Results in Applied Mathematics 17 (2023) 100351

Contents lists available at ScienceDirect

Results in Applied Mathematics

journal homepage: www.elsevier.com/locate/results-in-applied-mathematics

Weak convergence of the finite element method for semilinear parabolic SPDEs driven by additive noise

Antoine Tambue [a,*], Jean Daniel Mukam [b]

[a] Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Inndalsveien 28, 5063 Bergen, Norway
[b] Department of Mathematics, Bielefeld University, 33501 Bielefeld, Germany

ARTICLE INFO

Article history:
Received 6 October 2022
Received in revised form 3 December 2022
Accepted 19 December 2022
Available online 24 December 2022

Keywords:
Semilinear parabolic partial differential equations
Finite element method
Weak convergence
Additive noise

ABSTRACT

This paper aims to investigate the finite element weak convergence rate for semilinear parabolic stochastic partial differential equations(SPDEs) driven by additive noise. In contrast to many results in the current scientific literature, we investigate the more general case where the nonlinearity is allowed to be of Nemytskii-type and the linear operator is not necessarily self-adjoint, which is more challenging and models more realistic phenomena such as convection–reaction–diffusion processes. Using Malliavin calculus, Kolmogorov equations and by splitting the linear operator into a self-adjoint and non self-adjoint parts, we prove the convergence of the finite element approximation and obtain a weak convergence rate that is twice the strong convergence rate.

1. Introduction

The purpose of this paper is to analyze the weak convergence of the finite element method for the second order semilinear parabolic SPDEs of the following type in the Hilbert space $\mathcal{H} = L^2(\Lambda)$

$$dX(t) = [AX(t) + F(X(t))]\,dt + dW(t), \quad X(0) = X_0, \quad t \in (0, T]. \tag{1}$$

The model problem (1) is viewed in the Itô sense and the mild solution $X(t)$ is sought in the Hilbert space $\mathcal{H} = L^2(\Lambda)$ where $\Lambda \subset \mathbb{R}^d$ $(d = 1, 2, 3)$ is bounded with smooth boundary or is a convex polygon. In (1), the second order linear operator A is not necessary self-adjoint and generates an analytic semigroup, F is a nonlinear operator satisfying a Lipschitz condition. For technical reasons the initial data $X_0 \in L^2(\Lambda)$ is assumed to be deterministic. The norm and the inner product in $\mathcal{H}$ are denoted respectively by $\|.\|$ and $\langle ., .\rangle$. Let $Q : \mathcal{H} \longrightarrow \mathcal{H}$ be a positive definite self-adjoint operator. In the model problem (1), W is a Q-Wiener process in the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Note that the noise W can be represented as follows [1,2]

$$W(t, x) = \sum_{i=1}^{\infty} \sqrt{q_i} e_i(x)\beta_i(t), \quad t \in [0, T], \quad x \in \Lambda, \tag{2}$$

where e_i and q_i, $i \in \mathbb{N}$ are respectively the eigenfunctions and eigenvalues of Q and β_i $(i \in \mathbb{N})$ are identically distributed standard real-valued Brownian motions. Equations of type (1) are used to model many real-world phenomena such as oil and gas recovery from hydrocarbon reservoirs and mining heat from geothermal reservoirs. However, analytical solutions

* Corresponding author.
E-mail addresses: antonio@aims.ac.za (A. Tambue), jean.d.mukam@aims-senegal.org (J.D. Mukam).

https://doi.org/10.1016/j.rinam.2022.100351

of many SPDEs are rarely known. Therefore, numerical solutions are usually used to provide realistic approximations. Stability and convergence are usually the main features of a numerical algorithm. Convergence of a numerical algorithm aims to ensure that the proposed scheme converges with a certain rate to the true solution with respect to an appropriate norm. The strong convergence is well understood (see e.g. [3–9]) while the weak convergence is not yet well understood, see e.g., [10–19]. The weak convergence of the finite element method for stationary problem driven by spatial white noise was investigated in [16]. In [10], the weak convergence of the finite element method was done for linear evolutive SPDE with a linear self-adjoint operator and the test functions that satisfy an extra condition, namely [10, (1.8)]. Weak convergence of the finite element method for linear SPDEs without the extra condition [10, (1.8)] was later investigated in [20]. The weak convergence in space of semilinear SPDEs driven by multiplicative noise with nonlinear functions F or B is more complicated than the linear case and was investigated for instance in [11] with nonlinear drift function F such that $F \in C_b^2(\mathcal{H})$. Note that the Malliavin calculus and the self-adjointness of the linear operator A is crucial while proving the results in the above mentioned papers. For instance, the proofs in [11] use the projections onto the eigenspaces of A. The drawback was that those projectors do not commute with the L^2-projection P_h in (19). In addition, note that if F is a Nemytskii-type operator (i.e., is a mapping in the form $u \mapsto F(u) = f(u(.))$, where $f \in C^2(\mathbb{R})$), then in general $F \notin C_b^1(\mathcal{H})$, see [11, Section 2.3]. This excludes many interesting nonlinearities F. Therefore, it is interesting to investigate the case where the nonlinearity can be of Nemytskii-type. Recently, Jianbo and Jialin [12] examined the weak convergence in space under one-sided Lipschitz and polynomial growth of the nonlinear drift function F. However, the convergence analysis in [12] was done only for one-dimensional SPDE with linear self-adjoint operator. The case of SPDEs with cubic polynomial nonlinearity has been investigated in [21], but by the means of a spectral Galerkin method. Here the space discretization is performed using the finite element method, more adapted to realistic applications. Andersson et al. [22] used a more general setting by avoiding the Kolmogorov equation and by using the following linearization of the weak error

