq i,j\leq n}$, and $\kappa > 0$ is determined by the eigenvalues of $\widetilde{\mathbb{Q}}$. Thus, for $\mathscr{F}_{\widetilde{t}_j} := \sigma\{r^\varepsilon(s)\colon\ 0 \leq s \leq \widetilde{t}_j\}$, using the basic property of conditional expectation, we deduce that

$$\begin{aligned}
\big|\mathbb{E}\big\{(\mathbb{1}_{\{r^\varepsilon(u)=i\}} - \nu_i)(\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i)\big\}\big| &\leq \mathbb{E}\Big(\big|\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i\big| \cdot \big|\big(\mathbb{E}(\mathbb{1}_{\{r^\varepsilon(u)=i\}} - \nu_i)|\mathscr{F}_{\widetilde{t}_j}\big)\big|\Big) \\
&\lesssim \mathbb{E}\Big(\big|\big(\mathbb{E}(\mathbb{1}_{\{r^\varepsilon(u)=i\}} - \nu_i)|\mathscr{F}_{\widetilde{t}_j}\big)\big|\Big) \\
&\lesssim \left(\varepsilon + \exp\left(-\frac{\kappa(u-\widetilde{t}_j)}{\varepsilon}\right)\right) \\
&\lesssim \left(\varepsilon + \exp\left(-\frac{\kappa}{\varepsilon^{1-\beta}}\right)\right) \\
&\lesssim \varepsilon,
\end{aligned}$$

where in the last third step we used the fact (2.23), the last two step is due to $u > \bar{t}_{[(t-t_j)/\varepsilon^\beta]}$, while the last one owes to $\exp(-\frac{\kappa}{\varepsilon^{1-\beta}}) \lesssim \varepsilon$ for sufficiently small $\varepsilon \in (0,1)$. Hence,

$$\eta^\varepsilon(t) \lesssim \varepsilon^{\beta+\rho}. \tag{2.24}$$

Note that from (2.20) and the uniform boundedness of $|\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i| \le 1$,

$$\left(\mathbb{E}\left|\int_{t_j}^{t_{j+1}} \{\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i\}\,\mathrm{d}s\right|^{(p(1+\delta))/\delta}\right)^{\delta/(p(1+\delta))} \le \varepsilon^{\rho-(2\delta\rho)/(p(1+\delta))}\left(\mathbb{E}\left|\int_{t_j}^{t_{j+1}} \{\mathbb{1}_{\{r^\varepsilon(s)=i\}} - \nu_i\}\,\mathrm{d}s\right|^2\right)^{\delta/(p(1+\delta))}.$$

Then claim (2.21) follows from (2.24). Putting (2.21) and (2.22) into (2.19), we arrive at

$$\Upsilon_{3i}(t) \lesssim_T \sum_{j=0}^{\lfloor t/\varepsilon^\rho\rfloor} \mathrm{e}^{-\lambda_1(t-t_j)}\varepsilon^{\rho+(\beta\delta-2\delta\rho)/(p(1+\delta))} \lesssim_T \left(\mathrm{e}^{\lambda_1\varepsilon^\rho}-1\right)^{-1}\varepsilon^{\rho+((\beta-\rho)\delta)/(p(1+\delta))} \lesssim_T \varepsilon^{((\beta-\rho)\delta)/(p(1+\delta))}$$

due to the fact that $\mathrm{e}^{\lambda_1\varepsilon^\rho} - 1 = \mathrm{O}(\lambda_1\varepsilon^\rho)$ for sufficiently small $\varepsilon \in (0,1)$. So, we get

$$\left(\mathbb{E}\left|X^\varepsilon(t) - \overline{X}(t)\right|_H^p\right)^{1/p} \le C_T\left(\varepsilon^{\rho\theta} + \varepsilon^{((\beta-\rho)\delta)/(p(1+\delta))}\right) + \sum_{i=1}^n K_i\int_0^t \mathrm{e}^{-\lambda_1(t-s)}\left(\mathbb{E}\left|X^\varepsilon(s) - \overline{X}(s)\right|_H^p\right)^{1/p}\mathrm{d}s.$$

It follows from the Gronwall inequality that

$$\left(\mathbb{E}\left\|X^\varepsilon(s) - \overline{X}(s)\right\|_H^p\right)^{1/p} \lesssim_T \left(\varepsilon^{\rho\theta} + \varepsilon^{((\beta-\rho)\delta)/(p(1+\delta))}\right).$$

Then the desired assertion holds by noting that $\rho < (\alpha-\rho)/(\alpha-\rho+p\theta\alpha)$ and choosing appropriate $\beta \in (\rho, 1)$. □

Remark 2.1. By a close inspection of argument of Theorem 2.1, if $\sum_{i=1}^n K_i < \lambda_1$, we can also derive a long-term error bound below

$$\sup_{t\ge 0}\left(\mathbb{E}\left|X^\varepsilon(t) - \overline{X}(t)\right|_H^p\right)^{1/p} \lesssim \varepsilon^{\rho\theta}, \qquad p \in (1,\alpha),$$

for any $\alpha \in (1,2)$ and sufficiently small $\varepsilon \in (0,1)$, where $\theta \in (0,1)$ is the constant such that (A3) holds and $\rho < (\alpha-p)/(\alpha-p+p\theta\alpha)$.

Remark 2.2. By means of the martingale problem formulation, the weak convergence of $(X^\varepsilon(t), \overline{r}^\varepsilon(t))$ for hybrid finite-dimensional systems were obtained in Yin and Zhang [34], Theorem 7.20, page 204. In the current framework, it only admits a unique mild solution rather than strong solution so that the martingale-problem method seems not to be available. However, in this subsection, we investigate the strong convergence (in moment-sense) of $\{X^\varepsilon(t)\}_{t\ge 0}$ to the averaging process $\{\overline{X}(t)\}_{t\ge 0}$ defined by (2.6) by the semigroup approach. We also provide a convergence rate in terms of error bounds. Moreover, even for $\alpha = 2$, that is, the Wiener noise case, our result still seems to be new for infinite-dimensional stochastic dynamical systems.

2.2. Two-time-scale Markov switching with multiple weakly irreducible classes

In this subsection, we proceed to investigate the averaging principle associated with (2.3), where the Markov chain $r^\varepsilon(t)$ has a large state space

$$\mathbb{S} := \mathbb{S}_1 \cup \mathbb{S}_2 \cup \cdots \cup \mathbb{S}_l$$

with $\mathbb{S}_i := \{s_{i1}, \ldots, s_{in_i}\}$ and $n := n_1 + n_2 + \cdots + n_l$. Assume that the generator $\mathbb{Q}^\varepsilon := (q^\varepsilon_{ij})_{n\times n}$ of $r^\varepsilon(t)$ admits the form

$$\mathbb{Q}^\varepsilon := \frac{1}{\varepsilon}\widetilde{\mathbb{Q}} + \widehat{\mathbb{Q}},$$

where $\widetilde{\mathbb{Q}} := (\widetilde{q}_{ij})_{n\times n} = \mathrm{diag}(\widetilde{\mathbb{Q}}_1, \ldots, \widetilde{\mathbb{Q}}_l)$ such that, for each $k \in \{1, \ldots, l\}$, $\widetilde{\mathbb{Q}}_k$ is irreducible and the generator of some Markov chain taking values in $\mathbb{S}_k$ with the corresponding stationary distribution $\mu_k := (\mu_{k1}, \ldots, \mu_{kn_k}) \in \mathbb{R}^{1\times n_k}$, and $\widehat{\mathbb{Q}} := (\widehat{q}_{ij})_{n\times n}$. Since the transitions within each group take place at a fast pace, whereas the interactions from one group to another are relatively infrequently, following the basic idea in Yin and Zhang [34], we lump the states in each $\mathbb{S}_k$ into a single state and then define an aggregated process $\overline{r}^\varepsilon(\cdot)$ by

$$\overline{r}^\varepsilon(t) = k \qquad \text{for } r^\varepsilon(t) \in \mathbb{S}_k$$

