SIAM J. NUMER. ANAL.
Vol. 57, No. 4, pp. 1815–1841

STRONG AND WEAK CONVERGENCE RATES OF A SPATIAL APPROXIMATION FOR STOCHASTIC PARTIAL DIFFERENTIAL EQUATION WITH ONE-SIDED LIPSCHITZ COEFFICIENT*

JIANBO CUI† AND JIALIN HONG‡

Abstract. Strong and weak approximation errors of a spatial finite element method are analyzed for the stochastic partial differential equations (SPDEs) with one-sided Lipschitz coefficients, including the stochastic Allen–Cahn equation, driven by additive noise. In order to give the strong convergence rate of the finite element method, we present an appropriate decomposition and some a priori estimates of the discrete stochastic convolution. To the best of our knowledge, there have been no essentially sharp weak convergence rates of spatial approximations for parabolic SPDEs with non-globally Lipschitz coefficients. To investigate the weak error, we first regularize the original equation by the splitting technique and obtain the regularity estimate the corresponding regularized Kolmogorov equation. Meanwhile, we present the refined estimates and the regularity estimate in the Malliavin sense of the finite element methods. Combining with the regularity of regularized Kolmogorov equation and Malliavin integration by parts, the weak convergence rate is shown to be twice the strong convergence rate.

Key words. one-sided Lipschitz coefficient, stochastic Allen–Cahn equation, finite element method, strong and weak convergence rate, Kolmogorov equation, Malliavin calculus

AMS subject classifications. 60H35, 60H15, 65M15

DOI. 10.1137/18M1215554

1. Introduction.

Both strong and weak convergence rates of numerical approximations for stochastic partial differential equations (SPDEs) with globally Lipschitz continuous and regular nonlinearities have been studied in recent decades. In contrast to the Lipschitz coefficient case, strong and weak convergence rates of numerical approximations for SPDEs with non-globally Lipschitz continuous nonlinearities, especially the stochastic Allen–Cahn equation, have become more involved recently (see, e.g., [3, 4, 5, 6, 8, 13, 23, 24, 25, 26, 33]) and are far from well understood. We refer to [4, 5, 6, 24, 25, 26, 33] and references therein for the strong convergence rate results of many different temporal and spatial approximations and to [9, 12] for the weak convergence rate results of temporal splitting type schemes. Up to now, there has been no essentially sharp weak convergence rate result of spatial approximation for parabolic SPDEs with non-globally Lipschitz coefficients. The present work makes further contributions on the strong and weak convergence rates of spatial approximations for SPDEs with non-globally Lipschitz continuous nonlinearities but one-sided Lipschitz nonlinearities driven by additive noise.

Let $\mathcal{O} = [0, L]$ and $\mathbb{H} = L^2(\mathcal{O})$ be the real separable Hilbert space endowed with the usual inner product. In this article, we mainly focus on the following semilinear parabolic SPDE:

*Received by the editors September 20, 2018; accepted for publication (in revised form) May 23, 2019; published electronically July 25, 2019.
https://doi.org/10.1137/18M1215554
Funding: The work of the authors was supported by National Natural Science Foundation of China grants 91630312, 91530118, 11021101, and 11290142.
†Corresponding author. LSEC, ICMSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and School of Mathematical Science, University of Chinese Academy of Sciences, Beijing 100049, China (jianbocui@lsec.cc.ac.cn).
‡LSEC, ICMSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and School of Mathematical Science, University of Chinese Academy of Sciences, Beijing, 100049, China (hjl@lsec.cc.ac.cn).

$$(1) \qquad \begin{aligned} dX(t) + AX(t)dt &= F(X(t))dt + dW(t), \quad t \in [0,T], \\ X(0) &= X_0, \end{aligned}$$

where $0 < T < \infty$, $-A$ is the Laplacian operator on $\mathcal{O}$ under homogeneous Dirichlet boundary condition, and F is the Nemytskii operator defined by $F(X)(\xi) := f(X(\xi)), \xi \in \mathcal{O}$, where f is a real-valued nonlinear function and satisfies Assumption 2.3. In particular, (1) is the stochastic Allen–Cahn equation if $F(X) = X - X^3$. The stochastic process $\{W(t)\}_{t\geq 0}$ is a generalized Q-Wiener process on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. Under further assumptions on X_0, Q, f, and $\|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0} < \infty$, $\beta \in (0,1]$, similar arguments in [6, 26] yield that there is a unique mild solution X of (1), which possesses the optimal spatial regularity $\mathbb{E}[\|X(t)\|_{\mathbb{H}^\beta}^p] \leq C(T, Q, X_0, p)$, $p \geq 1$. For the numerical study of SPDEs with one-sided Lipschitz coefficient driven by the multiplicative noise under enough spatial regularity assumptions, we refer to [19, 25]. In this work, we do not consider the case of the multiplicative noise with low spatial regularity assumption, since it is more involved and beyond the scope of this article.

One main contribution of this article is applying the variational approach, combined with an appropriate error decomposition, to deduce the strong convergence rate of the spatial finite element method for (1) with one-sided Lipschitz coefficients under the mild assumption on X_0. The corresponding finite element approximation X^h satisfies

$$(2) \qquad \begin{aligned} dX^h(t) + A_h X^h(t)dt &= P^h F(X^h(t))dt + P^h dW(t), \\ X(0) &= X_0^h, \end{aligned}$$

where P^h is the Galerkin finite element projection and A_h is the discretization of A. Recently, the authors in [19] proved strong convergence with rates of the finite element method for the stochastic Allen–Cahn equation with gradient-type multiplicative noise. The authors in [26] deduced the strong convergence rate of the finite element method for the stochastic Allen–Cahn equation driven by additive trace-class noise. To the best of our knowledge, there exists no sharp strong convergence rate result of the finite element method approximating (1) driven by general additive noise. As the considered noise in (1) could be very rough, an a priori estimate of stochastic convolution is needed. We make use of the properties of S^h and P^h to get the nonuniform estimate of the approximated stochastic convolution Z^h, and obtain the sharp strong convergence rate, for $X_0 \in \mathcal{C}(\mathcal{O})$, $T > 0$, $p \geq 1$,

$$\mathbb{E}\Big[\|X(T) - X^h(T)\|_{\mathbb{H}}^p\Big] \leq C(X_0, T, p, \gamma)\left(1 + T^{-\frac{\gamma}{2}}\right)^p h^{\gamma p},$$

where $\gamma \leq \beta$, if $\beta \in (\frac{1}{2}, 1]$ and $\gamma < \beta$, if $\beta \in (0, \frac{1}{2}]$. We remark that this approach to deducing the strong convergence rate of the numerical approximation is also available for the more general case (see Remark 3.3).

Another main contribution is about the weak convergence rate of the finite element method for (1) with one-sided Lipschitz coefficient. In recent years, there have been many different strategies for the weak error analysis for many different numerical schemes approximating parabolic SPDEs with Lipschitz and regular coefficients. We refer to, e.g., [2, 7, 18, 21, 22] for the weak error analysis based on the associated Kolmogorov equation, to, e.g., [10, 20] for applying the mild Itô formula approach, and to, e.g., [1, 32] for other techniques. However, no essentially sharp weak convergence rate of spatial approximation is established for parabolic SPDEs with non-globally

Lipschitz coefficients. There are three key points to deducing the weak convergence rate of numerical approximations for (1) with non-sided Lipschitz coefficients: to give the regularity estimates of the corresponding Kolmogorov equation, to deduce the uniform estimate of the spatial approximation, and to get rid of the irregular terms in the weak error estimate. Inspired by [9], where the authors shows the weak convergence order of the two temporal splitting type schemes approximating the stochastic Allen–Cahn equation driven by space-time white noise, we propose a regularizing procedure through a splitting strategy. Then we utilize the properties of S^h, P^h, and A_h (see section 2), as well as the nonuniform estimate of the approximated stochastic convolution Z^h, to get an a priori estimate of the finite element approximation. Finally, by using the Malliavin integration by parts, together with the regularity estimates of the regularized Kolmogorov equation and an a priori estimate of the finite element approximation, we derive the essentially sharp weak convergence rate result of X^h, for $\phi \in \mathcal{C}_b^2(\mathbb{H})$, $X_0 \in \mathcal{C}(\mathcal{O})$, $T > 0$, $\gamma < \beta$,

$$\Big|\mathbb{E}\Big[\phi(X(T)) - \phi(X^h(T))\Big]\Big| \le C(X_0, T, \gamma, \phi)(1 + T^{-\gamma})h^{2\gamma}.$$

The outline of this paper is as follows. In the next section, some preliminaries are listed. Section 3 is devoted to giving the a priori estimates of (1), the strong convergence rate of the finite element method, and the a priori estimates of the finite element method and semidiscretized stochastic convolution. In section 4, we propose a new regularizing procedure and give an approach to studying the weak convergence rate of the finite element method by Malliavin calculus.

2. Preliminaries. In this section, we give assumptions on A, F, and $W(t)$, the abstract functional analytical framework of the considered equation and the finite element method, and a brief introduction to Malliavin calculus.

Given two separable Hilbert spaces $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ and $(\widetilde{H}, \|\cdot\|_{\widetilde{H}})$, denote by $\mathcal{L}(\mathcal{H}, \widetilde{H})$ and $\mathcal{L}_1(\mathcal{H}, \widetilde{H})$ the Banach spaces of all linear bounded operators and the nuclear operators from $\mathcal{H}$ to $\widetilde{H}$, respectively. The trace of an operator $\mathcal{T} \in \mathcal{L}_1(\mathcal{H})$ is $\mathrm{tr}[\mathcal{T}] = \sum_{k\in\mathbb{N}^+}\langle \mathcal{T} f_k, f_k\rangle_{\mathcal{H}}$, where $\{f_k\}_{k\in\mathbb{N}^+}$ ($\mathbb{N}^+ = \{1, 2, \dots\}$) is any orthonormal basis of $\mathcal{H}$. In particular, if $\mathcal{T} \ge 0$, $\mathrm{tr}[\mathcal{T}] = \|\mathcal{T}\|_{\mathcal{L}_1}$. Denote by $\mathcal{L}_2(\mathcal{H}, \widetilde{H})$ the space of Hilbert–Schmidt operators from $\mathcal{H}$ into $\widetilde{H}$, equipped with the usual norm given by $\|\cdot\|_{\mathcal{L}_2(\mathcal{H},\widetilde{H})} = (\sum_{k\in\mathbb{N}^+}\|\cdot f_k\|^2_{\widetilde{H}})^{\frac12}$. The following useful property and inequality (see, e.g., [2]) hold:

$$\langle \mathcal{T}, \mathcal{S}\rangle_{\mathcal{L}_2(\mathcal{H},\widetilde{H})} = \mathrm{tr}[\mathcal{T}^*\mathcal{S}] = \mathrm{tr}[\mathcal{S}\mathcal{T}^*], \quad \mathcal{T},\ \mathcal{S} \in \mathcal{L}_2(\mathcal{H}, \widetilde{H}), \tag{3}$$
$$|\mathrm{tr}[\mathcal{S}\mathcal{T}^*]| \le \|\mathcal{S}\mathcal{T}^*\|_{\mathcal{L}_1} \le \|\mathcal{S}\|\|\mathcal{T}\|_{\mathcal{L}_1}, \quad \mathcal{S} \in \mathcal{L}(\mathcal{H}, \widetilde{H}),\ \mathcal{T} \in \mathcal{L}_1(\mathcal{H}, \widetilde{H}),$$

where $\mathcal{T}^*$ is the adjoint operator of $\mathcal{T}$. Let $\mathcal{C}_b^k(\mathcal{H}, \widetilde{H})$, $k \in \mathbb{N}^+$, be the space of k times continuous differentiable mappings from $\mathcal{H}$ to $\widetilde{H}$ with continuous and bounded derivatives up to order k. We endow $\mathcal{C}_b^k(\mathcal{H}, \widetilde{H})$ with the seminorm $|\cdot|_{\mathcal{C}_b^k(\mathcal{H},\widetilde{H})}$, which is defined as $|g|_{\mathcal{C}_b^k(\mathcal{H},\widetilde{H})}$ the smallest constant for $g \in \mathcal{C}_b^k(\mathcal{H}, \widetilde{H})$ such that

$$\sup_{x\in\mathcal{H}}\|D^m g(x)\cdot(\phi_1, \dots, \phi_m)\|_{\widetilde{H}} \le C\|\phi_1\|_{\mathcal{H}}\dots\|\phi_m\|_{\mathcal{H}} \quad \forall \phi_1, \cdots, \phi_m \in \mathcal{H}, m \le k.$$

Given a Banach space $(\mathcal{E}, \|\cdot\|_{\mathcal{E}})$, we denote by $\gamma(\mathcal{H}, \mathcal{E})$ the space of γ-radonifying operators endowed with the norm defined by $\|\mathcal{T}\|_{\gamma(\mathcal{H},\mathcal{E})} = (\widetilde{\mathbb{E}}\|\sum_{k\in\mathbb{N}^+}\gamma_k\mathcal{T} f_k\|^2_{\mathcal{E}})^{\frac12}$, where $\{\gamma_k\}_{k\in\mathbb{N}^+}$ is a Rademacher sequence on a probability space $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathbb{P}})$. For convenience, we denote $\|\cdot\| = \|\cdot\|_{\mathbb{H}}$ and $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_{\mathbb{H}}$. Let $L^q = L^q(\mathcal{O})$, $1 \le q < \infty$, and $E = \mathcal{C}(\mathcal{O})$ equipped with the usual norms. We also need the following Burkholder inequality in martingale-type 2 Banach spaces (see, e.g., [30, Lemma 2.1]): for $L^q, q \in [2,\infty)$ and $p \ge 1$, there exists $C_{p,q} \in (0,\infty) > 0$ such that

$$\begin{aligned}\left\|\sup_{t\in[0,T]}\left\|\int_0^t \phi(r)d\widetilde{W}(r)\right\|_{L^q}\right\|_{L^p(\Omega)} &\le C_{p,q}\|\phi\|_{L^p(\Omega;L^2([0,T];\gamma(\mathbb{H};L^q)))}\\ &\le C_{p,q}\left(\mathbb{E}\left(\int_0^T\left\|\sum_{k\in\mathbb{N}^+}(\phi(t)e_k)^2\right\|_{L^{\frac{q}{2}}}dt\right)^{\frac{p}{2}}\right)^{\frac{1}{p}},\end{aligned} \tag{4}$$

where $\{\widetilde{W}(t)\}_{t\ge 0}$ is the $\mathbb{H}$-valued cylindrical Wiener process and $\{e_k\}_{k\in\mathbb{N}^+}$ is an orthonormal basis of $\mathbb{H}$.