$$\mathbb{E}[\varphi(X)-\varphi(Y)] = \mathbb{E}[\langle \bar{\varphi}, X-Y\rangle], \quad \text{where} \quad \bar{\varphi} = \int_0^1 \varphi'(\varrho X + (1-\varrho)Y)d\varrho.$$

The function φ being a test function. This approach was introduced in [23] for stochastic differential equations and in [24] for stochastic partial differential equations. Although the setting in [22] seems to be more general, it is not clear how this technique can work for more general second order not necessarily self-adjoint operator in (1) or (16), since the application was only done for self-adjoint operator in [22, Section 5].

The aims of this paper is to fill that gap by providing an elegant weak convergence proof of the finite element method for (1) with not necessary self-adjoint linear operator A and under weaker assumptions on F, namely Assumption 2.2. To achieve our goal, we first write our weak error representation formula in an appropriate form (see Proposition 3.2) by using Kolmogorov equation. We then split the linear operator in self-adjoint and non self-adjoint parts, and use Malliavin calculus along with technical and careful estimates (see e.g. (63)–(70)). Although our approach follows some lines as in [11], error estimate is more challenging since here in addition to the fact our linear operator is not necessarily self-adjoint, we are working under weaker assumptions on the nonlinearity F. Our main result reveals how the weak convergence order depends on the regularity of the noise, and is twice that of the strong convergence. More precisely, we obtain convergence rate $\mathcal{O}\left(h^{2\beta-\epsilon}\right)$, where β is the parameter defined in Assumption 2.2 and $\epsilon > 0$ is an arbitrarily small number.

The rest of this paper is structured as follows. In Section 2 we recall some preliminaries and fundamental functional spaces. Section 3 is devoted to the finite element discretization and weak error representation formula in space. In Section 4, we investigate the weak error of the finite element method.

2. Mathematical setting and preliminaries

2.1. Preliminaries

Let Ω be a sample space and U a separable Hilbert space with norm $\|.\|_U$, we denote by $\mathcal{L}(U,\mathcal{H})$ the space of bounded linear mappings from U to $\mathcal{H}$ endowed with the usual norm $\|.\|_{\mathcal{L}(U,\mathcal{H})}$. We denote by $L^2(\Omega, U)$ the Hilbert space of all equivalence classes of square integrable U-valued random variables. By $C_b^k(\mathcal{H}, U)$ we denote the space of not necessarily bounded mappings g from $\mathcal{H}$ to U that have continuous and bounded Fréchet derivatives $Dg, D^2g, \dots, D^kg$. We endow $C_b^k(\mathcal{H}, U)$ with the semi norm $|.|_{C_b^k(\mathcal{H},U)}$, which for $g \in C_b^k(\mathcal{H}, U)$, $|g|_{C_b^k(\mathcal{H},U)}$ is the smallest constant K such that

$$\sup_{x\in\mathcal{H}} \|D^n g(x)(\phi_1,\dots,\phi_n)\|_U \leq K\|\phi_1\|\cdots\|\phi_n\|, \quad \phi_1,\dots,\phi_n \in \mathcal{H}, \quad n \leq k.$$

Let $\mathcal{G}^m(U, V)$ be the space of Gâteaux differentiable mappings with symmetric and strongly continuous derivative. We denote by $\mathcal{G}_p^m(U, V)$ the subspace of functions of $\mathcal{G}^m(U, V)$ with derivatives of polynomial growth, see e.g., [22, Section 2.1] for more details. We denote by $\mathcal{L}_1(U,\mathcal{H})$ the set of nuclear operators from U to $\mathcal{H}$, $\mathcal{L}_2(U,\mathcal{H}) := HS(U,\mathcal{H})$ the space of Hilbert–Schmidt operators from U to $\mathcal{H}$. As usual, $\mathcal{L}_1(U,\mathcal{H})$ is endowed with the nuclear norm $\|.\|_{\mathcal{L}_1(U,\mathcal{H})}$, see e.g., [25]. For the seek of ease notations, we write $\mathcal{L}(U, U) =: \mathcal{L}(U)$, $\mathcal{L}_1(U, U) =: \mathcal{L}_1(U)$ and $\mathcal{L}_2(U, U) =: \mathcal{L}_2(U)$.