with the associated state space $\overline{\mathbb{S}} := \{1, \ldots, l\}$. Let

$$\overline{\mathbb{Q}} := (\overline{q}_{ij})_{l\times l} = \widetilde{\mu}\widehat{\mathbb{Q}}\mathbf{I},$$

where $\widetilde{\mu} := \mathrm{diag}(\mu_1, \ldots, \mu_l) \in \mathbb{R}^{l\times n}$ and $\mathbf{I} := \mathrm{diag}(\mathbf{I}_{n_1}, \ldots, \mathbf{I}_{n_l})$ with $\mathbf{I}_{n_k} := (1, \ldots, 1)^T \in \mathbb{R}^{n_k\times 1}$, $k = 1, \ldots, l$. Recall from Yin and Zhang [34], Theorem 7.4, page 172, that $\overline{r}^\varepsilon(\cdot)$ converges weakly to the continuous-time Markov chain $\overline{r}(\cdot)$ with the state space $\overline{\mathbb{S}}$ and the generator $\overline{\mathbb{Q}}$ as $\varepsilon \to 0$, although generally $\overline{r}^\varepsilon(t)$ need not be a Markov chain. Our main result in this subsection is stated as follows.

Theorem 2.3. *Let* (A1)–(A3) *hold and suppose further that* $x \in \mathcal{D}((-A)^\theta)$. *Then*

$$\lim_{\varepsilon\to 0} \mathbb{E}\big|X^\varepsilon(t) - \overline{X}(t)\big|_H^p = 0, \qquad t \in [0,T] \text{ and } p \in (1,\alpha), \tag{2.25}$$

where $\overline{X}(t)$ *satisfies in the mild sense the following averaging equation*

$$\mathrm{d}\overline{X}(t) = \big\{A\overline{X}(t) + \overline{b}\big(\overline{X}(t), \overline{r}(t)\big)\big\}\,\mathrm{d}t + \mathrm{d}L(t), \qquad \overline{X}(0) = x, \overline{r}(0) = \overline{r}_0 \tag{2.26}$$

with $\overline{b}(y,i) := \sum_{j=1}^{n_i} \mu_{ij} b(y, s_{ij})$.

Proof. We only give an outline of the proof since it is very similar to that of Theorem 2.1. By (A1)–(A3), for any $p \in (1,\alpha)$, we deduce that

$$\sup_{0\le t\le T} \mathbb{E}\big|\overline{X}(t)\big|_H^p < \infty. \tag{2.27}$$

It is easy to see from (A1) and (A2) that

$$\begin{aligned}
&\big(\mathbb{E}\big|X^\varepsilon(t)-\overline{X}(t)\big|_H^p\big)^{1/p}\\
&\quad\le\sum_{i=1}^{l}\sum_{j=1}^{n_i}K_{s_{ij}}\int_0^t \mathrm{e}^{-\lambda_1(t-s)}\big(\mathbb{E}\big|X^\varepsilon(s)-\overline{X}(s)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\
&\quad\quad+\sum_{i=1}^{l}\sum_{j=1}^{n_i}\bigg(\mathbb{E}\bigg|\int_0^t \mathrm{e}^{(t-s)A}b\big(\overline{X}(s),s_{ij}\big)\{\mathbb{1}_{\{r^\varepsilon(s)=s_{ij}\}}-\mu_{ij}\mathbb{1}_{\{\overline{r}^\varepsilon(s)=i\}}\}\,\mathrm{d}s\bigg|_H^p\bigg)^{1/p}\\
&\quad\quad+\sum_{i=1}^{l}\sum_{j=1}^{n_i}\mu_{ij}\bigg(\mathbb{E}\bigg|\int_0^t \mathrm{e}^{(t-s)A}b\big(\overline{X}(s),s_{ij}\big)\{\mathbb{1}_{\{\overline{r}^\varepsilon(s)=i\}}-\mathbb{1}_{\{\overline{r}(s)=i\}}\}\,\mathrm{d}s\bigg|_H^p\bigg)^{1/p}\\
&\quad=:\Phi_1(t)+\Phi_2(t)+\Phi_3(t).
\end{aligned}$$

By the definition of $\overline{r}^\varepsilon(\cdot)$, one has

$$\big\{\overline{r}^\varepsilon(t)=i\big\}=\big\{r^\varepsilon(t)\in\mathbb{S}_i\big\}.$$

Then, in the same way as the proof of (2.22), we deduce from (2.27) and Yin and Zhang [34], Theorem 7.2, page 170, that

$$\Phi_2(t)\to 0\qquad\text{as }\varepsilon\to 0. \tag{2.28}$$

Next, applying the Hölder inequality, we find that

$$\begin{aligned}
\Phi_3(t)\ &\le\ \sum_{i=1}^{l}\sum_{j=1}^{n_i}\mu_{ij}\int_0^t \big\|\mathrm{e}^{(t-s)A}\big(\mathbf{1}-\mathrm{e}^{(s-\lfloor s\rfloor)A}\big)\big\|\big(\mathbb{E}\big|b\big(\overline{X}(s),s_{ij}\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\
&\quad+\sum_{i=1}^{l}\sum_{j=1}^{n_i}\mu_{ij}\int_0^t \big\|\mathrm{e}^{(t-\lfloor s\rfloor)A}\big\|\big(\mathbb{E}\big|b\big(\overline{X}(s),s_{ij}\big)-b\big(\overline{X}\big(\lfloor s\rfloor\big),s_{ij}\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\
&\quad+\sum_{i=1}^{l}\sum_{j=1}^{n_i}\mu_{ij}\sum_{k=0}^{\lfloor t/\varepsilon^\rho\rfloor}\big(\mathbb{E}\big|\mathrm{e}^{(t-t_k)A}b\big(\overline{X}(t_k),s_{ij}\big)\big|_H^{p(1+\delta)}\big)^{1/(p(1+\delta))}\\
&\qquad\qquad\times\bigg(\mathbb{E}\bigg|\int_{t_j}^{t_{j+1}}\{\mathbb{1}_{\{\overline{r}^\varepsilon(s)=i\}}-\mathbb{1}_{\{\overline{r}(s)=i\}}\}\,\mathrm{d}s\bigg|^{(p(1+\delta))/\delta}\bigg)^{\delta/(p(1+\delta))}\\
&=:\Theta_1(t)+\Theta_2(t)+\Theta_3(t),
\end{aligned}$$

where $\lfloor s\rfloor:=[s/\varepsilon^\rho]\varepsilon^\rho$ for $\rho\in(0,1)$, and $t_j:=j\varepsilon^\rho$, $j=0,\dots,[t/\varepsilon^\rho]$, and $t_{[t/\varepsilon^\rho]+1}:=t$. Moreover, carrying out similar arguments to those of (2.17) and (2.18) and utilizing (2.27) and Lemma 2.2 yields that

$$\Theta_1(t)+\Theta_2(t)\lesssim\varepsilon^{\rho\theta}. \tag{2.29}$$

From (A1) and (2.27), for sufficiently small $\varepsilon \in (0,1)$, it is seen that

$$\sum_{k=0}^{\lfloor t/\varepsilon^{\rho}\rfloor} \left(\mathbb{E}\left|\mathrm{e}^{(t-t_k)A} b\left(\overline{X}(t_k), s_{ij}\right)\right|_H^{p(1+\delta)}\right)^{1/(p(1+\delta))} \lesssim \varepsilon^{-\rho}.$$

On the other hand, by the weak convergence of $\overline{r}^{\varepsilon}(\cdot)$ to $\overline{r}(\cdot)$ (Yin and Zhang [34], Theorem 7.4, page 172), the Skorohod representation theorem (Yin and Zhang [34], Theorem 14.5, page 318), and the dominated convergence theorem, we have

$$\left(\mathbb{E}\left|\int_{t_j}^{t_{j+1}} \{\mathbb{1}_{\{\overline{r}^{\varepsilon}(s)=i\}} - \mathbb{1}_{\{\overline{r}(s)=i\}}\}\,\mathrm{d}s\right|^{(p(1+\delta))/\delta}\right)^{\delta/(p(1+\delta))} \lesssim \varepsilon^{\rho} g(\varepsilon),$$

where the positive function $g(\cdot)$ such that $g(\varepsilon) \to 0$ as $\varepsilon \to 0$. Then we obtain that

$$\Theta_3(t) \lesssim g(\varepsilon).$$

Henceforth the desired assertion follows from the Gronwall inequality. □

Remark 2.3. Unlike the case discussed in the previous subsection, it seems hard to give a strong convergence rate bound since the details on $\overline{r}(\cdot)$ are not enough, however the averaging equation (2.26) is explicitly dependent on the Markov chain $\overline{r}(\cdot)$, which is quite different from the case investigated in the last subsection.