2.1. Main assumptions. In this subsection, we introduce some useful notations and our main assumptions on A, F, and W. Throughout this article, the initial datum X_0 is assumed to be a deterministic function and belongs to E for convenience. We use C to denote a generic constant, independent of h, which differs from one place to another.

Assumption 2.1. Let $\mathcal{O} = (0,L), L > 0$, and $-A : D(A) \subset \mathbb{H} \to \mathbb{H}$ be the Laplacian operator under the homogeneous Dirichlet boundary condition, i.e., $Au = -\Delta u, u \in D(A)$.

Such an assumption implies that $-A$ generates an analytic and contraction C_0-semigroup $S(t), t \ge 0$, in $\mathbb{H}$ and L^q. It is also well known that the assumption on $\mathcal{O}$ implies that the existence of the eigensystem $\{\lambda_k, e_k\}_{k\in\mathbb{N}^+}$ of $\mathbb{H}$, such that $\lambda_k > 0$, $Ae_k = \lambda_k e_k$, and $\lim_{k\to\infty}\lambda_k = \infty$. Let $\mathbb{H}^r$ be the Hilbert space equipped with the norm $\|\cdot\|_{\mathbb{H}^r} := \|A^{\frac{r}{2}}\cdot\|_{\mathbb{H}}$ for the fractional power $A^{\frac{r}{2}}, r \ge 0$.

Assumption 2.2. Let $W(t)$ be a Wiener process with covariance operator Q, where Q is a bounded, linear, self-adjoint, and positive definite operator on $\mathbb{H}$. Assume that $\|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0} < \infty$ with $0 < \beta \le 1$, where $\mathcal{L}_2^0 = \mathcal{L}_2(U_0,\mathbb{H})$, $U_0 = Q^{\frac{1}{2}}(\mathbb{H})$. In the case that $\beta \le \frac{1}{2}$, in addition assume that Q commutes with A.

Assumption 2.3. Let $K \in \mathbb{N}^+$ and $L_f > 0$. Assume that $f : \mathbb{R} \to \mathbb{R}$ satisfies

$$|f(\xi)| \le L_f\left(1+|\xi|^K\right), \quad f'(\xi) \le L_f, \quad |f'(\xi)| \le L_f\left(1+|\xi|^{K-1}\right).$$

Let $F : L^{2K} \to \mathbb{H}$ be the Nemytskii operator defined by $F(X)(\xi) = f(X(\xi))$.

The above assumption ensures that $F : L^{2K} \to \mathbb{H}$ satisfies for some constant $L = L(L_f, K)$,

$$\begin{aligned}\langle F(u)-F(v), u-v\rangle &\le L_f\|u-v\|^2,\ u,v \in L^{2K},\\ \|F(u)-F(v)\| &\le L\left(1+\|u\|_E^{K-1}+\|v\|_E^{K-1}\right)\|u-v\|,\ u,v \in E,\end{aligned}$$

where $\|\cdot\|_E$ is the supremum norm. We remark that in the analysis of strong convergence rates, the assumption about the upper bound of the derivative of f could be weakened to the monotone condition. We also point out that when studying the weak convergence rates, a more restricted condition on F is needed. The typical example of f is a cubic polynomial

$$f(\xi) = a_3\xi^3 + a_2\xi^2 + a_1\xi + a_0,\ \ a_3 < 0,\ a_2,\ a_1,\ a_0 \in \mathbb{R}.$$

In this case, (1) corresponds to the stochastic Allen–Cahn equation. We remark that Assumptions 2.1–2.3 could be extend to $d \leq 3$ and the more general noise case (see Remark 3.3 in section 3). We also mention that the weak convergence rate of a full discretization of (1) is studied in [16].

2.2. Finite element method. Let $(T_h)_{h\in(0,1)}$ be a quasi-uniform family of triangulations of $\mathcal{O}$, i.e., T_h is a partition of $\mathcal{O}$, the parameter h is the mesh size of T_h, and the length of each subinterval is bounded below by ch for a constant $c > 0$. Let $(V_h)_{h\in(0,1)}$ be a family of spaces of continuous piecewise linear functions corresponding to $(T_h)_{h\in(0,1)}$, and let N_h be the dimension of V_h. Denote $P^h : \mathbb{H} \to V_h$ the orthogonal projection and $A_h : V_h \to V_h$ the discrete Laplacian satisfying $\langle A_h u, v\rangle = \langle \nabla u, \nabla v\rangle, u, v \in V_h$. It is well known that the semidiscretization $-A_h u^h = P^h f$ is a finite element approximation of $-Au = f$ and that $\|u - u^h\| = \|A_h^{-1}P^h f - A^{-1}f\| \leq Ch^2\|f\|$ (see, e.g., [28]). The operator $-A_h$ generates an analytic semigroup $(S^h(t))_{t\geq 0}$. In particular, there is an orthonormal eigenbasis $\{e_i^h\}_{i=1}^{N_h}$ in V^h equipped with the $\mathbb{H}$ norm, with eigenvalues $0 < \lambda_1^h \leq \lambda_2^h \leq \cdots \leq \lambda_{N_h}^h$ such that

$$S^h(t)v^h = \sum_{i=1}^{N_h} e^{-\lambda_i^h t}\langle v^h, e_i^h\rangle e_i^h, \quad v^h \in V_h, \ \ t \geq 0.$$

We will often use the equivalence of the two norms for $v^h \in V_h$, $\gamma \in [-\frac{1}{2}, \frac{1}{2}]$,

$$c\|A_h^\gamma v^h\| \leq \|A^\gamma v^h\| \leq C\|A_h^\gamma v^h\|, \tag{5}$$

the interpolation space $(\mathbb{H}_h^\beta)_{\beta\in[-1,1]}$, and the properties of the Ritz projection $R^h : \mathbb{H}^1 \to V_h$ and P^h (see, e.g., [2], [28, Chapter 3]),

$$\begin{aligned} &\|A^{\frac{s}{2}}(I - R^h)A^{-\frac{r}{2}}\|_{\mathcal{L}(\mathbb{H})} \leq Ch^{r-s}, \quad 0 \leq s \leq 1 \leq r \leq 2,\\ &\|A^{\frac{s}{2}}(I - P^h)A^{-\frac{r}{2}}\|_{\mathcal{L}(\mathbb{H})} \leq Ch^{r-s}, \quad 0 \leq s \leq 1, 0 \leq s \leq r \leq 2. \end{aligned} \tag{6}$$

In the setting of the strong convergence rate analysis, we will need the error of the semigroups $G^h(t) := S^h(t)P^h - S(t)$, $t \geq 0$ (see, e.g., [26, section 3] or [28, Chapter 3]) for $h \in (0, 1]$,

$$\begin{aligned} \|G^h(t)x\| &\leq Ch^u t^{-\frac{u-v}{2}}\|x\|_{\mathbb{H}^v}, \ x \in \mathbb{H}^v, \ t > 0, \ 0 \leq v \leq u \leq 2,\\ \|G^h(t)x\| &\leq Ct^{\frac{\rho}{2}}\|x\|_{\mathbb{H}^{-\rho}}, \ x \in \mathbb{H}^{-\rho}, t > 0, \ 0 \leq \rho \leq 1,\\ \|G^h(t)x\| &\leq Ct^{-1}h^{2-\rho}\|x\|_{\mathbb{H}^{-\rho}}, \ x \in \mathbb{H}^{-\rho}, t > 0, \ 0 \leq \rho \leq 1,\\ \left\|\int_0^t G^h(s)x ds\right\| &\leq Ch^{2-\rho}\|x\|_{\mathbb{H}^{-\rho}}, \ x \in \mathbb{H}^{-\rho}, \ t > 0, \ 0 \leq \rho \leq 1,\\ \left(\int_0^t \|G^h(s)x\|^2 ds\right)^{\frac{1}{2}} &\leq Ch^{1+\rho}\|x\|_{\mathbb{H}^\rho}, \ x \in \mathbb{H}^\rho, \ t > 0, \ 0 \leq \rho \leq 1. \end{aligned} \tag{7}$$

Besides the above properties of finite element methods, the other important parts for our analysis are the smoothy effect of S^h (see, e.g., [28, Chapter 3])

$$\|A_h^{\gamma} S^h(t) P_h\|_{\mathcal{L}(\mathbb{H})} \le C_\gamma t^{-\gamma}, \ \ \gamma \ge 0, \ t > 0, \tag{8}$$
$$\int_0^t \left\| A_h^{\frac{1}{2}} S^h P_h x \right\|^2 ds \le C\|x\|^2, \ x \in \mathbb{H},$$

and the boundedness of P^h (see, e.g., [29, Lemma 2.3]),

$$\|P^h\|_{\mathcal{L}(L^p)} \le C, \ \ 1 \le p < \infty, \quad \|P^h\|_{\mathcal{L}(E)} \le C.$$

2.3. Malliavin calculus. In order to get the weak convergence rate, we recall some preliminaries about Malliavin calculus in Hilbert space (see, e.g., [2, section 2]), which will be used to deal with the singular term appeared in the weak error. Since Q is a bounded, linear, self-adjoint, and positive definite operator on $\mathbb{H}$, the corresponding Cameron–Martin space is $U_0 = Q^{\frac{1}{2}}(\mathbb{H})$. Let $\mathcal{I} : L^2([0,T];U_0) \to L^2(\Omega)$ be an isonormal process, i.e., for any $\psi \in L^2([0,T];U_0)$, $\mathcal{I}(\psi)$ is the centered Gaussian variable and $\mathbb{E}[\mathcal{I}(\psi_1)\mathcal{I}(\psi_2)] = \langle \psi_1, \psi_2 \rangle_{L^2([0,T];U_0)}$, $\psi_1, \psi_2 \in L^2([0,T];U_0)$. Let $\mathcal{C}_p^\infty(\mathbb{R}^N)$ be the space of all real-valued $\mathcal{C}^\infty$ functions on $\mathbb{R}^N$ with polynomial growth. We denote the family of smooth real-valued cylindrical random variables by

$$\mathcal{S} = \left\{ \mathcal{X} = g(\mathcal{I}(\psi_1), \dots, \mathcal{I}(\psi_N)) : g \in \mathcal{C}_p^\infty(\mathbb{R}^N), \psi_j \in L^2([0,T];U_0), j = 1, \dots, N \right\}$$

and the family of smooth cylindrical $\mathbb{H}$-valued random variables by

$$\mathcal{S}(\mathbb{H}) = \left\{ G = \sum_{i=1}^M \mathcal{X}_i \otimes h_i : \mathcal{X}_i \in \mathcal{S}, h_i \in \mathbb{H}, M \ge 1 \right\}.$$

Then the Malliavin derivative of $G = \sum_{i=1}^M g_i(\mathcal{I}(\psi_1), \dots, \mathcal{I}(\psi_N) \otimes h_i$ is defined by

$$\mathcal{D}_s G = \sum_{i=1}^M \sum_{j=1}^N \partial_j g_i(\mathcal{I}(\psi_1), \dots, \mathcal{I}(\psi_N)) \otimes (h_i \otimes \psi_j(s)).$$

Since the derivative operator $\mathcal{D}$ is closable (see, e.g., [2, section 2]), we denote $\mathbb{D}^{1,2}(\mathbb{H})$ the closure of $\mathcal{S}(\mathbb{H})$ with respect to the Malliavin derivative equipped with the norm

$$\|G\|_{\mathbb{D}^{1,2}(\mathbb{H})} = \left(\mathbb{E}[\|G\|^2] + \mathbb{E}\left[\int_0^T \|\mathcal{D}_s G\|^2 ds \right] \right)^{\frac{1}{2}},$$

where $\mathcal{D}_s G$ is the Malliavin derivative of G. The key in the analysis of weak convergence rate is the following integration by parts formula (see, e.g., [18, section 2]). For any random variable $G \in \mathbb{D}^{1,2}(\mathbb{H})$ and any predictable process $\Theta \in L^2([0,T];\mathcal{L}_2^0)$, we have

$$\mathbb{E}\left[\left\langle \int_0^T \Theta(t) dW(t), G \right\rangle \right] = \mathbb{E}\left[\int_0^T \left\langle \Theta(t), \mathcal{D}_t G \right\rangle_{\mathcal{L}_2^0} dt \right]. \tag{9}$$

Moreover, we also need the chain rule of the Malliavin derivative. Let $\mathcal{V}$ be another separable Hilbert space and $\sigma \in \mathcal{C}_b^1(\mathbb{H}, \mathcal{V})$. Then we have $\sigma(G) \in \mathbb{D}^{1,2}(\mathcal{V})$,

$$\mathcal{D}_t^y(\sigma(G)) = \mathcal{D}\sigma(G) \cdot \mathcal{D}_t^y G, \quad y \in U_0, \quad G \in \mathbb{D}^{1,2}(\mathbb{H}),$$
$$\mathcal{D}_t(\sigma(G)) = \mathcal{D}\sigma(G) \mathcal{D}_t G, \quad G \in \mathbb{D}^{1,2}(\mathbb{H}),$$

where $\mathcal{D}_t^y G := \mathcal{D}_t G y$ is the derivative of G in the direction of $y \in U_0$.

3. A priori estimate and strong convergence rate. In this section, we present the strong convergence rate of the finite element method, as well as the a priori estimates of the discrete stochastic convolution and the finite element method.

3.1. A priori estimate. Combining the equivalence of (1) and the random PDE

$$\begin{aligned} dY + AY dt &= F(Y+Z)dt, \quad Y(0) = X_0, \\ dZ + AZ dt &= dW(t), \quad Z(0) = 0, \end{aligned}$$

with similar arguments in the proofs of [17, Theorem 7.7] and [6, Lemma 3.3], we get the following a priori estimate on the exact solution of (1).

LEMMA 3.1. *Under Assumptions* 2.1–2.3*, there exists a unique mild solution* X *of* (1)*. Moreover, for* $t \in (0,T]$*,* $p \geq 1$*, there exists a constant* $C(T,p) > 0$ *such that*

$$\begin{aligned} \sup_{s\in[0,t]} \mathbb{E}\Big[\|X(s)\|_E^p\Big] &\leq C(T,p)(1+\|X_0\|_E^p), \\ \mathbb{E}\Big[\|X(t)\|_{\mathbb{H}^\beta}^p\Big] &\leq (1+t^{-\frac{\beta p}{2}})C(T,p)(1+\|X_0\|^p). \end{aligned}$$

Now, we are in a position to derive an a priori estimate for the semidiscretization equation (2). First, we prove the smoothy property of $S^h(t)$, $t \geq 0$.