For $l \in \mathcal{L}_1(U)$ the trace of l is defined by

$$Tr(l) := \sum_{i=1}^{\infty} \langle le_i, e_i \rangle_U, \tag{3}$$

where $(e_i)_{i=1}^{\infty}$ is an orthonormal basis of U. The norm in $\mathcal{L}_2(U)$ is defined by

$$\|l\|_{\mathcal{L}_2(U)} := \left(\sum_{i=1}^{\infty} \|le_i\|_U^2 \right)^{\frac{1}{2}} < \infty. \tag{4}$$

The space of Hilbert–Schmidt operators from $U_0 := Q^{\frac{1}{2}}(\mathcal{H})$ to $\mathcal{H}$ is denoted by $\mathcal{L}_2^0 := \mathcal{L}_2(U_0, \mathcal{H}) = HS(U_0, \mathcal{H})$. As usual, $\mathcal{L}_2^0$ is equipped with the norm

$$\|l\|_{\mathcal{L}_2^0} := \|lQ^{\frac{1}{2}}\|_{HS} = \left(\sum_{i=1}^{\infty} \|lQ^{\frac{1}{2}}\phi_i\|^2 \right)^{\frac{1}{2}}, \quad l \in \mathcal{L}_2^0, \tag{5}$$

where $(\phi_i)_{i\in\mathbb{N}}$ is an orthonormal basis of $\mathcal{H}$. Note that definitions (3), (4) and (5) are independent of the orthonormal bases of U and $\mathcal{H}$. The space U_0 equipped with $\langle u, v \rangle_{U_0} := \left\langle Q^{-\frac{1}{2}}u, Q^{-\frac{1}{2}}v \right\rangle$ is a Hilbert space.

Since Malliavin calculus will be one of the key ingredients while examining weak convergence error, let us recall from [11, Section 2] and [12] some of their useful properties. Let $\mathcal{I} : L^2([0,T]; U_0) \longrightarrow L^2(\Omega)$ be an isonormal process, i.e. for any $\varphi \in L^2([0,T]; U_0)$, the random variable $\mathcal{I}(\varphi)$ is centered Gaussian and $\mathcal{I}$ has the following covariance structure

$$\mathbb{E}[\mathcal{I}(\varphi_1), \mathcal{I}(\varphi_2)] = \langle \varphi_1, \varphi_2 \rangle_{L^2([0,T];U_0)}, \quad \varphi_1, \varphi_1 \in L^2([0,T]; U_0).$$

For $u \in U_0$, the Q-Wiener process $W \in L^2\left([0,T] \times U_0, L^2(\Lambda)\right)$ by

$$W(t)u = \mathcal{I}\left(\chi_{[0,t]} \otimes u\right), \quad t \in [0,T], \quad u \in U_0,$$

where χ stands for the indicator function. For $u \in U_0$, the process $W(t)u$, $t \in [0,T]$ is a Brownian motion and satisfies

$$\mathbb{E}\left[W(t)uW(s)v\right] = \min(s,t)\langle u, v \rangle_{U_0}, \quad u, v \in U_0, \quad s, t \in [0,T].$$

For $N \in \mathbb{N}$, let $C_P^{\infty}(\mathbb{R}^N)$ be the space of real-valued C^{∞} functions on $\mathbb{R}^N$ with polynomial growth. We denote the family of smooth real-valued cylindrical random variables by

$$\mathcal{S} := \left\{\chi = g(\mathcal{I}(\varphi_1), \ldots, \mathcal{I}(\varphi_N)) : g \in C_P^{\infty}(\mathbb{R}^N);\ \varphi_j \in L^2([0,T]; U_0);\ j = 1, \ldots, N\right\}$$

and the corresponding family with values in $\mathcal{H}$ by

$$\mathcal{S}(\mathcal{H}) := \left\{ F = \sum_{i=1}^{L} \chi_i \otimes h_i : \chi_1, \ldots, \chi_L \in \mathcal{S},\ h_1, \ldots, h_L \in \mathcal{H},\ L \in \mathbb{N} \right\}.$$

The Malliavin derivative of the random variable $\chi = g\left(\mathcal{I}(\varphi_1), \ldots, \varphi_N\right) \in \mathcal{S}$ is defined as the $L^2([0,T]; U_0)$-valued random variable

$$\mathcal{D}\chi = \sum_{i=1}^{N} \partial_i g\left(\mathcal{I}(\varphi_1), \ldots, \mathcal{I}(\varphi_N)\right) \otimes \varphi_i.$$

Obviously $\mathcal{D}\chi$ is an U_0-valued stochastic process. We write for $t \geq 0$

$$\mathcal{D}_t\chi = \sum_{i=1}^{N} \partial_i g\left(\mathcal{I}(\varphi_1), \ldots, \mathcal{I}(\varphi_N)\right) \otimes \varphi_i(t).$$

The Malliavin derivative of $G = \sum_{i=1}^{L} g_i\left(\mathcal{I}(\varphi_1), \ldots, \mathcal{I}(\varphi_N)\right) \otimes h_i \in \mathcal{S}(\mathcal{H})$ is defined by

$$\mathcal{D}_s G = \sum_{i=1}^{L} \sum_{j=1}^{N} \partial_j g_i\left(\mathcal{I}(\varphi_1), \ldots, \mathcal{I}(\varphi_N)\right) \otimes \left(h_i \otimes \varphi_j(s)\right).$$