3. SPDEs with an additional fast-varying process driven by another cylindrical stable process

In this section, we work on another two-time-scale system, in which there is an additional random process that has a fast-varying component driven by another cylindrical stable process.

For a small parameter $\varepsilon > 0$, we consider the following stochastic fast–slow system

$$\mathrm{d}X^{\varepsilon}(t) = \left\{AX^{\varepsilon}(t) + b\left(X^{\varepsilon}(t), Y^{\varepsilon}(t)\right)\right\}\mathrm{d}t + \mathrm{d}L(t), \qquad X^{\varepsilon}(0) = x \in \mathcal{D}\left((-A)^{1/2}\right) \tag{3.1}$$

and

$$\mathrm{d}Y^{\varepsilon}(t) = \frac{1}{\varepsilon}\left\{BY^{\varepsilon}(t) + f\left(X^{\varepsilon}(t), Y^{\varepsilon}(t)\right)\right\}\mathrm{d}t + \frac{1}{\varepsilon^{1/\beta}}\,\mathrm{d}Z(t), \qquad Y^{\varepsilon}(0) = y \in H. \tag{3.2}$$

Throughout this section, we shall assume that:

(B1) $A : \mathcal{D}(A) \subset H \mapsto H$ is a linear unbounded operator such that (A1) and $B : \mathcal{D}(B) \subset H \mapsto H$ is a self-adjoint compact operator on H such that $-B$ has discrete spectrum $0 < \mu_1 < \mu_2 < \cdots < \mu_k < \cdots$ and $\lim_{k\to\infty} \mu_k = \infty$.

(B2) b is uniformly bounded and Lipschitzian, that is, there exist $M, K_1 > 0$ such that

$$\sup_{x,y\in H} \left|b(x,y)\right| \le M,$$

and, for $x_1, x_2, y_1, y_2 \in H$,

$$\left|b(x_1, y_1) - b(x_2, y_2)\right|^2 \le K_1\left(|x_1 - y_1|_H^2 + |x_2 - y_2|_H^2\right).$$

(B3) For any $x, y \in H$ and $h \in H$, there exist $K_2, K_3 > 0$ such that $|\nabla^{(1)} f(x, y) \cdot h| \le K_2|h|$ and $|\nabla^{(2)} f(x, y) \cdot h| \le K_3|h|$, where $\nabla^{(1)} f$ and $\nabla^{(2)} f$ denote the Gâteaux derivative w.r.t. the first variable and the second variable, respectively.

(B4) There exists $\theta \in (0, 1)$ such that $\alpha\theta \in (0, 1)$,

$$\kappa_1 := \sum_{k=1}^{\infty} \frac{\beta_k^\alpha}{\lambda_k^{1-\alpha\theta}} < \infty \quad \text{and} \quad \kappa_2 := \sum_{k=1}^{\infty} \frac{q_k^\beta}{\mu_k} < \infty.$$

Under (B1)–(B4), both (3.1) and (3.2) are well-posed in the mild sense. Consider an SPDE associated with the fast variable, where the slow variable is fixed and equal to $z \in H$,

$$\mathrm{d}Y^z(t; y) = \left\{BY^z(t; y) + f\left(z, Y^z(t; y)\right)\right\} \mathrm{d}t + \mathrm{d}Z(t), \qquad Y^z(0; y) = y \in H. \tag{3.3}$$

Under (B1), (B3) and (B4), (3.3) has a unique mild solution $\{Y^z(t; y)\}_{t\ge0}$. Moreover, as Lemma 3.3 below states, (3.3) admits a unique ergodic invariant measure $\pi^z(\cdot) \in \mathcal{P}(H)$, the family of all probability measures on H. Our main result in this section is as follows:

Theorem 3.1. *Let* (A1) *and* (B1)–(B4) *hold and assume further that* $K_3 < \mu_1$. *Then,*

$$\lim_{\varepsilon\to0} \mathbb{E}\left|X^\varepsilon(t) - \overline{X}(t)\right|_H^p = 0, \qquad t \in [0, T],\ p \in (1, \alpha), \tag{3.4}$$

where $\overline{X}(t)$ *is the mild solution of the averaging equation*

$$\mathrm{d}\overline{X}(t) = \left\{A\overline{X}(t) + \overline{b}\left(\overline{X}(t)\right)\right\} \mathrm{d}t + \mathrm{d}L(t), \qquad \overline{X}(0) = x \in H \tag{3.5}$$

with

$$\overline{b}(z) := \int_H b(z, u)\pi^z(\mathrm{d}u), \qquad z \in H. \tag{3.6}$$

To facilitate the proof of Theorem 3.3, we shall present several technical lemmas in this regards and then finish the corresponding argument.

Lemma 3.2. *Under the assumptions of Theorem* 3.1,

$$\sup_{t\ge0} \mathbb{E}\left|Y^\varepsilon(t)\right|_H^p < \infty, \qquad p \in (1, \alpha). \tag{3.7}$$

Proof. It is easy to see from Priola and Zabczyk [30], (4.12), (A1), and (B1), that

$$\sup_{t\ge0} \mathbb{E}\left|X^\varepsilon(t)\right|_H^p < \infty. \tag{3.8}$$

Let

$$\overline{Z}^{\varepsilon}(t) := \frac{1}{\varepsilon^{1/\beta}} \int_0^t \mathrm{e}^{(t-s)B/\varepsilon} \, \mathrm{d}Z(s).$$

By Priola and Zabczyk [30], (4.12) and (B4), one has

$$\mathbb{E}\big|\overline{Z}^{\varepsilon}(t)\big|_H^p \le \varepsilon^{-p/\beta} \left(\sum_{k=1}^{\infty} q_k^{\beta} \int_0^t \mathrm{e}^{-\beta\mu_k(t-s)/\varepsilon} \, \mathrm{d}s \right)^{p/\beta} \le \big(\beta^{-1}\kappa_2\big)^{p/\beta}. \tag{3.9}$$

In view of (B1), (B3), (3.8), and (3.9), we then derive that

$$\begin{aligned}
\big(\mathbb{E}\big|Y^{\varepsilon}(t)\big|_H^p\big)^{1/p} &\le |y|_H + \varepsilon^{-1} \int_0^t \big\|\mathrm{e}^{(t-s)B/\varepsilon}\big\| \big(\mathbb{E}\big|f\big(X^{\varepsilon}(s), Y^{\varepsilon}(s)\big)\big|_H^p\big)^{1/p} \, \mathrm{d}s + \big(\mathbb{E}\big|\overline{Z}^{\varepsilon}(t)\big|_H^p\big)^{1/p} \\
&\le |y|_H + \varepsilon^{-1} \int_0^t \mathrm{e}^{-\mu_1(t-s)/\varepsilon} \big\{c\big(1 + |z|_H\big) + K_3\big(\mathbb{E}\big|Y^{\varepsilon}(s)\big|_H^p\big)^{1/p}\big\} \, \mathrm{d}s \\
&\le c\big(1 + |y|_H + |z|_H\big) + \frac{K_3}{\mu_1} \sup_{t\ge 0}\big(\mathbb{E}\big|Y^{\varepsilon}(t)\big|_H^p\big)^{1/p}.
\end{aligned}$$