LEMMA 3.2. *For* $t > 0$ *and* $2 \leq p \leq \infty$*, there exists a positive constant* C *such that for* $f \in \mathbb{H}$*,*

$$\|S^h(t)P^h f\|_{L^p} \leq Ct^{-\frac{1}{2}(\frac{1}{2}-\frac{1}{p})}\|f\|.$$

Proof. Since $S^h(t)P^h f \in V^h$, we have

$$S^h(t)P^h f = \sum_{i=1}^{N^h} e^{-\lambda_i^h t}\langle f, e_i^h\rangle e_i^h.$$

Then the uniform boundness of e_i^h and $ci^2 \leq \lambda_i^h \leq Ci^2$, $1 \leq i \leq N^h$, in [2, section 2] yields that

$$\left\|S^h(t)P^h f\right\|_E = \left\|\sum_{i=1}^{N^h} e^{-\lambda_i^h t}\left\langle f, e_i^h\right\rangle e_i^h\right\|_E \leq \left(\sum_{i=1}^{N^h} e^{-2\lambda_i^h t}\right)^{\frac{1}{2}} \|f\| \leq Ct^{-\frac{1}{4}}\|f\|$$

and

$$\|S^h(t)P^h f\| \leq \left\|\sum_{i=1}^{N^h} e^{-\lambda_i^h t}\left\langle f, e_i^h\right\rangle e_i^h\right\| \leq C\|f\|.$$

The Riesz–Thorin interpolation theorem (see, e.g., [27]) leads to the desired result. ◻

The other tool to get the a priori estimate is the weak discrete maximum principle in [11, Lemma 3.4].

LEMMA 3.3. *Under the assumptions on* T^h *and* V_h*, there exists a positive constant* C *such that, for any* $v^h \in V_h$*,*

$$\|S^h(t)v^h\|_{L^\infty} \leq C\|v^h\|_{L^\infty}, \quad t > 0. \tag{10}$$

We remark that in the case of higher dimension, the similar boundedness results of finite element methods still hold (see, e.g., [28, Chapter 6]). Next, we give the a priori estimate of the semidiscretized stochastic convolution Z^h, which satisfies

$$dZ^h(t) + A_h Z^h(t) = P^h dW(t), \quad Z^h(0) = 0.$$

LEMMA 3.4. *Let $\mathcal{V} = E$ or L^{2q} $(q \geq 1)$. Under Assumptions* 2.1–2.2, *there exists a constant $C(T,p) > 0$ such that the discretized stochastic convolution Z^h satisfies that for $t \in (0,T]$ and $p \geq 1$,*

$$\mathbb{E}\left[\left\|Z^h(t)\right\|_{\mathcal{V}}^p\right] \leq C(T,p) \quad \text{if} \quad \beta > \frac{1}{2},$$
$$\mathbb{E}\left[\left\|Z^h(t)\right\|_{\mathcal{V}}^p\right] \leq C(T,p)\left(1 + \log\left(\frac{1}{h}\right)\right)^{\frac{p}{2}} \quad \text{if} \quad 0 < \beta \leq \frac{1}{2}.$$

Moreover, if $Q = I$, $\beta \in [0, \frac{1}{2})$, then for $t \in (0,T]$ and $p \geq 1$, there exists a constant $C(T,p)$ such that

$$\mathbb{E}\Big[\|Z^h(t)\|_{\mathcal{V}}^p\Big] \leq C(T,p).$$

Proof. The a priori estimate in the case that $\beta > \frac{1}{2}$ is directly proven by using the Sobolev embedding theorem, the Burkholder inequality, and the smoothing property of A_h (8). Now we focus on the case $\beta \leq \frac{1}{2}$ and take $\mathcal{V} = L^{2q}$, $q \geq 1$, as example. Similar arguments yield the case $\mathcal{V} = E$. Notice that $Z^h(t,\xi) = \sum_{k\in\mathbb{N}^+}\int_0^t \sum_{i=1}^{N^h} e^{-\lambda_i^h s}\langle\sqrt{q_k}e_k, e_i^h\rangle e_i^h(\xi)d\beta_k(s)$, where $\{e_k, q_k\}_{k\in\mathbb{N}^+}$ is the eigensystem of Q. The Fubini theorem, Fourier transform, Burkholder inequality (4), and uniform boundedness of $e_i^h, 1 \leq i \leq N^h$, in E (see, e.g., [31, Appendix]) yield that

$$\begin{aligned}
\mathbb{E}\Big[\|Z^h(t)\|_{L^{2q}}^p\Big] &\leq C\mathbb{E}\left[\left(\int_0^t \Big\|\sum_{k\in\mathbb{N}^+}\Big(\sum_{i=1}^{N^h} e^{-\lambda_i^h s}\langle\sqrt{q_k}e_k, e_i^h\rangle e_i^h\Big)^2\Big\|_{L^q} ds\right)^{\frac{p}{2}}\right] \\
&\leq C\mathbb{E}\left[\left(\int_0^t \Big\|\sum_{i,j=1}^{N^h} e^{-(\lambda_i^h+\lambda_j^h)s}\langle Q^{\frac{1}{2}}e_i^h, Q^{\frac{1}{2}}e_j^h\rangle e_i^h e_j^h\Big\|_{L^q} ds\right)^{\frac{p}{2}}\right] \\
&\leq C\mathbb{E}\left[\left(\int_0^t \sum_{i,j=1}^{N^h} e^{-(\lambda_i^h+\lambda_j^h)s}\|e_i^h\|_E\|e_i^h\|\|e_j^h\|_E\|e_j^h\| ds\right)^{\frac{p}{2}}\right] \\
&\leq C\mathbb{E}\left[\left(\int_0^t \sum_{i,j=1}^{N^h} e^{-(\lambda_i^h+\lambda_j^h)s} ds\right)^{\frac{p}{2}}\right].
\end{aligned}$$

By $ci^2 \leq \lambda_i^h \leq Ci^2, 1 \leq i \leq N^h$, and $N^h \leq \mathcal{O}(\frac{1}{h})$, we have

$$\begin{aligned}
\mathbb{E}\Big[\|Z^h(t)\|_{L^{2q}}^p\Big] &\leq C\mathbb{E}\left[\left(\int_0^t \sum_{i,j=1}^{N^h} e^{-(\lambda_i^h+\lambda_j^h)s} ds\right)^{\frac{p}{2}}\right] \leq C\mathbb{E}\left[\left(\int_0^t \sum_{i=1}^{(N^h)^2} e^{-ics} ds\right)^{\frac{p}{2}}\right] \\
&\leq C\mathbb{E}\left[\left(\int_0^{h^l} \frac{1}{h^2} ds\right)^{\frac{p}{2}}\right] + C\mathbb{E}\left[\left(\int_{h^l}^t \int_1^\infty e^{-c\xi s} d\xi ds\right)^{\frac{p}{2}}\right]
\end{aligned}$$

$$
\begin{aligned}
&\le C\left(h^{l-2}+\left(\log(1+t)+\log\left(\frac{1}{h}\right)\right)^{\frac{p}{2}}\right)\\
&\le C\left(1+(\log(1+t))^{\frac{p}{2}}+\left(\log\left(\frac{1}{h}\right)\right)^{\frac{p}{2}}\right)
\end{aligned}
$$

for a large $l\in\mathbb{N}^+$. In particular, if $Q=I$, then the logarithmic factor can be eliminated as

$$
\begin{aligned}
\mathbb{E}\Big[\|Z^h(t)\|_{L^{2q}}^p\Big] &\le C\mathbb{E}\left[\left(\int_0^t\left\|\sum_{k\in\mathbb{N}^+}\left(\sum_{i=1}^{N^h}e^{-\lambda_i^h s}\langle e_k,e_i^h\rangle e_i^h\right)^2\right\|_{L^q}ds\right)^{\frac{p}{2}}\right]\\
&\le C\mathbb{E}\left[\left(\int_0^t\left\|\sum_{k\in\mathbb{N}^+}\sum_{i,j=1}^{N^h}e^{-(\lambda_i^h+\lambda_j^h)s}\langle e_k,e_i^h\rangle\langle e_k,e_j^h\rangle e_i^h e_j^h\right\|_{L^q}ds\right)^{\frac{p}{2}}\right]\\
&\le C\mathbb{E}\left[\left(\int_0^t\left\|\sum_{i,j=1}^{N^h}e^{-(\lambda_i^h+\lambda_j^h)s}\langle e_i^h,e_j^h\rangle e_i^h e_j^h\right\|_{L^q}ds\right)^{\frac{p}{2}}\right]\\
&\le C\mathbb{E}\left[\left(\int_0^t\left\|\sum_{i=1}^{N^h}e^{-2\lambda_i^h s}(e_i^h)^2\right\|_{L^q}ds\right)^{\frac{p}{2}}\right]\le Ct^{\frac{p}{4}}.
\end{aligned}
$$

Summing up all the estimates, we finish the proof. ∎

The following a priori estimate is very useful for deducing the weak convergence rate in section 4 and has its own interest.

PROPOSITION 3.1. *Let $\mathcal{V}=E$ or L^{2q} $(q\ge 1)$. Under Assumptions* 2.1–2.3 *with $K<5$, there exists $C(p,T,X_0)>0$ such that the unique mild solution X^h of* (2) *satisfies*

$$
\sup_{t\in[0,T]}\mathbb{E}\Big[\|X^h(t)\|_{\mathcal{V}}^p\Big]\le C(p,T,X_0)\left(1+\log\left(\frac{1}{h}\right)\right)^{\frac{K^2p}{2}}\quad \textit{for}\ \ p\ge 1.
$$

Proof. For convenience, we only prove the case $\mathcal{V}=E$. By the equivalence of the SPDE

$$
dX^h+A_hX^hdt=P^hF(X^h)dt+P^hdW(t),X^h(0)=P^hX_0,
$$

and the random PDE

$$
\begin{aligned}
&dY^h+A_hY^hdt=P^hF(Y^h+Z^h)dt,Y^h(0)=P^hX_0,\\
&dZ^h+A_hZ^hdt=P^hdW(t),Z^h(0)=0,
\end{aligned}
$$

and Lemma 3.4, it suffices to bound $\mathbb{E}[\|Y^h(t)\|_E^p]$. The higher regularity of Y^h and the dissipativity of F imply that

$$\begin{aligned}
&\|Y^h(t)\|^2 + 2\int_0^t \left\langle \nabla Y^h, \nabla Y^h \right\rangle ds \\
&\quad = \|X^h(0)\|^2 + 2\int_0^t \left\langle F\left(Y^h + Z^h\right), Y^h \right\rangle ds \\
&\quad = \|X^h(0)\|^2 + 2\int_0^t \left\langle F\left(Y^h + Z^h\right) - F(Z^h), Y^h \right\rangle ds + 2\int_0^t \left\langle F\left(Z^h\right), Y^h \right\rangle ds \\
&\quad \leq \|X^h(0)\|^2 + C\int_0^t \left\|Y^h\right\|^2 ds + C\int_0^t \left(1 + \left\|Z^h\right\|_{L^{2K}}^{2K}\right) ds.
\end{aligned}$$

The p-moment boundedness of $\|Z^h\|_E$ and Gronwall's inequality yield that for $p \geq 1$,

$$\begin{aligned}
&\sup_{t\in[0,T]} \|Y^h(t)\|^{2p} + \left(\int_0^T \|\nabla Y^h\|^2 ds\right)^p \\
&\quad \leq C(p,T)\left(1 + \|X^h(0)\|^{2p} + \int_0^T \|Z^h\|_{L^{2K}}^{2Kp} ds\right).
\end{aligned}$$

Next, based on the above estimates and Lemma 3.4, we are in the position to prove the desired result. The mild form of Y^h and Lemmas 3.2 and 3.3, together with the Gagliardo–Nirenberg–Sobolev inequality $\|f\|_{L^{2K}} \leq C\|\nabla f\|^{\frac{K-1}{2K}}\|f\|^{\frac{K+1}{2K}}$, lead to

$$\begin{aligned}
\|Y^h(t)\|_E &\leq \|S^h(t)P^h X_0\|_E + \int_0^t \|S^h(t-s)P^h F\left(Y^h + Z^h\right)\|_E ds \\
&\leq C\|X_0\|_E + C\int_0^t (t-s)^{-\frac{1}{4}} \left\|F\left(Y^h + Z^h\right)\right\| ds \\
&\leq C\|X_0\|_E + C\int_0^t (t-s)^{-\frac{1}{4}} \left(1 + \|Y^h\|_{L^{2K}}^K + \|Z^h\|_{L^{2K}}^K\right) ds \\
&\leq C\|X_0\|_E + C\int_0^t (t-s)^{-\frac{1}{4}} \left(1 + \|Z^h\|_{L^{2K}}^K\right) ds \\
&\quad + C\int_0^t (t-s)^{-\frac{1}{4}} \|\nabla Y^h\|^{\frac{K-1}{2}} \|Y^h\|^{\frac{K+1}{2}} ds.
\end{aligned}$$

Taking the pth moment on both sides, together with the Hölder and Young inequalities and the boundedness of $\int_0^T \|\nabla Y^h\|^2 ds$ and Z^h, yields that for $p \geq 1$ and $K < 5$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]} \|Y^h(t)\|_E^p\right] \\
&\quad \leq C\|X_0\|_E^p + C(T,p)\left(1 + \log\left(\frac{1}{h}\right)\right)^{\frac{Kp}{2}} \\
&\qquad + C\mathbb{E}\left[\left(\int_0^t (t-s)^{-\frac{1}{4}} \|\nabla Y^h\|^{\frac{K-1}{2}} ds\right)^p \sup_{s\in[0,T]} \|Y^h(s)\|^{\frac{(K+1)p}{2}}\right] \\
&\quad \leq C\sqrt{\mathbb{E}\left[\left(\sup_{t\in[0,T]} \int_0^t (t-s)^{-\frac{1}{4}} \|\nabla Y^h\|^{\frac{K-1}{2}} ds\right)^{2p}\right]} \sqrt{\mathbb{E}\left[\sup_{s\in[0,T]} \|Y^h(s)\|^{(K+1)p}\right]} \\
&\qquad + C(X_0,T,p)\left(1 + \log\left(\frac{1}{h}\right)\right)^{\frac{Kp}{2}}
\end{aligned}$$

$$
\begin{aligned}
&\leq C(X_0,T,p)\left(1+\log\left(\frac{1}{h}\right)\right)^{\frac{(K+1)Kp}{4}}\sqrt{\mathbb{E}\left[\left(\sup_{t\in[0,T]}\int_0^t\|\nabla Y^h\|^2ds\right)^{\frac{(K-1)p}{2}}\right]}\\
&\quad+C(X_0,T,p)\left(1+\log\left(\frac{1}{h}\right)\right)^{\frac{Kp}{2}}\\
&\leq C(X_0,T,p)\left(1+\log\left(\frac{1}{h}\right)\right)^{\frac{K^2p}{2}}.
\end{aligned}
$$

Using the fact that

$$
\mathbb{E}\left[\sup_{t\in[0,T]}\|X^h(t)\|_E^p\right]\leq C_p\mathbb{E}\left[\sup_{t\in[0,T]}\|Y^h\|_E^p\right]+C_p\mathbb{E}\left[\sup_{t\in[0,T]}\|Z^h(t)\|_E^p\right],
$$

together with the a priori estimate on Y^h above and Z^h in Lemma 3.4, we complete the proof. ∎

Remark 3.1. Under the assumptions of Proposition 3.1, if in addition we assume that $\|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0}<\infty,\beta>\frac{1}{2}$ or that $Q=I$, we have the following optimal estimate:

$$
\sup_{t\in[0,T]}\mathbb{E}\left[\left\|X^h(t)\right\|_{\mathcal{V}}^p\right]\leq C(p,T,X_0).
$$

The above a priori estimate of X^h in $\mathcal{V}$ is a crucial part to deriving the weak convergence rate in section 4. This is the main reason why we require that Assumption 2.3 holds for $K<5$ in Proposition 3.1.