Since the Malliavin derivative operator $\mathcal{D}$ is closable (see e.g., [11, Section 2]), let us denote by $\mathbb{D}^{1,2}$ the closure of $\mathcal{S}(\mathcal{H})$ with respect to the Malliavin derivative equipped with the following norm

$$\|G\|_{\mathbb{D}^{1,2}(\mathcal{H})}^2 := \left(\mathbb{E}\left[\|G\|^2\right] + \mathbb{E}\left[\int_0^T \|\mathcal{D}_s G\|_{\mathcal{L}_2^0}^2 ds\right] \right)^{\frac{1}{2}}.$$

For any $\sigma \in C_b^1(\mathcal{H}, \mathcal{H})$, the following chain rules for Malliavin derivatives hold

$$\mathcal{D}_t(\sigma(G)) = \mathcal{D}\sigma(G)\mathcal{D}_t G, \ \mathcal{D}_t^u(\sigma(G)) = \mathcal{D}\sigma(G).\mathcal{D}_t^u G, \ u \in U_0, \ G \in \mathbb{D}^{1,2}(\mathcal{H}), \tag{6}$$

where $\mathcal{D}_t^u := \mathcal{D}_t G u$ is the derivative in the direction u, with

$$\mathcal{D}_t G u = \sum_{i=1}^{L}\sum_{j=1}^{N} \langle u, \varphi_j\rangle_{U_0} \partial_j f_i \left(\mathcal{I}(\varphi_1), \ldots, \mathcal{I}(\varphi_N)\right) \otimes h_i,$$

where we equipped U_0 with the inner product $\langle u, v\rangle_{U_0} := \left\langle Q^{-\frac{1}{2}} u, Q^{-\frac{1}{2}} v\right\rangle$.

Moreover, for any random variable $G \in \mathbb{D}^{1,2}(\mathcal{H})$ and any predictable stochastic process $\Phi \in L^2([0,T], \mathcal{L}_2^0)$, the following holds

$$\mathbb{E}\left[\left\langle \int_0^T \Phi(t)dW(t), G\right\rangle\right] = \mathbb{E}\left[\int_0^T \langle \Phi(t), \mathcal{D}_t G\rangle_{\mathcal{L}_2^0} dt\right]. \tag{7}$$

In the rest of this paper, we consider $\mathcal{H} = L^2(\Lambda)$.

2.2. Main assumptions and well-defined problem

To guarantee a unique mild solution of (1) and for the purpose of the convergence analysis, we make the following assumptions.

Assumption 2.1. The linear operator $A : \mathcal{D}(A) \subset \mathcal{H} \longrightarrow \mathcal{H}$ is negative and generates an analytic semigroup $S(t) =: e^{At}$.

Assumption 2.2. The covariance operator $Q : \mathcal{H} \longrightarrow \mathcal{H}$ satisfies the following estimate

$$\left\|(-A)^{\frac{\beta-1}{2}} Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})} \leq C, \quad \beta \in [0,1].$$

Assumption 2.3. The nonlinear operator $F : \mathcal{H} \longrightarrow \mathcal{H}$ is Lipschitz continuous and twice Fréchèt differentiable, with derivatives satisfying

$$\|(-A)^{-\frac{\delta}{2}} F'(v)u\| \leq L\|u\|, \quad \|(-A)^{-\eta} F''(v)(u_1, u_2)\| \leq L\|u_1\|\|u_2\|, \quad u, v, u_1, u_2 \in \mathcal{H},$$

for some $\delta \in (0, \beta)$ and $\eta \in (\frac{1}{2}, 1)$.

The following theorem guarantees the existence of the unique mild solution to (1).

Theorem 2.1 ([1, *Theorem 7.2*]). *Let Assumptions 2.1, 2.3 and 2.2 be fulfilled. Then the SPDE (1) has up to modifications a unique mild solution $X : [0,T] \times \Omega \longrightarrow \mathcal{H}$, which takes the following form*

$$X(t) = S(t)X_0 + \int_0^t S(t-s)F(X(s))ds + \int_0^t S(t-s)dW(s) \tag{8}$$

$\mathbb{P}$-a.s. and satisfies

$$\mathbb{P}\left[\int_0^T \|X(s)\|^2 ds < \infty\right] = 1.$$

Moreover, for all $p \geq 1$ there exists a positive constant C such that.

$$\sup_{0\leq t\leq T} \|X(t)\|_{L^2(\Omega,\mathcal{H})} \leq C\left(1 + \|X_0\|\right), \quad \sup_{0\leq t\leq T} \|X(t)\|^{\frac{p}{2}}_{L^2(\Omega,\mathcal{H})} \leq C\left(1 + \|X_0\|^{\frac{p}{2}}\right), \tag{9}$$

For an $\mathcal{L}_2^0$-valued predictable stochastic process $\phi : [0,T] \times \Omega \longrightarrow \mathcal{L}_2^0$ such that

$$\int_0^t \mathbb{E}\|\phi(s)Q^{\frac{1}{2}}\|^2_{\mathcal{L}_2(\mathcal{H})} ds < \infty, \quad t \in [0,T],$$

the following relation called Itô isometry holds

$$\mathbb{E}\left\|\int_0^t \phi(s)dW(s)\right\|^2 = \int_0^t \mathbb{E}\|\phi(s)Q^{\frac{1}{2}}\|^2_{\mathcal{L}_2(\mathcal{H})} ds, \quad t \in [0,T], \tag{10}$$

see e.g., [1, Step 2 in Section 2.3.2] or [2, Proposition 2.3.5].