This therefore leads to (3.7) due to $K_3 < \mu_1$. □

Lemma 3.3. *Assume that the assumptions of Theorem 3.1 hold. Then (3.3) admits a unique ergodic invariant measure $\pi^z(\cdot) \in \mathcal{P}(H)$ such that*

$$\big|\mathbb{E}b\big(Y^z(t; y)\big) - \overline{b}(z)\big|_H \lesssim \mathrm{e}^{-(\mu_1 - K_3)t}\big(1 + |y|_H + |z|_H\big). \tag{3.10}$$

Proof. We adopt the remote start method to show existence of an invariant measure for (3.3). Let $\widehat{Z}(t) := \sum_{k=1}^{\infty} q_k \widehat{Z}_k(t) e_k$, where $\{\widehat{Z}_k(t)\}_{k\ge1}$ is an independent copy of $\{Z_k(t)\}_{k\ge1}$, and $\{\widetilde{Z}(t)\}_{t\ge0}$ be a double-sided cylindrical β-stable process defined by

$$\widetilde{Z}(t) := \begin{cases} Z(t), & t \ge 0, \\ \widehat{Z}(-t), & t < 0 \end{cases}$$

with the filtration

$$\overline{\mathscr{F}}_t := \bigcap_{s>t} \overline{\mathscr{F}}_s^0,$$

where $\overline{\mathscr{F}}_s^0 := \sigma(\{\widetilde{Z}(r_2) - \widetilde{Z}(r_1)\colon -\infty < r_1 \le r_2 \le s, \Gamma\}, \mathscr{N})$ and $\mathscr{N} := \{A \in \mathscr{F} | \mathbb{P}(A) = 0\}$. Next, consider (3.3), for arbitrary $s \in (-\infty, t]$ with $t \in \mathbb{R}$,

$$\mathrm{d}Y^z(t; s, y) = \big\{BY^z(t; s, y) + f\big(z, Y^z(t; s, y)\big)\big\} \, \mathrm{d}t + \mathrm{d}\widetilde{Z}(t), \qquad Y^z(s; s, y) = y \in H. \tag{3.11}$$

Set $\Gamma^z(t; y) := Y^z(t; -\lambda, y) - Y^z(t; -\gamma, y)$ for $-\lambda \in (-\gamma, t]$. By (B1) and (B3), it follows that

$$\big(\mathbb{E}\big|\Gamma^z(t; y)\big|_H^p\big)^{1/p} \le \mathrm{e}^{-\mu_1(t+\lambda)}\big(\mathbb{E}\big|\Gamma^z(-\lambda; y)\big|_H^p\big)^{1/p} + K_3 \int_{-\lambda}^t \mathrm{e}^{-\mu_1(t-s)}\big(\mathbb{E}\big|\Gamma^z(s; y)\big|_H^p\big)^{1/p} \, \mathrm{d}s.$$

Multiplying $\mathrm{e}^{\mu_1 t}$ on both sides leads to

$$\mathrm{e}^{\mu_1 t}\big(\mathbb{E}\big|\Gamma^z(t;y)\big|_H^p\big)^{1/p} \le \mathrm{e}^{-\mu_1\lambda}\big(\mathbb{E}\big|\Gamma^z(-\lambda;y)\big|_H^p\big)^{1/p} + K_3\int_{-\lambda}^t \mathrm{e}^{\mu_1 s}\big(\mathbb{E}\big|\Gamma^z(s;y)\big|_H^p\big)^{1/p}\,\mathrm{d}s.$$

Thus we get from the Gronwall inequality that

$$\big(\mathbb{E}\big|\Gamma^z(t;y)\big|_H^p\big)^{1/p} \le \mathrm{e}^{-(\mu_1-K_3)(t+\lambda)}\big(\mathbb{E}\big|\Gamma^z(-\lambda;y)\big|_H^p\big)^{1/p}. \tag{3.12}$$

Moreover, carrying out an argument of Lemma 3.2, we have

$$\sup_{t\ge s}\big(\mathbb{E}\big|Y^z(t;s,y)\big|_H^p\big)^{1/p} \lesssim 1+|y|_H+|z|_H, \qquad s\in\mathbb{R}. \tag{3.13}$$

For $t=0$ and $-\lambda\in(-\mu,0]$, we deduce from (3.12) and (3.13) that

$$\big(\mathbb{E}\big|Y^z(0;-\lambda,y)-Y^z(0;-\gamma,y)\big|_H^p\big)^{1/p} \lesssim \big(1+|y|_H+|z|_H\big)\mathrm{e}^{-(\mu_1-K_3)\lambda}.$$

From the estimate above, we conclude that $\{Y^z(0;-t,y)\}_{t\ge 0}$ is a Cauchy sequence in $L^p(\Omega;H)$, and therefore it is convergent to a random variable $\eta_z(y)\in L^p(\Omega;H)$, which is independent of $y\in H$, and denoted by $\eta_z\in L^p(\Omega;H)$. Then, following a standard procedure (see, e.g., Prévôt and Röckner [29], pages 109–110), we deduce that $\mathscr{L}(\eta_z)=:\pi^z(\cdot)$ is an invariant measure of (3.3).

Next, following an argument of (3.12), we obtain that

$$\big(\mathbb{E}\big|Y^z(t;y_1)-Y^z(t;y_2)\big|_H^p\big)^{1/p} \le \mathrm{e}^{-(\mu_1-K_3)t}|y_1-y_2|_H. \tag{3.14}$$

This, together with (3.13), implies that

$$\mathbb{E}\big|Y^z(t;y)\big|_H^p \le \mathrm{e}^{-p(\mu_1-K_3)t}|y|_H^p + c\big(1+|z|_H^p\big). \tag{3.15}$$

Furthermore, by virtue of (3.15) and using a stationary solution $Y^z(t,y)$ with invariant law $\pi^z(\cdot)$, we obtain that

$$\mathbb{E}^z|y|_H^p = \mathbb{E}^z\big|Y^z(t,y)\big|_H^p \le \mathrm{e}^{-p(\mu_1-K_3)t}\mathbb{E}^z|y|_H^p + c\big(1+|z|_H^p\big), \qquad t\ge 0, \tag{3.16}$$

where $\mathbb{E}^z$ is the mathematical expectation operator w.r.t. $\pi^z(\cdot)$. (3.16) further gives that

$$\pi^z\big(|\cdot|_H^p\big) \lesssim 1+|z|_H^p. \tag{3.17}$$

Consequently, (3.14) and (3.17) yield the uniqueness of invariant measure. Indeed, if $\widetilde{\pi}^z(\cdot)\in\mathcal{P}(H)$ is also an invariant measure, for any $\psi\in C_b^2(H;\mathbb{R})$, by the invariance of $\pi^z(\cdot)$ and $\widetilde{\pi}^z(\cdot)$, we deduce from (3.14) and (3.17) that

$$\begin{aligned}\big|\pi^z(\psi)-\widetilde{\pi}^z(\psi)\big| &\le c\mathrm{e}^{-(\mu_1-K_3)t}\big\{\pi^z\big(|\cdot|_H\big)+\widetilde{\pi}\big(|\cdot|_H\big)\big\}\\ &\le c\mathrm{e}^{-(\mu_1-K_3)t}\big\{1+|z|_H\big\}\to 0 \qquad \text{as } t\uparrow\infty.\end{aligned}$$

That is, for any $\psi \in C_b^2(H;\mathbb{R})$, $\pi^z(\psi) = \widetilde{\pi}^z(\psi)$, which shows that $\pi \equiv \widetilde{\pi}$ due to Ikeda and Watanabe [18], Proposition 2.2, page 3.

Finally, (3.10) follows by noting from the invariance of $\pi^z(\cdot)$, (3.17) and the Lipschitz property of b. □

Applying the Lipschitzian property of b, the ergodic property of invariant measure $\pi^z(\cdot) \in \mathcal{P}(H)$ due to Lemma 3.3 and the uniform boundedness of the directional derivative $\nabla_h Y^z(t;y)$ with respect to $z \in H$ along the direction $h \in H$, and adopting a similar argument in Cerrai and Freidlin [7], (5.4), we deduce that $\overline{b}$ is Lipschitzian, which is stated as the following corollary for citation convenience.