3.2. Strong convergence rate. In this subsection, we aim to give the strong convergence result of the finite element method. We also remark that this approach to getting strong convergence rates does not require the additional a priori estimate of the spatial approximation X^h.

THEOREM 3.1. *Under Assumptions* 2.1–2.3*, the finite element approximation $X^h(t)$ is strongly convergent to $X(t)$, $t\in(0,T]$, and satisfies, for $p\geq 1$,*

$$
\begin{aligned}
&\mathbb{E}\Big[\|X(t)-X^h(t)\|^p\Big]\leq C(X_0,T,p)\left(1+t^{-\frac{\beta}{2}}\right)^p h^{\beta p}\ \ \textit{for } \beta>\frac{1}{2},\\
&\mathbb{E}\Big[\|X(t)-X^h(t)\|^p\Big]\leq C(X_0,T,p)\left(1+t^{-\frac{\beta}{2}}+\left(\log\left(\frac{1}{h}\right)\right)^{\frac{(K-1)}{2}}\right)^p h^{\beta p}\ \ \textit{for } \beta\leq\frac{1}{2}.
\end{aligned}
$$

Proof. Since A does not commute with P^h, we could not use the usual strategy which divides the strong error $X(t)-X^h(t)$ into $(I-P^h)X(t)$ and $P^hX(t)-X^h(t)$. Thus, we introduce a new auxiliary process $\widetilde{Y}^h$ which satisfies

$$
d\widetilde{Y}^h+A_h\widetilde{Y}^hdt=P^hF(Y+Z)dt,\quad \widetilde{Y}^h(0)=X^h(0).
$$

Now, we split strong error as

$$
\begin{aligned}
X(t)-X^h(t)&=Y(t)-Y^h(t)+Z(t)-Z^h(t)\\
&=\left(Y(t)-\widetilde{Y}^h(t)\right)+\left(\widetilde{Y}^h(t)-Y^h(t)\right)+(Z(t)-Z^h(t)),
\end{aligned}
$$

and estimate the three terms, respectively. Using the estimates (6) of $G^h(t) := S^h(t)P^h - S(t)$, $t \geq 0$, and the Burkholder inequality, we get

$$\begin{aligned}
&\mathbb{E}\Big[\|Z(t) - Z^h(t)\|^p\Big] \\
&\leq C_p \mathbb{E}\left[\left(\int_0^t \|G^h(t-s)\|^2_{\mathcal{L}_2^0} ds\right)^{\frac{p}{2}}\right] \\
&\leq C_p \mathbb{E}\left[\left(\int_0^t \|(R^h - I)S(t-s)\|^2_{\mathcal{L}_2^0} ds\right)^{\frac{p}{2}}\right] \\
&\leq C_p \mathbb{E}\left[\left(\int_0^t \left\|(R^h - I)A^{-\frac{\beta}{2}}\right\|^2_{\mathcal{L}} \left\| A^{\frac{1}{2}} S(t-s) A^{\frac{\beta-1}{2}}\right\|^2_{\mathcal{L}_2^0} ds\right)^{\frac{p}{2}}\right] \\
&\leq C(T,p) \left\|A^{\frac{\beta-1}{2}}\right\|^p_{\mathcal{L}_2^0} h^{\beta p}.
\end{aligned}$$

The mild forms of Y and $\widetilde{Y}^h$, together with the a priori estimates of Y and Z, and the properties (7) of the finite element method, yield that for $0 \leq u < 2$,

$$\begin{aligned}
&\mathbb{E}\Big[\|Y(t) - \widetilde{Y}^h(t)\|^p\Big] \\
&\leq C_p \mathbb{E}\left[\|G^h(t)X_0\|^p\right] + C_p \mathbb{E}\left[\left\|\int_0^t G^h(t-s)F(Y+Z)ds\right\|^p\right] \\
&\leq C_p h^{up} t^{-\frac{up}{2}} \|X_0\|^p + C_p \mathbb{E}\left[\left\|\int_0^t G^h(t-s)F(Y+Z)ds\right\|^p\right] \\
&\leq C_p h^{up} t^{-\frac{up}{2}} \|X_0\|^p + C(p,T)\mathbb{E}\left[\left(\int_0^t h^u (t-s)^{-\frac{u}{2}} \|F(Y(s)+Z(s))\| ds\right)^p\right] \\
&\leq C_p h^{up} t^{-\frac{up}{2}} \|X_0\|^p + C(p,T) h^{up} \mathbb{E}\left[\sup_{s\in[0,T]} \left(1 + \|Y(s)\|^K_{L^{2K}} + \|Z(s)\|^K_{L^{2K}}\right)^p\right] \\
&\leq C(X_0,T,p) h^{up}\left(1 + t^{-\frac{up}{2}}\right).
\end{aligned}$$

Notice that arguments similar to those in the proof of [6, Lemma 3.1] yield that $\mathbb{E}[\|\widetilde{Y}^h(t)\|^p_E] \leq C(p,T,X_0)$. Next we deal with the term $\widetilde{Y}^h(t) - Y^h(t)$. The random PDE forms of $\widetilde{Y}^h(t)$ and $Y^h(t)$ lead to

$$\begin{aligned}
&\left\|\widetilde{Y}^h(t) - Y^h(t)\right\|^2 \\
&\leq -2\int_0^t \left\|\nabla\left(\widetilde{Y}^h(s) - Y^h(s)\right)\right\|^2 ds \\
&\quad + \int_0^t 2\left\langle F(Y(s)+Z(s)) - F(Y^h(s)+Z^h(s)), \widetilde{Y}^h(s) - Y^h(s)\right\rangle ds \\
&\leq \int_0^t 2\left\langle F(Y(s)+Z(s)) - F(\widetilde{Y}^h(s)+Z^h(s)), \widetilde{Y}^h(s) - Y^h(s)\right\rangle ds \\
&\quad + \int_0^t 2\left\langle F(\widetilde{Y}^h(s)+Z^h(s)) - F(Y^h(s)+Z^h(s)), \widetilde{Y}^h(s) - Y^h(s)\right\rangle ds.
\end{aligned}$$

For $\beta \le \frac{1}{2}$, by the monotonicity of F and Lemma 3.4, we have

$$
\begin{aligned}
&\sup_{s\in[0,t]} \left\|\widetilde{Y}^h(s) - Y^h(s)\right\|^2 \\
&\le C\int_0^t \sup_{s\in[0,r]} \|\widetilde{Y}^h(s) - Y^h(s)\|^2 dr + C\Bigg(\int_0^t \left(\|Y(r) - \widetilde{Y}^h(r)\| + \|Z(r) - Z^h(r)\|\right) \\
&\quad \times \left(1+\|Y(r)\|_E^{K-1}+\|\widetilde{Y}^h(r)\|_E^{K-1}+\|Z(r)\|_E^{K-1}+\|Z^h(r)\|_E^{K-1}\right)\|\widetilde{Y}^h(r)-Y^h(r)\|dr\Bigg) \\
&\le C\int_0^t \sup_{s\in[0,r]} \|\widetilde{Y}^h(s) - Y^h(s)\|^2 dr + \epsilon \sup_{r\in[0,t]} \|\widetilde{Y}^h(r) - Y^h(r)\|^2 \\
&\quad + C(\epsilon)\Bigg(\int_0^t \left(\|Y(r) - \widetilde{Y}^h(r)\| + \|Z(r) - Z^h(r)\|\right) \\
&\qquad \times \left(1+\|Y(r)\|_E^{K-1}+\|\widetilde{Y}^h(r)\|_E^{K-1}+\|Z(r)\|_E^{K-1}+\|Z^h(r)\|_E^{K-1}\right)dr\Bigg)^2.
\end{aligned}
$$

Taking the pth moment yields that for $0 \le \mu < 2$,

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{s\in[0,t]} \|\widetilde{Y}^h(s) - Y^h(s)\|^{2p}\right] \\
&\le C_p \int_0^t \mathbb{E}\left[\sup_{s\in[0,r]} \|\widetilde{Y}^h(s) - Y^h(s)\|^{2p}\right] dr + C(X_0,T,p)\left(1+\log\left(\frac{1}{h}\right)\right)^{(K-1)p} \\
&\quad \times \left(\left(\int_0^t h^{\mu}(1+s^{-\frac{\mu}{2}})ds\right)^{2p} + h^{2\beta p}\right) \\
&\le C\int_0^t \mathbb{E}\left[\sup_{s\in[0,r]} \|\widetilde{Y}^h(s) - Y^h(s)\|^{2p}\right] dr + C(X_0,T,p)\left(1+\log\left(\frac{1}{h}\right)\right)^{(K-1)p} h^{2\beta p}.
\end{aligned}
$$

Then the Gronwall inequality leads to

$$
\mathbb{E}\Big[\|\widetilde{Y}^h(t) - Y^h(t)\|^{2p}\Big] \le C(X_0,T,p)\left(1+\log\left(\frac{1}{h}\right)\right)^{(K-1)p} h^{2\beta p}.
$$

For $\beta > \frac{1}{2}$, similar arguments, together with Lemma 3.4 and the boundedness of $\|\widetilde{Y}^h(t)\|_E$, imply that

$$
\mathbb{E}\Big[\|\widetilde{Y}^h(t) - Y^h(t)\|^{2p}\Big] \le C(X_0,T,p) h^{2\beta p}.
$$

Combining the strong error estimates of $Y(t)-\widetilde{Y}^h(t)$, $\widetilde{Y}^h(t)-Y^h(t)$ and $Z(t)-Z^h(t)$ together, we finish the proof. ∎

Remark 3.2. Under the assumptions of Theorem 3.1, if in addition $X_0 \in \mathbb{H}^{\beta}$, then the term $t^{-\frac{\beta}{2}}$ in the strong convergence rate result can be eliminated. When $Q = I$, the logarithmic factor in the strong error estimate can also be eliminated. We also remark that the approach to deducing strong convergence rates of numerical schemes is also available for SPDEs with nonmonotone coefficients (see, e.g., [15]).

Remark 3.3. Assume that $\mathcal{O} \subset \mathcal{R}^d, d \le 3$, is a bounded open domain with smooth boundary, $\{W(t)\}_{t\ge 0}$ satisfies $\|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0} < \infty$ for some $\beta \in [\frac{(K-1)d}{2K}, 2]$, and

$\sup_{t\in[0,T]}\|\int_0^t S(t-s)dW(s)\|_{L^{p_0}(\Omega,L^{2K^2})}<\infty$ for a sufficient large number $p_0\in\mathbb{N}^+$. Then for $X_0\in\mathbb{H}^\beta$, $p\geq 1$, it holds that

$$\mathbb{E}\Big[\|X(t)-X^h(t)\|^p\Big]\leq C(X_0,T,p)h^{\beta p}.$$

In this case, the proof of the strong convergence rate result does not rely on the a priori estimates of the finite element method. The key lies in using the Sobolev embedding $\mathbb{H}^\gamma\hookrightarrow L^{2K}, \gamma\in[\frac{(K-1)d}{2K},1]$, and the dissipativity of $-A$ and $-A_h$.

4. Regularity of Kolmogorov equation and weak convergence rate.

4.1. Regularity of Kolmogorov equation. Denote $\mathcal{C}_b^2(\mathbb{H}):=\mathcal{C}_b^2(\mathbb{H},\mathbb{R})$. Set $U(t,x)=\mathbb{E}[\phi(X(t,x))]$; then formally, U is the solution of the Kolmogorov equation associated with (1):

$$\frac{\partial U(t,x)}{\partial t}=\langle -Ax+F(x),DU(t,x)\rangle+\operatorname{tr}\left[Q^{\frac12}D^2U(t,x)Q^{\frac12}\right].$$

To give the rigorous meaning of the Kolmogorov equation, we follow the approach in [9]. We first apply the splitting strategy inspired by [6, 8] to regularize the original equation and get a regularized Kolmogorov equation. Then making use of the regularity of the regularized Kolmogorov equation and integration by parts formula in the Malliavin sense, we obtain the weak convergence rate of the finite element method.

Now, we are in a position to give the rigorous meaning of the regularized Kolmogorov equation and regularity estimates of DU and D^2U. The following lemma is useful in constructing the regularized PDE and its corresponding Kolmogorov equation. For a function f on $\mathbb{R}$, we denote the first derivative and second derivative by f' and f''.