In the sequel of this paper, the following proposition will be used.

Proposition 2.1 (*[25]*). *Let l, $l_1.l_2$ be three operators in Hilbert spaces, the following holds*

(i) If $l \in \mathcal{L}_1(U)$ then

$$|Tr(l)| \leq \|l\|_{\mathcal{L}_1(U)}. \tag{11}$$

(ii) If $l_1 \in \mathcal{L}(\mathcal{H})$ and $l_2 \in \mathcal{L}_1(\mathcal{H})$, then $l_1 l_2, l_2 l_1 \in \mathcal{L}_1(\mathcal{H})$ and

$$Tr(l_1 l_2) = Tr(l_2 l_1). \tag{12}$$

(iii) If $l_1 \in \mathcal{L}_2(U, \mathcal{H})$ and $l_2 \in \mathcal{L}_2(\mathcal{H}, U)$, then $l_1 l_2 \in \mathcal{L}_1(\mathcal{H})$ and

$$\|l_1 l_2\|_{\mathcal{L}_1(\mathcal{H})} \leq \|l_1\|_{\mathcal{L}_2(U,\mathcal{H})} \|l_2\|_{\mathcal{L}_2(\mathcal{H},U)}. \tag{13}$$

(iv) If $l \in \mathcal{L}(U, \mathcal{H})$ and $l_j \in \mathcal{L}_j(U)$, $j = 1, 2$, then $ll_j \in \mathcal{L}_j(U, \mathcal{H})$ and

$$\|ll_j\|_{\mathcal{L}_j(U,\mathcal{H})} \leq \|l\|_{\mathcal{L}(U,\mathcal{H})} \|l_j\|_{\mathcal{L}_j(U)}, \quad j = 1, 2. \tag{14}$$

(v) If $l \in \mathcal{L}_2(U, \mathcal{H})$, then its adjoint $l^* \in \mathcal{L}_2(\mathcal{H}, U)$ and

$$\|l^*\|_{\mathcal{L}_2(\mathcal{H},U)} = \|l\|_{\mathcal{L}_2(U,\mathcal{H})}. \tag{15}$$

We equip $\dot{\mathcal{H}}^\alpha := \mathcal{D}((-A))^{\frac{\alpha}{2}}$, $\alpha \in \mathbb{R}$ with the norm $\|u\|_\alpha := \|(-A)^{\frac{\alpha}{2}} u\|$.

2.3. Second order semilinear parabolic SPDE

In the rest of this paper, we assume the linear operator A to be of second order. More precisely, we assume that our SPDE (1) is a second order semilinear parabolic of the form

$$dX(t,x) = [\nabla \cdot (\mathbf{D}\nabla X(t,x)) - \mathbf{q} \cdot \nabla X(t,x) + f(x, X(t,x))]dt + dW(t,x), \tag{16}$$

for $x \in \Lambda$ and $t \in [0,T]$, where the function $f : \Lambda \times \mathbb{R} \longrightarrow \mathbb{R}$ is twice continuously differentiable with globally bounded derivatives. In the abstract framework (1), the linear operator A is the $L^2(\Lambda)$ realization [26, p. 812] of the following differential operator

$$\mathcal{A}u = -\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(D_{ij}(x)\frac{\partial u}{\partial x_j}\right) + \sum_{i=1}^{d} q_i(x)\frac{\partial u}{\partial x_i}, \quad \mathbf{D} := \left(D_{i,j}\right)_{1\leq i,j\leq d}, \tag{17}$$

$\mathbf{q} := (q_i)_{1\leq i\leq d}$, where $D_{ij} \in L^\infty(\Lambda)$, $q_i \in L^\infty(\Lambda)$ and there exists a constant $c_1 > 0$ such that

$$\sum_{i,j=1}^{d} D_{ij}(x)\xi_i\xi_j \geq c_1|\xi|^2, \quad \xi \in \mathbb{R}^d, \quad x \in \overline{\Lambda},$$

the nonlinear function $F : \mathcal{H} \longrightarrow \mathcal{H}$ is given by

$$(F(v))(x) = f(x, v(x)), \quad x \in \Lambda, \quad v \in \mathcal{H}, \tag{18}$$

where $f : \Lambda \times \mathbb{R} \longmapsto \mathbb{R}$ is a smooth nonlinear function. If there exists $c_f \geq 0$ such that

$$|f(\xi,z)| \leq c_f(|z|+1), \quad \left|\frac{\partial f}{\partial z}(\xi,z)\right| \leq c_f, \quad \left|\frac{\partial^2 f}{\partial \xi \partial z}(\xi,z)\right| \leq c_f, \quad \left|\frac{\partial^2 f}{\partial^2 z}(\xi,z)\right| \leq c_f,$$

for all $z \in \mathbb{R}$, $i = 1,\ldots,d$, $\xi = (\xi_1, \xi_2, \ldots, \xi_d)^T \in \Lambda$, then Assumption 2.3 is fulfilled. See e.g., [19, Example 3.2] or [27, Example 5.1] for details.