Corollary 3.4. *Under the assumptions of Theorem 3.1, $\overline{b}: H \to H$ is Lipschitzian.*

To reveal the error analysis between the slow component $\{X^\varepsilon(t)\}_{t\ge 0}$ and the averaging process $\{\overline{X}(t)\}_{t\ge 0}$, determined by (3.5), we further need to define the following two auxiliary processes:

$$\widetilde{Y}^\varepsilon(t) := \mathrm{e}^{tB/\varepsilon} y + \frac{1}{\varepsilon}\int_0^t \mathrm{e}^{(t-s)B/\varepsilon} f\big(X^\varepsilon\big(\lfloor s/\delta\rfloor\delta\big), \widetilde{Y}^\varepsilon(s)\big)\,\mathrm{d}s + \frac{1}{\varepsilon^{1/\beta}}\int_0^t \mathrm{e}^{(t-s)B/\varepsilon}\,\mathrm{d}Z(s) \tag{3.18}$$

and

$$\widetilde{X}^\varepsilon(t) := \mathrm{e}^{tA} x + \int_0^t \mathrm{e}^{(t-s)A} b\big(X^\varepsilon\big(\lfloor s/\delta\rfloor\delta\big), \widetilde{Y}^\varepsilon(s)\big)\,\mathrm{d}s + \int_0^t \mathrm{e}^{(t-s)A}\,\mathrm{d}L(s), \tag{3.19}$$

where $\delta \in (\varepsilon, 1)$ is some constant to be chosen.

Lemma 3.5. *Assume that the assumptions of Theorem 3.1 hold. Then, for any $p \in (1,\alpha)$,*

$$\int_0^T \big(\mathbb{E}\big|Y^\varepsilon(s) - \widetilde{Y}^\varepsilon(s)\big|_H^p\big)^{1/p}\,\mathrm{d}s \lesssim \frac{\varepsilon}{\delta} + \varepsilon\delta^{-(1-\theta)}\mathrm{e}^{K_3\delta/\varepsilon} \tag{3.20}$$

and

$$\int_0^T \big(\mathbb{E}\big|X^\varepsilon(s) - \widetilde{X}^\varepsilon(s)\big|_H^p\big)^{1/p}\,\mathrm{d}s \lesssim \delta^\theta + \frac{\varepsilon}{\delta} + \varepsilon\delta^{-(1-\theta)}\mathrm{e}^{K_3\delta/\varepsilon}, \tag{3.21}$$

where $\theta \in (0,1)$ is the constant such that (B4).

Proof. For notation simplicity, we set $\Lambda^\varepsilon(t) := Y^\varepsilon(t) - \widetilde{Y}^\varepsilon(t)$. By Lemma 2.2, for any $t \in [0,T]$ it follows from (B1) and (B2) that

$$\begin{aligned}\big(\mathbb{E}\big|X^\varepsilon(t) - \widetilde{X}^\varepsilon(t)\big|_H^p\big)^{1/p} &\le \int_0^t \big(\mathbb{E}\big|b\big(X^\varepsilon(s), Y^\varepsilon(s)\big) - b\big(X^\varepsilon\big(\lfloor s/\delta\rfloor\delta\big), \widetilde{Y}^\varepsilon(s)\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s \\ &\lesssim \int_0^t \big(\mathbb{E}\big|X^\varepsilon(s) - X^\varepsilon\big(\lfloor s/\delta\rfloor\delta\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s + \int_0^t \big(\mathbb{E}\big|\Lambda^\varepsilon(s)\big|_H^p\big)^{1/p}\,\mathrm{d}s \\ &\lesssim_T \delta^\theta + \int_0^t \big(\mathbb{E}\big|\Lambda^\varepsilon(s)\big|_H^p\big)^{1/p}\,\mathrm{d}s.\end{aligned}$$

Therefore, to complete the proof of Lemma 3.5, it is sufficient to show (3.20). Carrying out similar arguments to those of (3.7) and (3.8), we also deduce that

$$\sup_{t\geq 0}\mathbb{E}\big|\widetilde{X}^\varepsilon(t)\big|_H^p \vee \sup_{t\geq 0}\mathbb{E}\big|\widetilde{Y}^\varepsilon(t)\big|_H^p < \infty. \tag{3.22}$$

For any $t \in [0, T]$, there exists an integer $k \geq 0$ such that $t \in [k\delta, (k+1)\delta)$. From (B3) and Lemma 2.2, we derive that

$$\begin{aligned}
&\big(\mathbb{E}\big|\Lambda^\varepsilon(t)\big|_H^p\big)^{1/p}\\
&\quad\leq \mathrm{e}^{-\mu_1(t-k\delta)/\varepsilon}\big(\mathbb{E}\big|\Lambda^\varepsilon(k\delta)\big|_H^p\big)^{1/p}\\
&\qquad+\frac{1}{\varepsilon}\int_{k\delta}^t \mathrm{e}^{-\mu_1(t-s)/\varepsilon}\big(\mathbb{E}\big|f\big(X^\varepsilon(s),Y^\varepsilon(s)\big)-f\big(X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(s)\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\
&\quad\leq \mathrm{e}^{-\mu_1(t-k\delta)/\varepsilon}\big(\mathbb{E}\big|\Lambda^\varepsilon(k\delta)\big|_H^p\big)^{1/p}\\
&\qquad+\frac{1}{\varepsilon}\int_{k\delta}^t \mathrm{e}^{-\mu_1(t-s)/\varepsilon}\big\{K_2\big(\mathbb{E}\big|X^\varepsilon(s)-X^\varepsilon(k\delta)\big|_H^p\big)^{1/p}+K_3\big(\mathbb{E}\big|\Lambda^\varepsilon(s)\big|_H^p\big)^{1/p}\big\}\,\mathrm{d}s.
\end{aligned}$$

This, together with the combined use of (3.7) and (3.22), yields that

$$\begin{aligned}
&\mathrm{e}^{\mu_1 t/\varepsilon}\big(\mathbb{E}\big|\Lambda^\varepsilon(t)\big|_H^p\big)^{1/p}\\
&\quad\leq c\mathrm{e}^{\mu_1 k\delta/\varepsilon}+\frac{c}{\varepsilon}\int_{k\delta}^t \mathrm{e}^{\mu_1 s/\varepsilon}\big(\mathbb{E}\big|X^\varepsilon(s)-X^\varepsilon(k\delta)\big|_H^p\big)^{1/p}+\frac{K_3}{\varepsilon}\int_{k\delta}^t \mathrm{e}^{\mu_1 s/\varepsilon}\big(\mathbb{E}\big|\Lambda^\varepsilon(s)\big|_H^p\big)^{1/p}\,\mathrm{d}s.
\end{aligned}$$

Then, applying the Gronwall inequality and using Lemma 2.2, we obtain that

$$\begin{aligned}
&\big(\mathbb{E}\big|\Lambda^\varepsilon(t)\big|_H^p\big)^{1/p}\\
&\quad\leq c\mathrm{e}^{-(\mu_1-K_3)(t-k\delta)/\varepsilon}\\
&\qquad+\frac{K_3}{\varepsilon}\int_{k\delta}^t \mathrm{e}^{(-(\mu_1-K_3)t-K_3k\delta+\mu_1 s)/\varepsilon}\big(\mathbb{E}\big|X^\varepsilon(s)-X^\varepsilon(k\delta)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\
&\quad\leq c\mathrm{e}^{-(\mu_1-K_3)(t-k\delta)/\varepsilon}-\frac{cK_3\delta^\theta}{\lambda_1}\mathrm{e}^{-(\mu_1-K_3)(t-k\delta)/\varepsilon}+\frac{cK_3\delta^\theta}{\mu_1}\mathrm{e}^{K_3(t-k\delta)/\varepsilon}\\
&\quad\lesssim \mathrm{e}^{-(\mu_1-K_3)(t-k\delta)/\varepsilon}+\frac{K_3\delta^\theta}{\mu_1}\mathrm{e}^{K_3(t-k\delta)/\varepsilon}.
\end{aligned}$$