LEMMA 4.1. *Let $L_f>0$, $K\in\mathbb{N}^+$, and $f:\mathbb{R}\to\mathbb{R}$ satisfy*

$$|f(\xi)|\leq L_f\left(1+|\xi|^K\right),\quad |f'(\xi)|\leq L_f\left(1+|\xi|^{K-1}\right),$$
$$f'(\xi)\leq L_f,\quad |f''(\xi)|\leq L_f\left(1+|\xi|^{(K-2)\vee 0}\right).$$

Then the phase flow Φ_t of the differential equation

$$dx(t)=f(x(t))dt,\quad x(0)=\xi\in\mathbb{R}, \tag{11}$$

satisfies for all $\xi\in\mathbb{R}$,

$$|\Phi_t(\xi)|\leq C(f,t)(1+|\xi|),\quad \Phi_t'(\xi)\leq C(f,t),\quad |\Phi_t''(\xi)|\leq C(f,t)\left(1+|\xi|^{(K-2)\vee 0}\right).$$

Proof. From the properties of f and the Young inequality, it follows that

$$\begin{aligned}|x(t)|^2&=|\xi|^2+\int_0^t 2(f(x(s))-f(0))x(s)ds+\int_0^t 2f(0)x(s)ds\\&\leq|\xi|^2+\int_0^t\left(L_f^2+(1+2L_f)|x(s)|^2\right)ds.\end{aligned}$$

Then Gronwall's inequality implies that $|\Phi_t(\xi)|=|x(t)|\leq C(f,t)(1+|\xi|)$. Similarly, using the differentiable dependence on initial datum, we obtain

$$\Phi_t'(\xi)=1+\int_0^t f'(\Phi_s(\xi))\Phi_s'(\xi)ds,$$

which, together with Gronwall's inequality, yields that $0 \le \Phi_t'(\xi) \le e^{L_f t}$. Similar arguments lead to

$$\begin{aligned}
|\Phi_t''(\xi)|^2 &\le 2\int_0^t f'(\Phi_s(\xi))|\Phi_s''(\xi)|^2 ds + 2\int_0^t f''(\Phi_s(\xi))(\Phi_s'(\xi))^2\Phi_s''(\xi)ds \\
&\le \int_0^t (2L_f+1)|\Phi_s''(\xi)|^2 ds + \int_0^t |f''(\Phi_s(\xi))|^2|\Phi_s'(\xi)|^4 ds,
\end{aligned}$$

which indicates that $|\Phi_t''(\xi)| \le C(f,t)(1+|\xi|^{(K-2)\vee 0})$. ∎

With the help of Lemma 4.1, we introduce our regularizing procedures. Based on the strategy of the splitting method in [8], we split (1) into two subsystems,

$$dX_1 = F(X_1)dt, \quad dX_2 = -AX_2 dt + dW(t).$$

Then given a fixed time step size $\delta t > 0$, the splitting method is defined as

$$\begin{aligned}
\widetilde{X}_{n+1} &:= S(\delta t)\Phi_{\delta t}(\widetilde{X}_n) + \int_{t_n}^{t_{n+1}} S(t_{n+1}-s)dW(s) \\
&= S_{\delta t}\widetilde{X}_n + \delta t S_{\delta t}\Psi_{\delta t}(\widetilde{X}_n) + \int_{t_n}^{t_{n+1}} S(t_{n+1}-s)dW(s)
\end{aligned}$$

for $0 \le n \le N-1, N\delta t = T, t_n = n\delta t$, where $\Psi_t(x) := \frac{\Phi_t(x)-x}{t}$, $t > 0$, and $\Psi_0(x) = F(x)$. Notice that the splitting method can be also used to approximate SPDE with non-monotone coefficients in strong and weak convergence senses (see, e.g., [12, 14]). Based on the idea that $\{\widetilde{X}_n\}_{n=1,\dots,N}$ is the exponential Euler method applied to the SPDE in [6, 8], we introduce the auxiliary problem as

$$dX^{\delta t} + AX^{\delta t}dt = \Psi_{\delta t}(X^{\delta t})dt + dW(t), \quad X^{\delta t}(0) = X_0. \tag{12}$$

The differentiability of $\Psi_{\delta t}$ is given in the following lemma, which generalizes the case in [8, Lemma 2.1].

LEMMA 4.2. *Under the conditions of Lemma* 4.1, *for* $\delta t_0 \in (0,1]$, *there exists* $C(\delta t_0, f) > 0$ *such that for all* $\delta t \in [0, \delta t_0]$ *and* $\xi \in \mathbb{R}$,

$$\begin{aligned}
&\Psi_{\delta t}'(\xi) \le e^{C\delta t_0}, &&|\Psi_{\delta t}'(\xi)| \le C(\delta t_0)\left(1+|\xi|^{K-1}\right), \\
&|\Psi_{\delta t}''(\xi)| \le C(\delta t_0)\left(1+|\xi|^{(2K-3)\vee(K-1)}\right), &&|\Psi_{\delta t}(\xi)-\Psi_0(\xi)| \le C(\delta t_0)\delta t\left(1+|\xi|^{2K-1}\right).
\end{aligned}$$

Proof. By the definition of $\Psi_{\delta t}$ and the properties of $\Phi_{\delta t}$ in Lemma 4.1, we have

$$\begin{aligned}
\Psi_{\delta t}'(\xi) &= \frac{\Phi_{\delta t}'(\xi)-1}{\delta t} = \frac{\int_0^{\delta t} f'(\Phi_s(\xi))\Phi_s'(\xi)ds}{\delta t} \le C(f,\delta t_0), \\
|\Psi_{\delta t}'(\xi)| &\le \frac{|\int_0^{\delta t} f'(\Phi_s(\xi))\Phi_s'(\xi)ds|}{\delta t} \le C(f,\delta t_0)\sup_{s\in[0,\delta t]}\left(1+|\Phi_s(\xi)|^{K-1}\right) \\
&\le C(f,\delta t_0)(1+|\xi|^{K-1}), \\
|\Psi_{\delta t}''(\xi)| &\le \frac{|\int_0^{\delta t} f''(\Phi_s(\xi))(\Phi_s'(\xi))^2 + f'(\Phi_s(\xi))\Phi_s''(\xi)ds|}{\delta t} \\
&\le C(f,\delta t_0)\left(1+|\xi|^{(2K-3)\vee(K-1)}\right),
\end{aligned}$$

$$
\begin{aligned}
|\Psi_{\delta t}(\xi)-\Psi_0(\xi)| &\le \frac{|\int_0^{\delta t}\int_0^1 f'(\theta\Phi_s(\xi)+(1-\theta)\xi)(\Phi_s(\xi)-\xi)d\theta ds|}{\delta t}\\
&\le \sup_{s\in[0,\delta t]}\int_0^1 |f'(\theta\Phi_s(\xi)+(1-\theta)\xi)|d\theta \sup_{s\in[0,\delta t]}|(\Phi_s(\xi)-\xi)|\\
&\le C(f,\delta t_0)\delta t\left(1+|\xi|^{2K-1}\right). \qquad \square
\end{aligned}
$$

Based on Lemma 4.2, the coefficient $\Psi_{\delta t}(\cdot)$ of (12) is globally Lipschitz due to the fact that $\Psi_t(\xi)=\frac{\Phi_t(\xi)-\xi}{t}$. However, the Lipschitz constants of $\Psi_t, t\ge 0$, are not uniformly bounded with respect to t (see, e.g., [9]). Indeed, the solution of (12) is strongly convergent to that of (1), whose proof is similar to in that of [8, Proposition 4.8].

LEMMA 4.3. *Let Assumptions* 2.1–2.3 *hold. Then the solution* $X^{\delta t}$ *of* (12) *is strongly convergent to the solution* X *of* (1) *and satisfies, for any* $p\ge 1$,

$$
\begin{aligned}
&\mathbb{E}\Big[\|X^{\delta t}(t)\|_E^p\Big]\le C(T,p)\left(1+\|X_0\|_E^p\right),\\
&\left\|\sup_{t\in[0,T]}\left\|X^{\delta t}(t)-X(t)\right\|\right\|_{L^p(\Omega)}\le C(X_0,T,p)\delta t.
\end{aligned}
$$

The idea of deducing the sharp weak convergence rate lies on the decomposition of $\mathbb{E}[\phi(X(t))-\phi(X^h(t))]$ into $\mathbb{E}[\phi(X(t))-\phi(X^{\delta t}(t))]$ and $\mathbb{E}[\phi(X^{\delta t}(t))-\phi(X^h(t))]$. The first term is estimated by Lemma 4.3 and possesses the strong convergence rate with respect to the parameter δt. The second error is estimated by utilizing the regularity of Kolmogorov equation with respect to (12) and integration by parts in the sense of Malliavin calculus. Similar to [9, 10], to get the rigorous regularity result of the Kolmogorov equation, the noise term $dW(t)$ is regularized as $e^{\delta A}dW(t)$, $\delta>0$. For convenience, we omit the procedure of regularizing the noise since the following proposition allows us to take the limit $\delta\to 0$.

Next, we give the regularity estimate of Kolmogorov equation with respect to (12),

$$
\begin{aligned}
\frac{\partial U^{\delta t}(t,x)}{\partial t} &= \mathcal{L}^{\delta t}U^{\delta t}(t,x)=\left\langle -Ax+\Psi_{\delta t}(x), DU^{\delta t}(t,x)\right\rangle\\
&\quad+\frac{1}{2}\mathrm{tr}\left[Q^{\frac12}D^2U^{\delta t}(t,x)Q^{\frac12}\right].
\end{aligned}\tag{13}
$$

PROPOSITION 4.1. *Let* $\phi\in\mathcal{C}_b^2(\mathbb{H})$. *For every* $\alpha,\beta,\gamma\in[0,1)$, $\beta+\gamma<1$, *there exist* $C(T,\delta t_0,\alpha)$ *and* $C(T,\delta t_0,\beta,\gamma)$ *such that for* $\delta t\in[0,\delta t_0]$, $x\in E, y,z\in\mathbb{H}$, *and* $t\in(0,T]$,

$$
|DU^{\delta t}(t,x).y|\le\frac{C(T,\delta t_0,\alpha)\left(1+|x|_E^{K-1}\right)|\phi|_{\mathcal{C}_b^1}}{t^\alpha}\|A^{-\alpha}y\|,\tag{14}
$$

$$
|D^2U^{\delta t}(t,x).(y,z)|\le\frac{C(T,\delta t_0,\beta,\gamma)\left(1+|x|_E^{(5K-6)\vee(4K-4)}\right)|\phi|_{\mathcal{C}_b^2}}{t^{\beta+\gamma}}\|A^{-\beta}y\|\|A^{-\gamma}z\|.\tag{15}
$$

Proof. Similar arguments in [9, Theorem 4.1] yield (14) and that for $0\le\alpha<1$,

$$
\|\eta^y(t,x)\|\le\frac{C(T,\delta t_0,\alpha)}{t^\alpha}\sup_{t\in[0,T]}\|\Psi'_{\delta t}(X^{\delta t}(t,x))\|_E\|A^{-\alpha}y\|,\tag{16}
$$

where η^y satisfies

$$d\eta^y(t,x) = (-A + \Psi'_{\delta t}(X^{\delta t}(t,x))\eta^y(t,x)dt,\ \eta^y(t,x) = y.$$

Here we give a short proof for (15) which is different from the dual argument in [9]. Notice that

$$\begin{aligned} D^2U^{\delta t}(t,x).(y,z) = {} & \mathbb{E}[D\phi(X^{\delta t}(t,x)).\zeta^{y,z}(t,x)] \\ & + \mathbb{E}[D^2\phi(X^{\delta t}(t,x)).(\eta^y(t,x),\eta^z(t,x))], \end{aligned}$$

where $\zeta^{y,z}$ satisfies

$$d\zeta^{y,z}(t,x) = (-A + \Psi'_{\delta t}(X^{\delta t}(t,x))\zeta^{y,z}(t,x)dt + \Psi''_{\delta t}(X^{\delta t}(t,x))\eta^y(t,x)\eta^z(t,x).$$

Thus it suffices to prove the regularity of $\zeta^{y,z}$ thanks to (16). Due to the fact that

$$\zeta^{y,z}(t,x) = \int_0^t V(t,s)\Big(\Psi''_{\delta t}(X^{\delta t}(s,x))\eta^y(s,x)\eta^z(s,x)\Big)ds,$$

where

$$dV(t,s)z = (-A + \Psi'_{\delta t}(X^{\delta t}(t,x)))V(t,s)zdt, \quad V(s,s)z = z,$$

we need to deduce a more refined estimate of $V(t,s)z,\ 0 \le s < t \le T$. The property of $\Psi'_{\delta t}$ in Lemma 4.2, combined with an energy estimate, yields that $\|V(t,s)z\|^2 \le C(T,\delta t_0)\|z\|^2$. Moreover, we claim that for $0 \le s < t \le T$, $0 \le \alpha < 1$,

$$\|V(t,s)y\| \le \frac{C(T,\delta t_0,\alpha)}{(t-s)^\alpha} \sup_{r\in[0,T]} \|\Psi'_{\delta t}(X^{\delta t}(r,x))\|_E \|A^{-\alpha}y\|. \tag{17}$$

Indeed, let $\widetilde{V}(t,s)y = V(t,s)y - e^{-(t-s)A}y, 0 \le s \le t \le T$. Then we have for $t > s$,

$$\begin{aligned} & d\widetilde{V}(t,s)y = (-A + \Psi'_{\delta t}(X^{\delta t}(t,x)))\widetilde{V}(t,s)ydt + \Psi'_{\delta t}(X^{\delta t}(t,x))e^{-(t-s)A}ydt, \\ & \widetilde{V}(s,s)y = 0, \end{aligned}$$

and

$$\begin{aligned} \|\widetilde{V}(t,s)y\| &\le \int_s^t \Big\|V(t,r)\Big(\Psi'_{\delta t}(X^{\delta t}(r,x))e^{-(r-s)A}y\Big)\Big\|dr \\ &\le C(T,\delta t_0) \sup_{r\in[0,T]} \|\Psi'_{\delta t}(X^{\delta t}(r,x))\|_E \int_s^t \|e^{-(r-s)A}y\|dr \\ &\le C(T,\delta t_0,\alpha) \sup_{r\in[0,T]} \|\Psi'_{\delta t}(X^{\delta t}(r,x))\|_E (t-s)^{1-\alpha}\|A^{-\alpha}y\|, \end{aligned}$$

which implies that the estimate (17) holds. Now, we are in a position to prove (15). Based on (17) and the Sobolev embedding theorem, we have for $\alpha > \frac{1}{4}$,

$$\begin{aligned} \|\zeta^{y,z}(t,x)\| &\le \int_0^t \|V(t,s)\Psi''_{\delta t}(X^{\delta t}(s,x))\eta^y(s,x)\eta^z(s,x)\|ds \\ &\le C(T,\delta t_0,\alpha) \int_0^t \sup_{r\in[0,T]} \|\Psi'_{\delta t}(X^{\delta t}(r,x))\|_E (t-s)^{-\alpha} \\ &\quad \times \Big\|A^{-\alpha}\Big(\Psi''_{\delta t}(X^{\delta t}(s,x))\eta^y(s,x)\eta^z(s,x)\Big)\Big\|ds \end{aligned}$$

$$\leq C(T,\delta t_0,\alpha)\int_0^t \sup_{r\in[0,T]}\|\Psi'_{\delta t}(X^{\delta t}(r,x))\|_E (t-s)^{-\alpha}\|\Psi''_{\delta t}(X^{\delta t}(s,x))\|_E \times \|\eta^y(s,x)\|\|\eta^z(s,x)\|ds$$
$$\leq C(T,\delta t_0,\alpha)\sup_{r\in[0,T]}\|\Psi'_{\delta t}(X^{\delta t}(r,x))\|_E \sup_{r\in[0,T]}\|\Psi''_{\delta t}(X^{\delta t}(r,x))\|_E \times \int_0^t (t-s)^{-\alpha}\|\eta^y(s,x)\|\|\eta^z(s,x)\|ds.$$