As in [7,26], we introduce two spaces $\mathbb{H}$ and V, such that $\mathbb{H} \subset V$; the two spaces depend on the boundary conditions of Λ and the domain of the operator A. For Dirichlet (or first-type) boundary conditions we take

$$V = \mathbb{H} = H_0^1(\Lambda) = \{v \in H^1(\Lambda) : v = 0 \quad \text{on} \quad \partial\Lambda\}.$$

For Robin (third-type) boundary condition and Neumann (second-type) boundary condition, which is a special case of Robin boundary condition, we take $V = H^1(\Lambda)$ and

$$\mathbb{H} = \{v \in H^2(\Lambda) : \partial v/\partial \mathrm{v}_{\mathcal{A}} + \alpha_0 v = 0, \quad \text{on} \quad \partial\Lambda\}, \quad \alpha_0 \in \mathbb{R},$$

where $\partial v/\partial \mathrm{v}_{\mathcal{A}}$ is the normal derivative of v and $\mathrm{v}_{\mathcal{A}}$ is the exterior pointing normal $n = (n_i)$ to the boundary of $\mathcal{A}$, given by

$$\partial v/\partial \mathrm{v}_{\mathcal{A}} = \sum_{i,j=1}^{d} n_i(x) D_{ij}(x)\frac{\partial v}{\partial x_j}, \qquad x \in \partial\Lambda.$$

Using Green's formula and the boundary conditions, the corresponding bilinear form associated to $-\mathcal{A}$ is given by

$$a(u,v)=\int_{\Lambda}\left(\sum_{i,j=1}^{d}D_{ij}\frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j}+\sum_{i=1}^{d}q_i\frac{\partial u}{\partial x_i}v\right)dx,\quad u,v\in V,$$

for Dirichlet and Neumann boundary conditions, and

$$a(u,v)=\int_{\Lambda}\left(\sum_{i,j=1}^{d}D_{ij}\frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_j}+\sum_{i=1}^{d}q_i\frac{\partial u}{\partial x_i}v\right)dx+\int_{\partial\Lambda}\alpha_0 uv dx,\quad u,v\in V.$$

for Robin boundary conditions. Note that A A is the infinitesimal generator of a bounded analytic semigroup $S(t)=e^{tA}$ on $L^2(\Lambda)$ such that

$$S(t)=e^{tA}=\frac{1}{2\pi i}\int_{\mathcal{C}}e^{t\lambda}(\lambda I-A)^{-1}d\lambda,\qquad t>0,$$

where $\mathcal{C}$ denotes a path that surrounds the spectrum of A, see e.g., [26,28].

3. Finite element discretization and weak error representation formula

Let us now turn our attention to the discretization of problem (1). We start by splitting the domain Λ in finite triangles. Let $\mathcal{T}_h$ be the triangulation with maximal length h satisfying the usual regularity assumptions. Let $V_h\subset V$ be the space of continuous functions that are piecewise linear over the triangulation $\mathcal{T}_h$. We consider the projection P_h from $\mathcal{H}$ to V_h defined by

$$\langle P_h u,\chi\rangle=\langle u,\chi\rangle,\quad u\in\mathcal{H},\quad \chi\in V_h. \tag{19}$$

The discrete operator $A_h: V_h\longrightarrow V_h$ is defined by

$$\langle -A_h\phi,\chi\rangle=\langle(-A)^{1/2}\phi,(-A)^{*1/2}\chi\rangle=a(\phi,\chi),\quad \phi,\chi\in V_h, \tag{20}$$

Note that $(-A)^{*1/2}$ stands for the adjoint of $(-A)^{1/2}$. Like A, A_h is also a generator of analytic semigroup $S_h(t):=e^{tA_h}$. As any semigroup and its generator, A_h and $S_h(t)$ satisfy the smoothing properties of Theorem 2.1 with a uniform constant C, independent of h. The semi-discrete version in space of problem (1) consists of finding $X^h(t)\in V_h$, $t\in(0,T]$ such that $X^h(0)=P_hX_0$ and

$$dX^h(t)=[A_hX^h(t)+P_hF(X^h(t))]dt+P_hdW(t),\quad t\in(0,T]. \tag{21}$$

We note that A_h and P_hF satisfy the same assumptions as A and F respectively. Therefore, Theorem 2.1 ensures the existence of the unique mild solution $X^h(t)$ of (21) such that

$$\|X^h(t)\|_{L^2(\Omega,\mathcal{H})}\leq C\left(1+\|X_0\|\right),\quad \|X^h(t)\|^{\frac{p}{2}}_{L^2(\Omega,\mathcal{H})}\leq C\left(1+\|X_0\|^{\frac{p}{2}}\right),\quad t\in[0,T], \tag{22}$$

for all $p\geq 2$. The mild solution of (21) is given by

$$X^h(t)=S_h(t)X^h(0)+\int_0^t S_h(t-s)P_hF(X^h(s))ds+\int_0^t S_h(t-s)P_hdW(s). \tag{23}$$