Integrating from $k\delta$ to $(k+1)\delta$ with respect to the variable t in the above leads to

$$\begin{aligned}
\int_{k\delta}^{(k+1)\delta}\big(\mathbb{E}\big|\Lambda^\varepsilon(t)\big|_H^p\big)^{1/p}\,\mathrm{d}t &\lesssim \int_{k\delta}^{(k+1)\delta}\left\{\mathrm{e}^{-(\mu_1-K_3)(t-k\delta)/\varepsilon}+\frac{K_3\delta^{1/2}}{\lambda_1}\mathrm{e}^{K_3(t-k\delta)/\varepsilon}\right\}\mathrm{d}t\\
&\lesssim \varepsilon+\varepsilon\delta^\theta\mathrm{e}^{K_3\delta/\varepsilon}.
\end{aligned}$$

Thus, (3.20) follows. □

Remark 3.1. Bréhier [5], Lemma 3.1, confined Lemma 3.5 on the case $p = 1$, which is not sufficient for our purposes, and the techniques used there does not work for our model. On the other hand, for finite-dimensional jump–diffusion processes, Givon [14], Lemma 2.4, gives a similar estimate making use of the Itô formula, which is unavailable for our framework since the noise process does not admits second moments.

With the previous lemmas at hand, we now can complete the proof of Theorem 3.1.

Proof of Theorem 3.1. The proof is inspired by Khasminskii [20]. According to (B2), Lemmas 2.2 and 3.5, it then follows that

$$\begin{aligned}\big(\mathbb{E}\big|X^\varepsilon(t)-\overline{X}(t)\big|_H^p\big)^{1/p} &\lesssim \big(\mathbb{E}\big|\widetilde{X}^\varepsilon(t)-\overline{X}(t)\big|_H^p\big)^{1/p}+\int_0^t\big(\mathbb{E}\big|X^\varepsilon(s)-X^\varepsilon\big(\lfloor s/\delta\rfloor\delta\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\&\quad+\int_0^t\big(\mathbb{E}\big|Y^\varepsilon(s)-\widetilde{Y}^\varepsilon(s)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\&\lesssim \delta^\theta+\frac{\varepsilon}{\delta}+\varepsilon\delta^{-(1-\theta)}\mathrm{e}^{K_3\delta/\varepsilon}+\big(\mathbb{E}\big|\widetilde{X}^\varepsilon(t)-\overline{X}(t)\big|_H^p\big)^{1/p}.\end{aligned}$$

Therefore, to get the desired assertion, it is sufficient to show that

$$\big(\mathbb{E}\big|\widetilde{X}^\varepsilon(t)-\overline{X}(t)\big|_H^p\big)^{1/p}\lesssim \delta^\theta+\frac{\varepsilon}{\delta}+\sqrt{\frac{\varepsilon}{\delta}}+\varepsilon\delta^{-(1-\theta)}\mathrm{e}^{K_3\delta/\varepsilon}. \tag{3.23}$$

By the Lipschitz property of $\overline{b}$ due to Corollary 3.4, Lemmas 2.2 and 3.5, we deduce that

$$\begin{aligned}&\big(\mathbb{E}\big|\widetilde{X}^\varepsilon(t)-\overline{X}(t)\big|_H^p\big)^{1/p}\\&\quad\le\bigg(\mathbb{E}\bigg|\int_0^t\mathrm{e}^{(t-s)A}\big\{b\big(X^\varepsilon\big(\lfloor s/\delta\rfloor\delta\big),\widetilde{Y}(s)\big)-\overline{b}\big(X^\varepsilon\big(\lfloor s/\delta\rfloor\delta\big)\big)\big\}\,\mathrm{d}s\bigg|_H^p\bigg)^{1/p}\\&\qquad+\int_0^t\big(\mathbb{E}\big|\overline{b}\big(X^\varepsilon\big(\lfloor s/\delta\rfloor\delta\big)\big)-\overline{b}\big(X^\varepsilon(s)\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\&\qquad+\int_0^t\big(\mathbb{E}\big|\overline{b}\big(X^\varepsilon(s)\big)-\overline{b}\big(\widetilde{X}^\varepsilon(s)\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s+\int_0^t\big(\mathbb{E}\big|\overline{b}\big(\widetilde{X}^\varepsilon(s)\big)-\overline{b}\big(\overline{X}(s)\big)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\&\quad\lesssim\delta^\theta+\frac{\varepsilon}{\delta}+\varepsilon\delta^{-(1-\theta)}\mathrm{e}^{K_3\delta/\varepsilon}+\int_0^t\big(\mathbb{E}\big|\widetilde{X}^\varepsilon(s)-\overline{X}(s)\big|_H^p\big)^{1/p}\,\mathrm{d}s\\&\qquad+\bigg(\mathbb{E}\bigg|\int_0^t\mathrm{e}^{(t-s)A}\big\{b\big(X^\varepsilon\big(\lfloor s/\delta\rfloor\delta\big),\widetilde{Y}(s)\big)-\overline{b}\big(X^\varepsilon\big(\lfloor s/\delta\rfloor\delta\big)\big)\big\}\,\mathrm{d}s\bigg|_H^p\bigg)^{1/p}.\end{aligned}\tag{3.24}$$

Furthermore, noting that

$$\left|\int_0^t h(s)\,\mathrm{d}s\right|_H^2 = 2\int_0^t\int_s^t \langle h(r), h(s)\rangle_H\,\mathrm{d}r\,\mathrm{d}s$$

for a locally integrable function $h:[0,\infty)\mapsto H$, we obtain from Jensen's inequality that

$$\begin{aligned}
&\left(\mathbb{E}\left|\int_0^t \mathrm{e}^{(t-s)A}\big\{b\big(X^\varepsilon(\lfloor s/\delta\rfloor\delta),\widetilde{Y}(s)\big)-\overline{b}\big(X^\varepsilon(\lfloor s/\delta\rfloor\delta)\big)\big\}\,\mathrm{d}s\right|_H^p\right)^{1/p}\\
&\quad\le \sum_{k=0}^{\lfloor t/\delta\rfloor}\left(\mathbb{E}\left|\int_{k\delta}^{(k+1)\delta}\mathrm{e}^{(t-s)A}\big\{b\big(X^\varepsilon(k\delta),\widetilde{Y}(s)\big)-\overline{b}\big(X^\varepsilon(k\delta)\big)\big\}\,\mathrm{d}s\right|_H^p\right)^{1/p} \qquad (3.25)\\
&\quad\lesssim \varepsilon\sum_{k=0}^{\lfloor t/\delta\rfloor}\left(\int_0^{\delta/\varepsilon}\int_s^{\delta/\varepsilon}\mathcal{J}_k(r,s)\,\mathrm{d}r\,\mathrm{d}s\right)^{1/2},
\end{aligned}$$

where $t := (\lfloor t/\delta\rfloor + 1)\delta$ and

$$\begin{aligned}
\mathcal{J}_k(r,s) := \mathbb{E}\big\langle &\mathrm{e}^{(t-(k\delta+r\varepsilon))A}\big(b\big(X^\varepsilon(k\delta),\widetilde{Y}(r\varepsilon+k\delta)\big)-\overline{b}\big(X^\varepsilon(k\delta)\big)\big),\\
&\mathrm{e}^{(t-(k\delta+s\varepsilon))A}\big(b\big(X^\varepsilon(k\delta),\widetilde{Y}(s\varepsilon+k\delta)\big)-\overline{b}\big(X^\varepsilon(k\delta)\big)\big)\big\rangle_H.
\end{aligned}$$