Now using the estimation (16), the growth of $\Psi_{\delta t}$, and the stability of $X^{\delta t}$, we obtain

$$\begin{aligned}\mathbb{E}[\|\zeta^{y,z}(t,x)\|] &\leq C(T,\delta t_0,\alpha,\beta,\gamma)\int_0^t (t-s)^{-\alpha}s^{-\beta-\gamma}ds\|A^{-\beta}y\|\|A^{-\gamma}y\| \\ &\quad\times\mathbb{E}\left[\sup_{r\in[0,T]}\|\Psi'_{\delta t}(X^{\delta t}(r,x))\|_E^3 \sup_{r\in[0,T]}\|\Psi''_{\delta t}(X^{\delta t}(r,x))\|_E\right] \\ &\leq C(T,\delta t_0,\alpha,\beta,\gamma)t^{-\beta-\gamma}\left(1+\|x\|_E^{(5K-6)\vee(4K-4)}\right)\|A^{-\beta}y\|\|A^{-\gamma}y\|,\end{aligned}$$

which completes the proof. ∎

Remark 4.1. The Sobolev embedding inequality $\|y\|_{L^\infty}\leq C\|y\|_{\mathbb{H}^{\frac{d}{2}+\epsilon}}$, $y\in\mathbb{H}^{\frac{d}{2}+\epsilon}$, $\epsilon>0$, $d\leq 3$, yields that $\|A^{-\frac{d}{2}-\epsilon}y\|\leq C\|y\|_{L^1}$. Thus the regularity result of the Kolmogorov equation in Proposition 4.1 can be generalized to the higher dimensional case ($d=2,3$) and the more regular noise case.

4.2. Weak convergence rate. Before studying the weak convergence rate, we show that the numerical solution X^h is differentiable in the Malliavin sense and prove some estimates of X^h needed later similar to [2, Lemma 3.1].

PROPOSITION 4.2. *Let Assumptions* 2.1–2.3 *hold. Then the Malliavin derivative of X^h satisfies, for some constant $C(T,\beta,X_0,K)$,*

$$\mathbb{E}\left[\left\|A_h^{\frac{\beta-1}{2}}\mathcal{D}_sX^h(t)\right\|^2_{\mathcal{L}_2^0}\right]\leq C(T,\beta,X_0,K)\left(1+\log\left(\frac{1}{h}\right)\right)^{2(K-1)K^2},\quad 0\leq s\leq t\leq T.$$

Proof. Similar to the well-posedness of (2), we have that for $0\leq s\leq t\leq T$, $y\in U_0$,

$$\mathcal{D}_s^yX^h(t)=S^h(t-s)P^hy+\int_s^t S^h(t-r)P^hDF(X^h(r))\cdot\mathcal{D}_s^yX^h(r)dr$$

satisfies

$$\begin{aligned}d\mathcal{D}_s^yX^h(t)&=-A_h\mathcal{D}_s^yX^h(t)dt+P^hDF(X^h(t))\cdot\mathcal{D}_s^yX^h(t)dt,\\ \mathcal{D}_s^yX^h(s)&=P^hy.\end{aligned}\tag{18}$$

In order to get the estimate of $\|A_h^{\frac{\beta-1}{2}}\mathcal{D}_sX^h(t)\|_{\mathcal{L}_2^0}$, we first estimate $\|A_h^{-\gamma}\mathcal{D}_sX^h(t)y\|$, $0\leq\gamma\leq\frac{1}{2}$, and define $\widetilde{\eta}_s(t,y)=\mathcal{D}_sX^h(t)y-S^h(t-s)P^hy$. Then $\widetilde{\eta}_s(t,y)$ satisfies the equation

$$\begin{aligned}d\widetilde{\eta}_s(t,y)&=-A_h\widetilde{\eta}_s(t,y)dt+P^h(DF(X^h(t))\cdot\widetilde{\eta}_s(t,y))dt\\ &\quad+P^h(DF(X^h(t))\cdot S^h(t-s)P^hy)dt,\\ \widetilde{\eta}_s(s,y)&=0,\end{aligned}$$

and

$$\widetilde{\eta}_s(t,y) = \int_s^t \widehat{V}(t,r)P^h\left(DF(X^h(r))\cdot S^h(r-s)P^h y\right)dr,$$

where $\widehat{V}(t,r)z$ solves for $z \in V^h$,

$$d\widehat{V}(t,r)z = -A_h\widehat{V}(t,r)zdt + P^h\left(DF(X^h(t))\widehat{V}(t,r)z\right)dt, \quad \widehat{V}(r,r)z = z.$$

The energy estimate, combined with Gronwall's inequality, yields that for $s \le r \le t$,

$$\|\widehat{V}(t,r)z\|^2 \le C(T)\|z\|^2.$$

This implies that

$$\begin{aligned}\|\widetilde{\eta}_s(t,y)\| &\le C(T)\int_s^t \|P^h\left(DF(X^h(r))\cdot S^h(r-s)P^h y\right)\|dr\\ &\le C(T,\gamma)\sup_{r\in[0,T]}\left[1+\|X^h(r)\|_E^{K-1}\right]\int_s^t (r-s)^{-\gamma}dr\|A^{-\gamma}y\|.\end{aligned}$$

Combining with the fact that for $0 \le \gamma \le \frac{1}{2}$,

$$\|S^h(t-s)P^h y\| \le C(T,\gamma)(t-s)^{-\gamma}\|A^{-\gamma}y\|,$$

we get

$$\|\mathcal{D}_s^y X^h(t)\| \le C(T,\gamma)\sup_{r\in[0,T]}\left[1+\|X^h(r)\|_E^{K-1}\right](t-s)^{-\gamma}\|A^{-\gamma}y\|.$$

Thus by the mild form of $\mathcal{D}_s X^h(t)y$ and the equivalence of norms in (5), we obtain for $0 \le \gamma \le \frac{1}{2}$,

$$\begin{aligned}\|A^{-\gamma}\mathcal{D}_s^y X^h(t)\| &\le \|A^{-\gamma}S^h(t-s)P^h y\|\\ &\quad + \int_s^t \|S^h(t-r)P^h DF(X^h(r))\cdot \mathcal{D}_s^y X^h(r)\|dr\\ &\le C(\gamma)\|A_h^{-\gamma}S^h(t-s)P^h y\|\\ &\quad + C(T,\gamma)\sup_{r\in[0,T]}\left[1+\|X^h(r)\|_E^{K-1}\right]\int_s^t \|\mathcal{D}_s^y X^h(r)\|dr\\ &\le C(T,\gamma)\|A^{-\gamma}y\|\left(1+\sup_{r\in[0,T]}\left[1+\|X^h(r)\|_E^{2K-2}\right]\int_s^t (r-s)^{-\gamma}ds\right)\\ &\le C(T,\gamma)\left(1+\sup_{r\in[0,T]}\|X^h(r)\|_E^{2K-2}\right)\|A^{-\gamma}y\|.\end{aligned}$$

Now, taking $-\gamma = \frac{\beta-1}{2}$, $0 < \beta \le 1$, $y = Q^{\frac{1}{2}}e_i, i \in \mathbb{N}^+$, together with the stability result of X^h in Proposition 3.1, yields that

$$\begin{aligned}\mathbb{E}\left[\left\|A^{\frac{\beta-1}{2}}\mathcal{D}_s X^h(t)\right\|_{\mathcal{L}_2^0}^2\right] &\le C(T,\beta)\sum_{i\in\mathbb{N}^+}\mathbb{E}\left[\left(1+\sup_{r\in[0,T]}\|X^h(r)\|_E^{4(K-1)}\right)\|A^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}e_i\|^2\right]\\ &\le C(T,X_0,\beta)\left(1+\left(\log\left(\frac{1}{h}\right)\right)^{2(K-1)K^2}\right),\end{aligned}$$

which completes the proof. ∎

Now, we turn to estimate the weak error of $|\mathbb{E}[\phi(X^{\delta t}(t)) - \phi(X^h(t))]|$.

THEOREM 4.1. *Let Assumptions* 2.1–2.3 *hold. Assume in addition that* $|f''(\xi)| \le L_f(1+|\xi|^{K-2})$, $2 \le K < 5$*; then for every test functions* $\phi \in \mathcal{C}_b^2(\mathbb{H})$, $T>0$, $\beta \in (0,1]$, *and* $\gamma < \beta$, *there exists* $C(X_0,T,\beta,\phi)$ *such that*

$$\left|\mathbb{E}\Big[\phi(X^{\delta t}(T)) - \phi(X^h(T))\Big]\right| \le C(X_0,T,\beta,\phi)\left(h^{2\gamma} + \delta t\left(\log\left(\frac{1}{h}\right)\right)^{\frac{(3K-2)K^2}{2}}\right).$$

Proof. Based on the property $\mathbb{E}[\phi(X^h(T))] = \mathbb{E}[U^{\delta t}(0,X^h(T))]$, we split the weak error as

$$\begin{aligned}\left|\mathbb{E}\Big[\phi(X^{\delta t}(T)) - \phi(X^h(T))\Big]\right| &\le \left|\mathbb{E}\Big[\phi(X^{\delta t}(T))\Big] - \mathbb{E}\Big[U^{\delta t}(T,X^h(0))\Big]\right| \\ &\quad + \left|\mathbb{E}\Big[U^{\delta t}(T,X^h(0))\Big] - \mathbb{E}\Big[\phi(X^h(T))\Big]\right|.\end{aligned}$$

By the regularity of $DU^{\delta t}(t,x)$ (14) in Proposition 4.1, we bound the first error as for $0 \le \alpha < 1$,

$$\begin{aligned}&\left|\mathbb{E}\Big[\phi(X^{\delta t}(T))\Big] - \mathbb{E}\Big[U^{\delta t}(T,X^h(0))\Big]\right| \\ &\quad = \left|U^{\delta t}(T,X(0)) - U^{\delta t}(T,X^h(0))\right| \\ &\quad = C(T,\delta t_0)\left|\int_0^1 DU^{\delta t}(T,\theta X(0) + (1-\theta)X^h(0))d\theta \cdot ((I-P^h)X(0))\right| \\ &\quad \le C(T,\delta t_0,\alpha,\phi)T^{-\alpha}\mathbb{E}\left[\left(1 + \|X_0\|_E^{K-1} + \|X^h(0)\|_E^{K-1}\right)\right]\|(-A)^{-\alpha}(I-P^h)X(0)\| \\ &\quad \le C(T,\delta t_0,\alpha,\phi,X_0)T^{-\alpha}h^{2\alpha},\end{aligned}$$

where we use the fact $\|A^{-\alpha}(I-P^h)\|_{\mathcal{L}(\mathbb{H})} = \|A^{-\alpha}(I-P^h))^*\|_{\mathcal{L}(\mathbb{H})} = \|(I-P^h)A^{-\alpha}\|_{\mathcal{L}(\mathbb{H})}$ and the estimation (6).

Next, we aim to estimate the left error $|\mathbb{E}[U^{\delta t}(T,X^h(0))] - \mathbb{E}[\phi(X^h(T))]|$. We recall the Markov generator $\mathcal{L}^h$ of X^h,

$$(\mathcal{L}^hU)(x) = \left\langle -A_hx + P^hF(x), DU(x)\right\rangle + \frac{1}{2}\mathrm{tr}\left[P^hQP^hD^2U(x)\right],$$

where $U \in \mathcal{C}^2(\mathbb{H},\mathbb{R}), x \in V_h$. Then the Itô formula and corresponding Kolmogorov equation (13) yield that

$$\begin{aligned}&\mathbb{E}\Big[U^{\delta t}(T,X^h(0))\Big] - \mathbb{E}\Big[\phi(X^h(T))\Big] \\ &\quad = \mathbb{E}\Big[U^{\delta t}(T,X^h(0)) - U^{\delta t}(0,X^h(T))\Big] \\ &\quad = -\mathbb{E}\left[\int_0^T \Big(-\dot{U}^{\delta t}(T-t,X^h(t)) + \mathcal{L}^hU^{\delta t}(T-t,X^h(t))\Big)dt\right] \\ &\quad = \mathbb{E}\left[\int_0^T \left(\left(\mathcal{L}^{\delta t} - \mathcal{L}^h\right)U^{\delta t}(T-t,X^h(t))\right)dt\right].\end{aligned}$$

Based on the expressions of $\mathcal{L}^{\delta t}$ and $\mathcal{L}^h$, we obtain

$$
\begin{aligned}
&\Big|\mathbb{E}\Big[U^{\delta t}(T,X^h(0))\Big]-\mathbb{E}\Big[\phi(X^h(T))\Big]\Big|\\
&\quad\le\left|\mathbb{E}\left[\int_0^T\Big\langle(A-A_h)X^h(t),DU^{\delta t}(T-t,X^h(t))\Big\rangle dt\right]\right|\\
&\qquad+\left|\mathbb{E}\left[\int_0^T\Big\langle\Psi_{\delta t}(X^h(t))-P^hF(X^h(t)),DU^{\delta t}(T-t,X^h(t))\Big\rangle dt\right]\right|\\
&\qquad+\frac{1}{2}\left|\mathbb{E}\left[\int_0^T\mathrm{tr}\Big\{QD^2U^{\delta t}(T-t,X^h(t))-P^hQP^hD^2U^{\delta t}(T-t,X^h(t))\Big\}dt\right]\right|\\
&\quad=:e^1(T)+e^2(T)+e^3(T).
\end{aligned}
$$