Let us introduce the Ritz representation operator $R^h: V\longrightarrow V_h$, defined by

$$\langle AR^hv,\chi\rangle=-\langle(-A)^{1/2}v,(-A)^{*1/2}\chi\rangle=-a(v,\chi),\quad v\in V,\quad \chi\in V_h. \tag{24}$$

As in [17,29], we split A as follows: $A=A_1+A_2$, where A_1 and A_2 are respectively the self-adjoint and the non-selfadjoint parts of A. Note that formally A_1 corresponds to the second order derivative part of (17) and A_2 corresponds to the first order derivative. Note that $\mathcal{D}((-A)^{\frac{1}{2}})=\mathcal{D}((-A_1)^{\frac{1}{2}})=\mathcal{D}(A_2)=V$, where $V=H^1_0(\Lambda)$ for Dirichlet boundary conditions and $V=H^1(\Lambda)$ for Neumann or Robin boundary condition, see e.g., [29–31]. Note that the range of A_2 is a subspace of $\mathcal{H}$. We denote by $A_{1,h}$ and $A_{2,h}$ the discrete version of A_1 and A_2 respectively, see e.g., [29]. Note that $A_{2,h}: V_h\longrightarrow V_h$ satisfies

$$\langle A_{2,h}v^h,\chi\rangle=\langle A_2v^h,\chi\rangle,\quad v^h,\chi\in V_h. \tag{25}$$

In the rest of this paper, for $u^h\in V_h$ and $v\in\mathcal{D}((-A)^{\frac{1}{2}})$, the notation $\langle -Au^h,v\rangle$ is understood in the following sense

$$\langle -Au^h,v\rangle:=\langle(-A)^{\frac{1}{2}}u^h,(-A^*)^{\frac{1}{2}}v\rangle,\quad u^h\in V_h,\quad v\in\mathcal{D}((-A^*)^{\frac{1}{2}}). \tag{26}$$

We also introduce the Ritz representation operator $R^h_1: V\longrightarrow V_h$, defined by

$$\langle A_1R^h_1v,\chi\rangle=-\langle(-A_1)^{1/2}v,(-A_1)^{1/2}\chi\rangle,\quad v\in V,\quad \chi\in V_h. \tag{27}$$

We introduce the operator $R_2^h : V \longrightarrow V_h$ defined by

$$R_{2,h}v = A_{2,h}^{-1}P_hA_2v, \quad \text{i.e.,} \quad A_{2,h}R_2^hv = P_hA_2v, \quad v \in V, \tag{28}$$

where $A_{2,h}^{-1}$ is the pseudo-inverse of $A_{2,h}$ [1, Appendix B.2]. From (25), (28) and the definition of P_h (cf. (19)), it follows that

$$\langle A_2R_2^hv, \chi\rangle = \langle A_{2,h}R_2^hv, \chi\rangle = \langle P_hA_2v, \chi\rangle = \langle A_2v, \chi\rangle, \quad v \in V, \ \chi \in V_h. \tag{29}$$

It is well known that the following estimate holds (see e.g., [11,26] or [32, Chapter 3])

$$\left\|(-A)^{\alpha}(P_h - \mathbf{I})(-A)^{-\eta}\right\|_{\mathcal{L}(\mathcal{H})} \leq Ch^{2\eta-2\alpha}, \quad 0 \leq \alpha \leq \eta \leq 1, \tag{30}$$

$$\left\|(-A)^{\delta}(R^h - \mathbf{I})(-A)^{-\gamma}\right\|_{\mathcal{L}(\mathcal{H})} \leq Ch^{2\gamma-2\delta}, \quad 0 \leq \delta \leq \frac{1}{2} \leq \gamma \leq 1, \tag{31}$$

$$\left\|(-A_1)^{\delta}(R_1^h - \mathbf{I})(-A_1)^{-\gamma}\right\|_{\mathcal{L}(\mathcal{H})} \leq Ch^{2\gamma-2\delta}, \quad 0 \leq \delta \leq \frac{1}{2} \leq \gamma \leq 1. \tag{32}$$

Let us recall the following lemma, which will be useful in the rest of the paper.

Lemma 3.1 (*[17, Lemma 3.1]*). *For $\rho \in [0, 1]$, $(-A)^{\rho}(-A^*)^{-\rho}$, $(-A^*)^{\rho}(-A)^{-\rho}$, $(-A_h)^{\rho}(-A_h^*)^{-\rho}$ and $(-A_h^*)^{\rho}(-A_h)^{-\rho}$ are bounded operators in $\mathcal{H}$.*

Throughout this paper, C is a generic constant that may change from one place to another but is independent of the space discretization parameter h.