For any $s\in(0,\delta)$, observe from (3.18) that

$$\begin{aligned}
\widetilde{Y}^\varepsilon(s+k\delta) = \mathrm{e}^{sB/\varepsilon}\widetilde{Y}^\varepsilon(k\delta) &+ \frac{1}{\varepsilon}\int_0^s \mathrm{e}^{(s-u)B/\varepsilon} f\big(X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta+u)\big)\,\mathrm{d}u\\
&+\frac{1}{\varepsilon^{1/\beta}}\int_0^s \mathrm{e}^{(s-u)B/\varepsilon}\,\mathrm{d}Z_1(u),
\end{aligned} \qquad (3.26)$$

where $Z_1(\cdot) := Z(\cdot + k\delta) - Z(k\delta)$ with filtration $\mathscr{F}_{\cdot+k\delta}$, which is again a cylindrical β-stable process. Let

$$Z_2(t) := \sum_{k=1}^{\infty} q_k \overline{Z}_k(t) e_k,$$

where $\{\overline{Z}_k(t)\}_{k\ge1}$ is a sequence of i.i.d. $\mathbb{R}$-valued symmetric β-stable Lévy processes defined on the filtered probability space $(\Omega,\mathscr{F},\{\mathscr{F}_t\}_{t\ge0},\mathbb{P})$ such that $\{Z_2(t)\}_{t\ge0}$ is independent of $\{L(t)\}_{\ge0}$ and $\{Z(t)\}_{t\ge0}$, respectively. For fixed $X^\varepsilon(k\delta)$ and the starting point $\widetilde{Y}^\varepsilon(k\delta)$, define the process $Y_s^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}$ by

$$\begin{aligned}
Y_{s/\varepsilon}^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)} := \mathrm{e}^{sB/\varepsilon}\widetilde{Y}^\varepsilon(k\delta) &+ \int_0^{s/\varepsilon}\mathrm{e}^{(s/\varepsilon-u)B} f\big(X^\varepsilon(k\delta), Y_u^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}\big)\,\mathrm{d}u\\
&+\int_0^{s/\varepsilon}\mathrm{e}^{(s/\varepsilon-u)B}\,\mathrm{d}Z_2(u).
\end{aligned} \qquad (3.27)$$

A simple calculation gives that

$$\begin{aligned} Y_{s/\varepsilon}^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)} &= \mathrm{e}^{sB/\varepsilon}\widetilde{Y}^\varepsilon(k\delta) + \frac{1}{\varepsilon}\int_0^s \mathrm{e}^{(s-u)B/\varepsilon} f\big(X^\varepsilon(k\delta), Y_{u/\varepsilon}^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}\big)\,\mathrm{d}u \\ &\quad + \frac{1}{\varepsilon^{1/\beta}}\int_0^s \mathrm{e}^{(s-u)B/\varepsilon}\,\mathrm{d}Z_3(u), \qquad s \in (0,\delta), \end{aligned} \tag{3.28}$$

where $Z_3(\cdot) := \varepsilon^{1/\beta} Z_2(\cdot/\varepsilon)$. By the self-similar property of stable Lévy processes (Applebaum [2], page 51), we conclude from (3.26) and (3.27) that

$$\mathscr{L}\big(\widetilde{Y}^\varepsilon(s+k\delta)\big) = \mathscr{L}\big(Y_{s/\varepsilon}^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}\big), \qquad s \in (0,\delta). \tag{3.29}$$

This further implies from (3.22) that

$$\sup_{s\in[0,\delta]} \mathbb{E}\big|Y_s^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}\big|_H^p < \infty. \tag{3.30}$$

Let

$$\mathscr{F}_s := \sigma\big\{Y_u^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}, u \le s\big\}.$$

Then $X^\varepsilon(k\delta) \in \mathscr{F}_s$. By the property of conditional expectation (Applebaum [2], Lemma 1.1.9), and the boundedness of b due to (B2), for $r > s$ we obtain from (3.29) that

$$\begin{aligned} \mathcal{J}_k(r,s) &= \mathbb{E}\big\langle \mathrm{e}^{(t-(k\delta+s\varepsilon))A}\big(b\big(X^\varepsilon(k\delta), Y_s^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}\big) - \overline{b}\big(X^\varepsilon(k\delta)\big)\big) \\ &\qquad \times \mathrm{e}^{(t-(k\delta+r\varepsilon))A}\big(\mathbb{E}\big(b\big(X^\varepsilon(k\delta), Y_r^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}\big) - \overline{b}\big(X^\varepsilon(k\delta)\big)\big)|\mathscr{F}_s\big)\big\rangle_H \\ &= \mathbb{E}\big\langle \mathrm{e}^{(t-(k\delta+s\varepsilon))A}\big(b\big(X^\varepsilon(k\delta), Y_s^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}\big) - \overline{b}\big(X^\varepsilon(k\delta)\big)\big) \\ &\qquad \times \mathrm{e}^{(t-(k\delta+r\varepsilon))A}\big(\mathbb{E}\big(b\big(z_1, Y_{r-s}^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)} + z_2\big) - \overline{b}(z_1)\big)\big)\big|_{\substack{z_1 = X^\varepsilon(k\delta) \\ z_2 = Y_s^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}}}\big\rangle_H \\ &\le \big(\mathbb{E}\big|b\big(z_1, Y_s^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}\big) - \overline{b}\big(z_1(\xi)\big)\big|_H^2\big)^{1/2} \\ &\qquad \times \big(\mathbb{E}\big|\big(\mathbb{E}\big(b\big(z_1, Y_{r-s}^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)} + z_2\big) - \overline{b}\big(z_1(\xi)\big)\big)\big)\big|_{\substack{z_1(\xi) = X^\varepsilon(k\delta) \\ z_2 = Y_s^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}}}\big|_H^2\big)^{1/2} \\ &\lesssim \mathbb{E}\big(\big|\big(\mathbb{E}\big(b\big(z_1, Y_{r-s}^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)} + z_2\big) - \overline{b}(z_1)\big)\big)\big|_{\substack{z_1 = X^\varepsilon(k\delta) \\ z_2 = Y_s^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}}}\big|_H\big), \end{aligned}$$

where in the last step we have used the boundedness of b due to (B2). The previous estimation, combining Lemma 3.3 with (3.8) and (3.30), yields that

$$\begin{aligned} \mathcal{J}_k(r,s) &\lesssim \mathrm{e}^{-(\mu_1-K_3)(r-s)}\mathbb{E}\big(1 + \big|X^\varepsilon(k\delta)\big|_H + \big|Y_s^{X^\varepsilon(k\delta),\widetilde{Y}^\varepsilon(k\delta)}\big|_H\big) \\ &\lesssim \mathrm{e}^{-(\mu_1-K_3)(r-s)}. \end{aligned} \tag{3.31}$$

Thus (3.23) follows by putting (3.31) into (3.25) and applying the Gronwall inequality in (3.24). Hence, we obtain that

$$\left(\mathbb{E}\left|X^\varepsilon(t) - \overline{X}(t)\right|_H^p\right)^{1/p} \lesssim \delta^\theta + \frac{\varepsilon}{\delta} + \sqrt{\frac{\varepsilon}{\delta}} + \varepsilon\delta^{-(1-\theta)}\mathrm{e}^{K_3\delta/\varepsilon}.$$

Letting $\delta := \varepsilon(-\ln\varepsilon)^{1/2}$ and then taking $\varepsilon \to 0$ yields the desired assertion, as required. □

Remark 3.2. In this section, we show an averaging result for a class of two-time-scale SPDEs driven by cylindrical stable noises in the abstract setting. Therefore, stochastic evolution equations of parabolic type with slow and fast time scales fit into our framework.