The relation $R^h=A_h^{-1}P^hA$ implies that

$$
\begin{aligned}
&\langle(A-A_h)X^h(t),DU^{\delta t}(T-t,X^h(t))\rangle\\
&\quad=\langle(AP^h-P^hA_h)X^h(t),DU^{\delta t}(T-t,X^h(t))\rangle\\
&\quad=\langle X^h,(P^hA-A_hP^h)DU^{\delta t}(T-t,X^h(t))\rangle\\
&\quad=\langle X^h,A_hP^h(A_h^{-1}P^hA-I)DU^{\delta t}(T-t,X^h(t))\rangle\\
&\quad=\langle X^h,A_hP^h(R^h-I)DU^{\delta t}(T-t,X^h(t))\rangle.
\end{aligned}
$$

The above equality and the mild form of X^h lead to

$$
\begin{aligned}
&e^1(T)\\
&\quad\le\left|\mathbb{E}\left[\int_0^T\Big\langle S^h(t)X^h(0),A_hP^h(R^h-I)DU^{\delta t}(T-t,X^h(t))\Big\rangle dt\right]\right|\\
&\quad+\left|\mathbb{E}\left[\int_0^T\left\langle\int_0^t S^h(t-s)P^hF(X^h(s))ds,A_hP^h(R^h-I)DU^{\delta t}(T-t,X^h(t))\right\rangle dt\right]\right|\\
&\quad+\left|\mathbb{E}\left[\int_0^T\left\langle\int_0^t S^h(t-s)P^hdW(s),A_hP^h(R^h-I)DU^{\delta t}(T-t,X^h(t))\right\rangle dt\right]\right|\\
&\quad=:e^{1,1}(T)+e^{1,2}(T)+e^{1,3}(T).
\end{aligned}
$$

Applying the regularity estimate of $DU^{\delta t}$ (14), the smoothing property of S^h (8), and the stability of X^h in Proposition (3.1), it follows that for small $\epsilon>0$, $\epsilon<\alpha<1$,

$$
\begin{aligned}
e^{1,1}(T)&=\left|\mathbb{E}\left[\int_0^T\Big\langle A_h^{1-\epsilon}S^h(t)X^h(0),A_h^{\epsilon}P^h(R^h-I)A^{-\alpha}\right.\right.\\
&\qquad\qquad\left.\left.A^{\alpha}DU^{\delta t}(T-t,X^h(t))\Big\rangle dt\right]\right|\\
&\le\mathbb{E}\left[\int_0^T\Big\|A_h^{1-\epsilon}S^h(t)X^h(0)\Big\|\Big\|A_h^{\epsilon}P^h(R^h-I)A^{-\alpha}\Big\|_{\mathcal{L}(\mathbb{H})}\right.\\
&\qquad\left.\times\Big\|A^{\alpha}DU^{\delta t}(T-t,X^h(t))\Big\|dt\right]
\end{aligned}
$$

$$\leq C(T,\epsilon)h^{2\alpha-2\epsilon}\int_0^T t^{-1+\epsilon}\|X_0\|\mathbb{E}\Big[\Big\|A^\alpha DU^{\delta t}(T-t,X^h(t))\Big\|\Big]dt$$
$$\leq C(T,\epsilon,\alpha,\phi)h^{2\alpha-2\epsilon}\int_0^T t^{-1+\epsilon}(T-t)^{-\alpha}\|X_0\|\sup_{t\in[0,T]}\mathbb{E}\left[1+\Big\|X^h(t)\Big\|_E^{K-1}\right]dt.$$

Similar arguments yield that

$$\begin{aligned}
e^{1,2}(T)&=\Bigg|\mathbb{E}\left[\int_0^T\left\langle\int_0^t A_h^{1-\epsilon}S^h(t-s)P^hF(X^h(s))ds, A_h^\epsilon P^h(R^h-I)A^{-\alpha}\right.\right.\\
&\qquad\left.\left. A^\alpha DU^{\delta t}(T-t,X^h(t))\right\rangle dt\right]\Bigg|\\
&\leq\mathbb{E}\left[\int_0^T\int_0^t\Big\|A_h^{1-\epsilon}S^h(t-s)P^hF(X^h(s))\Big\|ds\right.\\
&\qquad\left.\times\Big\|A_h^\epsilon P^h(R^h-I)A^{-\alpha}\Big\|_{\mathcal{L}(\mathbb{H})}\Big\|A^\alpha DU^{\delta t}(T-t,X^h(t))\Big\|dt\right]\\
&\leq C(T,\epsilon,\alpha)h^{2\alpha-2\epsilon}\sup_{t\in[0,T]}\sqrt{\mathbb{E}\Big[\|F(X^h(t))\|^2\Big]}\\
&\quad\times\int_0^T\int_0^t\sqrt{\mathbb{E}\Big[\|A^\alpha DU^{\delta t}(T-t,X^h(t))\|^2\Big]}(t-s)^{-1+\epsilon}dsdt\\
&\leq C(T,\epsilon,\alpha,\phi)h^{2\alpha-2\epsilon}\sup_{t\in[0,T]}\sqrt{\mathbb{E}\Big[1+\|X^h(t)\|_E^{2K}\Big]}\\
&\quad\times\sup_{t\in[0,T]}\sqrt{\mathbb{E}\Big[1+\|X^h(t))\|_E^{2K-2}\Big]}\int_0^T\int_0^t(T-t)^{-\alpha}(t-s)^{-1+\epsilon}dsdt.
\end{aligned}$$

To deal with $e^{1,3}(T)$, we make use of the integration by parts formula in the Malliavin sense (9) and the chain rule to get

$$\begin{aligned}
e^{1,3}(T)&=\left|\mathbb{E}\left[\int_0^T\left\langle\int_0^t S^h(t-s)P^hdW(s),A_hP^h(R^h-I)DU^{\delta t}(T-t,X^h(t))\right\rangle dt\right]\right|\\
&=\Bigg|\mathbb{E}\left[\int_0^T\int_0^t\Big\langle S^h(t-s)P^h,\right.\\
&\qquad\left. A_hP^h(R^h-I)D^2U^{\delta t}(T-t,X^h(t))\mathcal{D}_sX^h(t)\Big\rangle_{\mathcal{L}_2^0}dsdt\right]\Bigg|.
\end{aligned}$$

Then by the property of the Hilbert–Schmidt operator and the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&e^{1,3}(T)\\
&=\Bigg|\mathbb{E}\left[\int_0^T\int_0^t\Big\langle A_h^{\frac{1+\beta}{2}-\epsilon}S^h(t-s)A_h^{\frac{1-\beta}{2}}A_h^{\frac{\beta-1}{2}}P^h,\right.\\
&\quad\left. A_h^{\frac{1-\beta}{2}+\epsilon}P^h(R^h-I)A^{-\frac{1+\beta}{2}+\epsilon}A^{\frac{1+\beta}{2}-\epsilon}D^2U^{\delta t}(T-t,X^h(t))\mathcal{D}_sX^h(t)\Big\rangle_{\mathcal{L}_2^0}dsdt\right]\Bigg|
\end{aligned}$$

$$\leq \mathbb{E}\left[\int_0^T \int_0^t \left\|A_h^{1-\epsilon} S^h(t-s)P^h\right\|_{\mathcal{L}(\mathbb{H})} \|A_h^{\frac{\beta-1}{2}} P^h\|_{\mathcal{L}_2^0} \left\|A_h^{\frac{1-\beta}{2}+\epsilon} P^h(R^h - I)\right.\right.$$
$$\left.\left. A^{-\frac{\beta+1}{2}+\epsilon}\right\|_{\mathcal{L}(\mathbb{H})} \left\|A^{\frac{1+\beta}{2}-\epsilon} D^2 U^{\delta t}(T-t, X^h(t)) A^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathbb{H})} \left\|A^{\frac{\beta-1}{2}} \mathcal{D}_s X^h(t)\right\|_{\mathcal{L}_2^0} ds dt\right].$$

Combining the regularity result of $DU^{\delta t}$ (14) and $D^2U^{\delta t}$ (15), the smoothing property of S^h, $\|A_h^{\frac{\beta-1}{2}} P^h\|_{\mathcal{L}_2^0} \leq C\|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0}$, and the stability of X^h together, we obtain

$$\begin{aligned}
e^{1,3}(T) &\leq C(T,\epsilon) h^{2\beta-4\epsilon} \int_0^T \int_0^t (t-s)^{-1+\epsilon} \sqrt{\mathbb{E}\left[\left\|A^{\frac{\beta-1}{2}} \mathcal{D}_s X^h(t)\right\|_{\mathcal{L}_2^0}^2\right]} \\
&\qquad \sqrt{\mathbb{E}\left[\left\|A^{\frac{1+\beta}{2}-\epsilon} D^2 U^{\delta t}(T-t, X^h(t)) A^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathbb{H})}^2\right]} ds dt \\
&\leq C(T,\epsilon) h^{2\beta-4\epsilon} \int_0^T \int_0^t (t-s)^{-1+\epsilon} (T-t)^{-1+\epsilon} \sqrt{\mathbb{E}\left[\left\|A^{\frac{\beta-1}{2}} \mathcal{D}_s X^h(t)\right\|_{\mathcal{L}_2^0}^2\right]} \\
&\qquad \sup_{t\in[0,T]} \sqrt{\mathbb{E}\left[1 + \|X^h(t)\|_E^{10K-12}\right]} ds dt.
\end{aligned}$$

Proposition 4.2 yields that

$$e^{1,3}(T) \leq C(X_0, T, \epsilon, \beta) h^{2\beta-4\epsilon} \left(1 + \log\left(\frac{1}{h}\right)\right)^{\frac{(7K-8)K^2}{2}}.$$

Thus we conclude that $e^1(T) \leq C(X_0, T, \epsilon, \beta)(1 + T^{-\beta} + (\log(\frac{1}{h}))^{\frac{(7K-8)K^2}{2}}) h^{2\beta-4\epsilon}$.

Next, we turn to focus on $e^2(T)$. From $\Psi_0 = F$, it follows that

$$\begin{aligned}
e^2(T) &\leq \left|\mathbb{E}\left[\int_0^T \left\langle \Psi_{\delta t}(X^h(t)) - \Psi_0(X^h(t)), DU^{\delta t}(T-t, X^h(t))\right\rangle dt\right]\right| \\
&\quad + \left|\mathbb{E}\left[\int_0^T \left\langle (I-P^h)F(X^h(t)), DU^{\delta t}(T-t, X^h(t))\right\rangle dt\right]\right| \\
&=: e^{2,1}(T) + e^{2,2}(T).
\end{aligned}$$

By the continuity of Ψ_t with respect to t in Lemma 4.2 and the regularity of $DU^{\delta t}$(14), it leads to

$$\begin{aligned}
e^{2,1}(T) &\leq \mathbb{E}\left[\int_0^T \left|\left\langle \Psi_{\delta t}(X^h(t)) - \Psi_0(X^h(t)), DU^{\delta t}(T-t, X^h(t))\right\rangle\right| dt\right] \\
&\leq \mathbb{E}\left[\int_0^T \left\|DU^{\delta t}(T-t, X^h(t))\right\| \left\|\Psi_{\delta t}(X^h(t)) - \Psi_0(X^h(t))\right\| dt\right] \\
&\leq C(T, \delta t_0) \delta t \left(1 + \sup_{t\in[0,T]} \|X^h(t)\|_E^{3K-2}\right).
\end{aligned}$$

The regularity of $DU^{\delta t}$ (14), the estimate (6), the growth of F, and the stability of X^h yield that

$$e^{2,2}(T) \le \mathbb{E}\left[\int_0^T \left\|(I-P^h)A^{-1+\epsilon}\right\|_{\mathcal{L}(\mathbb{H})}\left\|A^{1-\epsilon}DU^{\delta t}(T-t,X^h(t))\right\|_{\mathcal{L}(\mathbb{H})} \times \|F(X^h(t))\|dt\right]$$
$$\le C(T,\epsilon)h^{2-2\epsilon}\int_0^T (T-t)^{-1+\epsilon}\sqrt{\mathbb{E}\Big[\|F(X^h(t))\|^2\Big]}\sqrt{\mathbb{E}\Big[1+\|X^h(t)\|_E^{2K-2}\Big]}dt$$
$$\le C(T,\epsilon)h^{2-2\epsilon}\sup_{t\in[0,T]}\mathbb{E}\Big[1+\|X^h(t)\|_E^{2K-1}\Big].$$

Summing up the estimations of $e^{2,1}$ and $e^{2,2}$, we deduce that

$$e^2(T) \le C(X_0,T,\epsilon)\left(h^{2-2\epsilon}\left(1+\left(\log\left(\frac{1}{h}\right)\right)^{\frac{(2K-1)K^2}{2}}\right) + \delta t\left(1+\left(\log\left(\frac{1}{h}\right)\right)^{\frac{(3K-2)K^2}{2}}\right)\right).$$

For the last term $e^3(T)$, we have

$$e^3(T) = \frac{1}{2}\left|\mathbb{E}\left[\int_0^T \mathrm{tr}\left\{\left(IQ(I-P^h)+(I-P^h)QP^h\right)D^2U^{\delta t}(T-t,X^h(t))\right\}dt\right]\right|$$
$$\le \frac{1}{2}\left|\mathbb{E}\left[\int_0^T \mathrm{tr}\Big\{IQ(I-P^h)D^2U^{\delta t}(T-t,X^h(t))\Big\}dt\right]\right|$$
$$+\frac{1}{2}\left|\mathbb{E}\left[\int_0^T \mathrm{tr}\Big\{(I-P^h)QP^hD^2U^{\delta t}(T-t,X^h(t))\Big\}dt\right]\right|$$
$$=: e^{3,1}(T)+e^{3,2}(T).$$