Remark 3.3. If $\alpha = 2$ and $\beta = 2$ in Theorem 3.1, which corresponds to the cylindrical Wiener noises, by reexamining the argument of Theorem 3.1, the boundedness of b can be removed by imposing, for example,

$$\left|f(x, y)\right|_H \le c_1 + c_2|y|, \qquad x, y \in H$$

for some appropriate constants $c_1, c_2 > 0$, that is, f is uniformly bounded w.r.t. the first variable. Moreover, by a close inspection of argument of Theorem 3.1, the boundedness of second moment of X^ε plays an important role in error analysis. However, for the case $\alpha, \beta \in (1, 2)$, $X^\varepsilon(\cdot)$ only has the pth moment with $p \in (1, \alpha)$. Therefore, for the technical reason, it seems hard to show Theorem 3.1 without the uniform boundedness of the nonlinearity. However, for the weak convergence (e.g., convergence in probability) of averaging principle for systems (3.1) and (3.2), the boundedness of the nonlinearity can be removed. Such result will be reported in our forthcoming paper.

Remark 3.4. In this section, we aim to obtaining averaging principles for a class of SPDEs driven by α-stable noise with $\alpha \in (1, 2]$. However, for the case $\alpha \in (0, 1)$ the method of this paper does not work. For such a case, it is necessary to find new approaches for the investigation.

Acknowledgements

We are indebted to the referee for his/her valuable comments which have greatly improved our earlier version.

The research of J. Bao was supported in part by the National Natural Science Foundation of China under Grant 11401592; the research of G. Yin was supported in part by the U.S. Army Research Office under Grant W911NF-15-1-0218; the research of C. Yuan was supported in part by the EPSRC and NERC.

References

[1] Anabtawi, M.J. (2011). Practical stability of nonlinear stochastic hybrid parabolic systems of Itô-type: Vector Lyapunov functions approach. *Nonlinear Anal. Real World Appl.* **12** 1386–1400. MR2781865

[2] Applebaum, D. (2009). *Lévy Processes and Stochastic Calculus*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **116**. Cambridge: Cambridge Univ. Press. MR2512800

[3] Bao, J., Mao, X. and Yuan, C. (2012). Lyapunov exponents of hybrid stochastic heat equations. *Systems Control Lett*. **61** 165–172. MR2878702

[4] Blömker, D., Hairer, M. and Pavliotis, G.A. (2007). Multiscale analysis for stochastic partial differential equations with quadratic nonlinearities. *Nonlinearity* **20** 1721–1744. MR2335080

[5] Bréhier, C.-E. (2012). Strong and weak orders in averaging for SPDEs. *Stochastic Process*. *Appl*. **122** 2553–2593. MR2926167

[6] Cerrai, S. (2009). A Khasminskii type averaging principle for stochastic reaction–diffusion equations. *Ann*. *Appl*. *Probab*. **19** 899–948. MR2537194

[7] Cerrai, S. and Freidlin, M. (2009). Averaging principle for a class of stochastic reaction–diffusion equations. *Probab*. *Theory Related Fields* **144** 137–177. MR2480788

[8] Da Prato, G., Jentzen, A. and Röckner, M. A mild Itô formula for SPDEs. Preprint. Available at arXiv:1009.3526v4.

[9] Dong, Z., Xu, L. and Zhang, X. (2014). Exponential ergodicity of stochastic Burgers equations driven by α-stable processes. *J*. *Stat*. *Phys*. **154** 929–949. MR3164597

[10] E, W., Liu, D. and Vanden-Eijnden, E. (2005). Analysis of multiscale methods for stochastic differential equations. *Comm*. *Pure Appl*. *Math*. **58** 1544–1585. MR2165382

[11] Freidlin, M.I. and Wentzell, A.D. (1998). *Random Perturbations of Dynamical Systems*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* **260**. New York: Springer. MR1652127

[12] Fu, H. and Duan, J. (2011). An averaging principle for two-scale stochastic partial differential equations. *Stoch*. *Dyn*. **11** 353–367. MR2836531

[13] Fu, H. and Liu, J. (2011). Strong convergence in stochastic averaging principle for two time-scales stochastic partial differential equations. *J*. *Math*. *Anal*. *Appl*. **384** 70–86. MR2822851

[14] Givon, D. (2007). Strong convergence rate for two-time-scale jump–diffusion stochastic differential systems. *Multiscale Model*. *Simul*. **6** 577–594 (electronic). MR2338495

[15] Griego, R.J. and Hersh, R. (1969). Random evolutions, Markov chains, and systems of partial differential equations. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **62** 305–308. MR0270207

[16] Hamilton, J.D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *J*. *Econometrics* **64** 307–333.

[17] Hersh, R. (1974). Random evolutions: A survey of results and problems. *Rocky Mountain J*. *Math*. **4** 443–477. MR0394877

[18] Ikeda, N. and Watanabe, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed. *North-Holland Mathematical Library* **24**. Amsterdam: North-Holland. MR1011252

[19] Khasminskii, R.Z. (1966). On stochastic processes defined by differential equations with a small parameter. *Theory Probab*. *Appl*. **11** 211–228.

[20] Khasminskii, R.Z. (1968). On an averaging principle for Itô stochastic differential equations. *Kibernetica* **4** 260–279.

[21] Khasminskii, R.Z. and Yin, G. (2004). On averaging principles: An asymptotic expansion approach. *SIAM J*. *Math*. *Anal*. **35** 1534–1560 (electronic). MR2083789

[22] Khasminskii, R.Z. and Yin, G. (2005). Limit behavior of two-time-scale diffusions revisited. *J*. *Differential Equations* **212** 85–113. MR2130548

[23] Kuksin, S.B. and Piatnitski, A.L. (2008). Khasminskii–Whitham averaging for randomly perturbed KdV equation. *J*. *Math*. *Pures Appl*. (9) **89** 400–428. MR2401144

[24] Kushner, H.J. (1984). *Approximation and Weak Convergence Methods for Random Processes*, *with Applications to Stochastic Systems Theory*. *MIT Press Series in Signal Processing*, *Optimization*, *and Control*, 6. Cambridge, MA: MIT Press. MR0741469

[25] Mao, X. and Yuan, C. (2006). *Stochastic Differential Equations with Markovian Switching*. London: Imperial College Press. MR2256095

[26] Mariton, M. (1990). *Jump Linear Systems in Automatic Control*. New York: Dekker.

[27] Maslowski, B., Seidler, J. and Vrkoč, I. (1991). An averaging principle for stochastic evolution equations. II. *Math*. *Bohem*. **116** 191–224. MR1112004

[28] Ouyang, S.-X. (2009). Harnack inequalities and applications for stochastic equations. Ph.D. thesis, Bielefeld Univ.

[29] Prévôt, C. and Röckner, M. (2007). *A Concise Course on Stochastic Partial Differential Equations*. *Lecture Notes in Math*. **1905**. Berlin: Springer. MR2329435

[30] Priola, E. and Zabczyk, J. (2011). Structural properties of semilinear SPDEs driven by cylindrical stable processes. *Probab. Theory Related Fields* **149** 97–137. MR2773026

[31] Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes*: *Stochastic Models with Infinite Variance*. *Stochastic Modeling*. New York: Chapman & Hall. MR1280932

[32] Sethi, S.P. and Zhang, Q. (1994). *Hierarchical Decision Making in Stochastic Manufacturing Systems*. *Systems & Control*: *Foundations & Applications*. Boston, MA: Birkhäuser. MR1301778

[33] Wang, F.-Y. (2011). Gradient estimate for Ornstein–Uhlenbeck jump processes. *Stochastic Process*. *Appl*. **121** 466–478. MR2763092

[34] Yin, G.G. and Zhang, Q. (1998). *Continuous-Time Markov Chains and Applications*: *A Singular Perturbation Approach*. *Applications of Mathematics* (*New York*) **37**. New York: Springer. MR1488963

[35] Yin, G.G. and Zhu, C. (2010). *Hybrid Switching Diffusions*: *Properties and Applications*. *Stochastic Modelling and Applied Probability* **63**. New York: Springer. MR2559912

[36] Zhang, X. (2013). Derivative formulas and gradient estimates for SDEs driven by α-stable processes. *Stochastic Process*. *Appl*. **123** 1213–1228. MR3016221

Received July 2013 and revised March 2014