The properties of the trace and the Hilbert–Schmidt operator lead to

$$e^{3,1}(T) = \frac{1}{2}\left|\mathbb{E}\left[\int_0^T \mathrm{tr}\Big\{IQ(I-P^h)D^2U^{\delta t}(T-t,X^h(t))A^{\frac{1-\beta}{2}}A^{\frac{\beta-1}{2}}\Big\}dt\right]\right|$$
$$= \frac{1}{2}\left|\mathbb{E}\left[\int_0^T \mathrm{tr}\Big\{A^{\frac{\beta-1}{2}}Q(I-P^h)D^2U^{\delta t}(T-t,X^h(t))A^{\frac{1-\beta}{2}}\Big\}dt\right]\right|$$
$$= \frac{1}{2}\left|\mathbb{E}\left[\int_0^T \mathrm{tr}\Big\{A^{\frac{\beta-1}{2}}QA^{\frac{\beta-1}{2}}A^{\frac{1-\beta}{2}}(I-P^h)A^{-\frac{1+\beta}{2}+\epsilon} A^{\frac{1+\beta}{2}-\epsilon}D^2U^{\delta t}(T-t,X^h(t))A^{\frac{1-\beta}{2}}\Big\}dt\right]\right|$$
$$\le \frac{1}{2}\int_0^T \mathbb{E}\Big[\|A^{\frac{\beta-1}{2}}\|^2_{\mathcal{L}_2^0}\left\|A^{\frac{1-\beta}{2}}(I-P^h)A^{-\frac{1+\beta}{2}+\epsilon}\right\|_{\mathcal{L}(\mathbb{H})} \times \left\|A^{\frac{1+\beta}{2}-\epsilon}D^2U^{\delta t}(T-t,X^h(t))A^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathbb{H})}\Big]dt.$$

Then the regularity of $D^2U^{\delta t}$ (15), the estimate (6), and Proposition 3.1 yield that

$$
\begin{aligned}
e^{3,1}(T) &\le C(T,\beta,\epsilon)\|A^{\frac{\beta-1}{2}}\|^2_{\mathcal{L}_2^0} h^{2\beta-2\epsilon}\int_0^T (T-t)^{-1+\epsilon}\sup_{t\in[0,T]}\mathbb{E}\Big[1+\|X^h(t)\|_E^{5K-6}\Big]dt\\
&\le C(T,X_0,\beta,\epsilon)h^{2\beta-2\epsilon}\left(1+\left(\log\left(\frac{1}{h}\right)\right)^{\frac{(5K-6)K^2}{2}}\right).
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
e^{3,2}(T) &= \frac{1}{2}\left|\mathbb{E}\left[\int_0^T \mathrm{tr}\Big\{A^{-\frac{\beta+1}{2}+\epsilon}(I-P^h)A^{\frac{1-\beta}{2}}A^{\frac{\beta-1}{2}}QA^{\frac{\beta-1}{2}}\right.\right.\\
&\qquad \left.\left. A^{\frac{1-\beta}{2}}P^hA^{\frac{\beta-1}{2}}A^{\frac{1-\beta}{2}}D^2U^{\delta t}(T-t,X^h(t))A^{\frac{1+\beta}{2}-\epsilon}\Big\}\,dt\right]\right|\\
&\le C\mathbb{E}\left[\int_0^T \|A^{\frac{\beta-1}{2}}\|^2_{\mathcal{L}_2^0}\left\|A^{\frac{1-\beta}{2}}(I-P^h)A^{-\frac{1+\beta}{2}+\epsilon}\right\|_{\mathcal{L}(\mathbb{H})}\right\|A^{\frac{1-\beta}{2}}P^h\\
&\qquad \left. A^{\frac{\beta-1}{2}}\right\|_{\mathcal{L}(\mathbb{H})}\left\|A^{\frac{1-\beta}{2}}D^2U^{\delta t}(T-t,X^h(t))A^{\frac{1+\beta}{2}-\epsilon}\right\|_{\mathcal{L}(\mathbb{H})}dt\Bigg]\\
&\le C(T,X_0,\beta,\epsilon)h^{2\beta-2\epsilon}\int_0^T\left\|A^{\frac{1-\beta}{2}}P^hA^{\frac{\beta-1}{2}}\right\|_{\mathcal{L}(\mathbb{H})}(T-t)^{-1+\epsilon}dt\\
&\le C(T,X_0,\beta,\epsilon)h^{2\beta-2\epsilon}\left(1+\left(\log\left(\frac{1}{h}\right)\right)^{\frac{(5K-6)K^2}{2}}\right),
\end{aligned}
$$

where we use the property $\|A^{\frac{\beta-1}{2}}P^hA^{\frac{1-\beta}{2}}\|_{\mathcal{L}(\mathbb{H})}\le C$ proven by the equivalence of norms (5),

$$
\left\|A^{\frac{\beta-1}{2}}P^hA^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathbb{H})}\le C\left\|A_h^{\frac{\beta-1}{2}}P^hA^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathbb{H})}\le C\left\|A^{\frac{\beta-1}{2}}A^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathbb{H})}\le C.
$$

The estimations of $e^{3,1}(T)$ and $e^{3,2}(T)$ indicate

$$
e^3(T)\le C(T,X_0,\beta,\epsilon)h^{2\beta-2\epsilon}\left(1+\left(\log\left(\frac{1}{h}\right)\right)^{\frac{(5K-6)K^2}{2}}\right).
$$

Summing up all the estimations of $\left|\mathbb{E}\left[\phi(X^{\delta t}(T))\right]-\mathbb{E}\left[U^{\delta t}(T,X^h(0))\right]\right|$, $e^1(T)$, $e^2(T)$ and $e^3(T)$, we obtain that for any small $\epsilon_1>\epsilon$,

$$
\begin{aligned}
&\left|\mathbb{E}\Big[\phi(X^{\delta t}(T))-\phi(X^h(T))\Big]\right|\\
&\quad\le C(X_0,T,\epsilon,\beta)h^{2\beta-4\epsilon}\left(1+T^{-\beta}+\left(\log\left(\frac{1}{h}\right)\right)^{\frac{(7K-8)K^2}{2}}\right)\\
&\qquad + C(X_0,T,\epsilon)\delta t\left(1+\left(\log\left(\frac{1}{h}\right)\right)^{\frac{(3K-2)K^2}{2}}\right)\\
&\quad\le C(X_0,T,\beta,\epsilon)\left(h^{2\beta-4\epsilon_1}+\delta t\left(\log\left(\frac{1}{h}\right)\right)^{\frac{(3K-2)K^2}{2}}\right),
\end{aligned}
$$

which, combined with a standard argument, finishes the proof. ∎

Combining Lemma 4.3 with Theorem 4.1, we deduce the essentially sharp weak convergence rate of the finite element method approximating (1). The essentially sharp weak convergence rate is in the sense that the weak convergence rate is essentially twice the strong convergence rate. We remark that even if the logarithmic factor can be eliminated when $\beta > \frac{1}{2}$ or $Q = I$, the weak convergence rate cannot be improved (see, e.g., [2, Theorem 1.1]).

THEOREM 4.2. *Let $T > 0$. Under the assumptions of Theorem 4.1, for $\phi \in \mathcal{C}_b^2(\mathbb{H})$, $\beta \in [0, 1)$, $\gamma < \beta$, there exists $C(X_0, T, \beta, \phi) > 0$ such that*

$$\left|\mathbb{E}\Big[\phi(X(T)) - \phi(X^h(T))\Big]\right| \le C(X_0, T, \beta, \phi)h^{2\gamma}.$$

Proof. By the triangle inequality, Lemma 4.3, and Theorem 4.1 and taking $\delta t = \mathcal{O}(h^{2\beta})$, we have

$$\begin{aligned}
\left|\mathbb{E}\Big[\phi(X(T)) - \phi(X^h(T))\Big]\right| &\le \left|\mathbb{E}\Big[\phi(X(T)) - \phi(X^{\delta t}(T))\Big]\right| \\
&\quad + \left|\mathbb{E}\Big[\phi(X^{\delta t}(T)) - \phi(X^h(T))\Big]\right| \\
&\le C(X_0, T, p)\delta t \\
&\quad + C(X_0, T, \beta)\left(h^{2\gamma} + \delta t\left(\log\left(\frac{1}{h}\right)\right)^{\frac{(3K-2)K^2}{2}}\right) \\
&\le C(X_0, T, p)h^{2\gamma},
\end{aligned}$$

which completes the proof. ∎

Acknowledgment. The authors would like to thank Charles-Edouard Bréhier for his comments based on carefully reading the manuscript.

REFERENCES

[1] A. Andersson, R. Kruse, and S. Larsson, *Duality in refined Sobolev-Malliavin spaces and weak approximation of SPDE*, Stoch. Partial Differ. Equ. Anal. Comput., 4 (2016), pp. 113–149.

[2] A. Andersson and S. Larsson, *Weak convergence for a spatial approximation of the nonlinear stochastic heat equation*, Math. Comp., 85 (2016), pp. 1335–1358.

[3] R. Anton, D. Cohen, and L. Quer-Sardanyons, *A fully discrete approximation of the one-dimensional stochastic heat equation*, IMA J. Numer. Anal., https://doi.org/10.1093/imanum/dry060 (2018).

[4] S. Becker, B. Gess, A. Jentzen, and P. E. Kloeden, *Strong Convergence Rates for Explicit Space-Time Discrete Numerical Approximations of Stochastic Allen–Cahn Equations*, arXiv:1711.02423, 2017.

[5] S. Becker and A. Jentzen, *Strong convergence rates for nonlinearity-truncated Euler-type approximations of stochastic Ginzburg-Landau equations*, Stochastic Process. Appl., 129 (2019), pp. 28–69.

[6] C. E. Bréhier, J. Cui, and J. Hong, *Strong convergence rates of semi-discrete splitting approximations for stochastic Allen–Cahn equation*, IMA J. Numer. Anal., https://doi.org/10.1093/imanum/dry052 (2018).

[7] C. E. Bréhier and A. Debussche, *Kolmogorov equations and weak order analysis for SPDEs with nonlinear diffusion coefficient*, J. Math. Pures Appl. (9), 119 (2018), pp. 193–254.

[8] C. E. Bréhier and L. Goudenège, *Analysis of some splitting Schemes for the stochastic Allen–Cahn equation*, Discrete Contin. Dyn. Syst. Ser. B to appear.

[9] C. E. Bréhier and L. Goudenège, *Weak Convergence Rates of Splitting Schemes for the Stochastic Allen–Cahn Equation*, arXiv:1804.04061, 2018.

[10] D. Conus, A. Jentzen, and R. Kurniawan, *Weak convergence rates of spectral Galerkin approximations for SPDEs with nonlinear diffusion coefficients*, Ann. Appl. Probab., 29 (2019), pp. 653–716, https://doi.org/10.1214/17-AAP1352.

[11] M. Crouzeix, S. Larsson, and V. Thomée, *Resolvent estimates for elliptic finite element operators in one dimension*, Math. Comp., 63 (1994), pp. 121–140.

[12] J. Cui and J. Hong, *Analysis of a splitting scheme for damped stochastic nonlinear Schrödinger equation with multiplicative noise*, SIAM J. Numer. Anal., 56 (2018), pp. 2045–2069.

[13] J. Cui, J. Hong, and Z. Liu, *Strong convergence rate of finite difference approximations for stochastic cubic Schrödinger equations*, J. Differential Equations, 263 (2017), pp. 3687–3713.

[14] J. Cui, J. Hong, Z. Liu, and W. Zhou, *Strong convergence rate of splitting schemes for stochastic nonlinear Schrödinger equations*, J. Differential Equations, 266 (2019), pp. 5625–5663, https://doi.org/10.1016/j.jde.2018.10.034.

[15] J. Cui, J. Hong, and L. Sun, *Strong Convergence Rate of a Full Discretization for Stochastic Cahn–Hilliard Equation Driven by Space-Time White Noise*, arXiv:1812.06289, 2018.

[16] J. Cui, J. Hong, and L. Sun, *Weak Convergence and Invariant Measure of a Full Discretization for Parabolic SPDEs with Non-Globally Lipschitz Coefficients*, arXiv:1811.04075, 2018.

[17] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, 2nd ed., Encyclopedia Math. Appl. 152, Cambridge University Press, Cambridge, UK, 2014.

[18] A. Debussche, *Weak approximation of stochastic partial differential equations: The nonlinear case*, Math. Comp., 80 (2011), pp. 89–117.

[19] X. Feng, Y. Li, and Y. Zhang, *Finite element methods for the stochastic Allen–Cahn equation with gradient-type multiplicative noise*, SIAM J. Numer. Anal., 55 (2017), pp. 194–216.

[20] M. Hefter, A. Jentzen, and R. Kurniawan, *Weak Convergence Rates for Numerical Approximations of Stochastic Partial Differential Equations with Nonlinear Diffusion Coefficients in UMD Banach Spaces*, arXiv:1612.03209, 2016.

[21] M. Kovács, S. Larsson, and F. Lindgren, *Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise*, BIT, 52 (2012), pp. 85–108.

[22] M. Kovács, S. Larsson, and F. Lindgren, *Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise* II. *Fully discrete schemes*, BIT, 53 (2013), pp. 497–525.

[23] M. Kovács, S. Larsson, and F. Lindgren, *On the discretisation in time of the stochastic Allen–Cahn equation*, Math. Nachr., 291 (2018), pp. 966–995.

[24] Z. Liu and Z. Qiao, *Strong approximation of stochastic Allen–Cahn equation with white noise*, IMA J. Numer. Anal., https://doi.org/10.1093/imanum/dry088 (2019).

[25] A. K. Majee and A. Prohl, *Optimal strong rates of convergence for a space-time discretization of the stochastic Allen–Cahn equation with multiplicative noise*, Comput. Methods Appl. Math., 18 (2018), pp. 297–311.

[26] R. Qi and X. Wang, *Optimal error estimates of Galerkin finite element methods for stochastic Allen–Cahn equation with additive noise*, J. Sci. Comput., https://doi.org/10.1007/s10915-019-00973-8 (2019).

[27] E. M. Stein, *Interpolation of linear operators*, Trans. Amer. Math. Soc., 83 (1956), pp. 482–492, https://doi.org/10.2307/1992885.

[28] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, 2nd ed., Springer Ser. Comput. Math. 25, Springer-Verlag, Berlin, 2006.

[29] V. Thomée and L. B. Wahlbin, *Maximum-norm stability and error estimates in Galerkin methods for parabolic equations in one space variable*, Numer. Math., 41 (1983), pp. 345–371.

[30] J. van Neerven, M. C. Veraar, and L. Weis, *Stochastic evolution equations in UMD Banach spaces*, J. Funct. Anal., 255 (2008), pp. 940–993.

[31] J. B. Walsh, *Finite element methods for parabolic stochastic PDE's*, Potential Anal., 23 (2005), pp. 1–43.

[32] X. Wang, *Weak error estimates of the exponential Euler scheme for semi-linear SPDEs without Malliavin calculus*, Discrete Contin. Dyn. Syst., 36 (2016), pp. 481–497.

[33] X. Wang, *An Efficient Explicit Full Discrete Scheme for Strong Approximation of Stochastic Allen–Cahn Equation*, arXiv:1802.09413, 2018.