Assumption 3.1. The test function $\varphi \in \mathcal{G}_b^2(\mathcal{H}, \mathbb{R})$ is such that there exists some constants $m \geq 2$ and $C \geq 0$ such that

$$\|\varphi^{(j)}(u)\|_{\mathcal{L}^{[j]}(\mathcal{H};\mathbb{R})} \leq C\left(1 + \|u\|^{m-j}\right), \quad u \in \mathcal{H}, \quad j = 1, 2.$$

Let us now move to the weak error representation formula. For $\varphi \in \mathcal{G}_b^2(\mathcal{H}, \mathbb{R})$, we define

$$\mu(t, \psi) := \mathbb{E}\left(\varphi(X(t, \psi))\right), \tag{33}$$

where $X(t, \psi)$ is the mild solution of (1) with initial value ψ.

Proposition 3.1. *Under Assumptions 2.1, 2.2 and 2.3, $\mu(t, \psi)$ (33) is the unique strict solution to the following deterministic PDE, called Kolmogorov equation*

$$\begin{cases} \frac{\partial \mu}{\partial t}(t, \psi) = \langle A\psi + F(\psi), D\mu(t, \psi)\rangle + \frac{1}{2}Tr\left[D^2\mu(t, \psi)Q\right], \quad \psi \in \mathcal{D}(A), \\ \mu(0, \psi) = \varphi(\psi), \quad \psi \in \mathcal{D}(A). \end{cases} \tag{34}$$

Proof. The proof goes along the same lines as that of [1, Theorem 9.25]. Note that compared to [1, Theorem 9.25], here we have assumed weaker assumptions on the nonlinearity F and on the noise W, which is assumed to be of trace class in [1, Theorem 9.25]. However, the proof is exactly as in [1, Theorem 9.25]. In fact, the main ingredients in proving [1, Theorem 9.25] are [1, Theorems 9.8 & 9.9], which themselves are consequences of [1, Lemma 9.2, Propositions 9.5 & 9.6]. Note that [1, Lemma 9.2] only requires Lipschitz condition on F and does not involves their first and second derivatives, so it remains valid in our setting. As far as it concerns [1, Proposition 9.5 & Proposition 9.6], the main point is to estimate

$$\sup_{t\in[0,T]} \mathbb{E}\left(\int_0^t \|S(t-s)[F'(s, X_n(s)) - F'(s, X)].Y(s)\|ds\right)^p,$$

where X_n is a sequence of processes converging to X in $\mathscr{H}_p$, with $\mathscr{H}_p$ being the Banach space of all (equivalence classes of) $\mathcal{H}$-valued predictable stochastic processes Y defined on the time interval $[0, T]$ with the norm $|||Y||| := \left(\sup_{t\in[0,T]} \|Y(t)\|^p\right)^{\frac{1}{p}}$. In our setting, using Assumption 2.3 leads to

$$\begin{aligned}
&\sup_{t\in[0,T]} \mathbb{E}\left(\int_0^t \|S(t-s)[F'(s, X_n(s)) - F'(s, X)].Y(s)\|ds\right)^p \\
&\leq \sup_{t\in[0,T]} \mathbb{E}\left(\int_0^t \|(-A)^{\frac{\delta}{2}}S(t-s)\|_{\mathcal{L}(\mathcal{H})}\|(-A)^{-\frac{\delta}{2}}[F'(s, X_n(s)) - F'(s, X)].Y(s)\|ds\right)^p \\
&\leq \sup_{t\in[0,T]} \mathbb{E}\left(\int_0^t (t-s)^{-\frac{\delta}{2}}\|Y(s)\|ds\right)^p \leq \sup_{t\in[0,T]} \mathbb{E}[\sup_{r\in[0,t]} \|Y(r)\|^p]\left(\int_0^t (t-s)^{-\frac{\delta}{2}}ds\right)^p \\
&\leq C\mathbb{E}[\sup_{t\in[0,T]} \|Y(t)\|^p] < \infty,
\end{aligned} \tag{35}$$

where at the last step we used the fact that $Y \in \mathscr{H}_p$. This then shows the boundedness of the term I_n^1 in [1, Theorem 9.6] and then the dominated convergence theorem can be applied. In our setting, since we are dealing with additive noise, the term I_n^2 in [1, theorem 9.6] vanishes and we get rid of the noise. The same argument as above applies in the rest of the theory and we will not repeat it here. ■

Remark 3.1. Along the same lines as the proof of [1, Theorem 9.25] (or Proposition 3.1), one can prove that for $\psi \in \mathcal{D}((-A)^{\frac{1}{2}})$, $\mu(t,\psi)$ is a strict solution to the following deterministic PDE

$$\begin{cases} \frac{\partial \mu}{\partial t}(t,\psi) = -\left\langle (-A)^{\frac{1}{2}}\psi, (-A^*)^{\frac{1}{2}}D\mu(t,\psi)\right\rangle + \langle F(\psi), D\mu(t,\psi)\rangle + \frac{1}{2}Tr\left[D^2\mu(t,\psi)Q\right], \\ \mu(0,\psi) = \varphi(\psi), \quad \psi \in \mathcal{D}((-A)^{\frac{1}{2}}). \end{cases} \tag{36}$$

In fact, (36) is just a weak formulation of (34) by noticing that $\langle A\psi, D\mu(t,\psi)\rangle = -\langle -A\psi, D\mu(t,\psi)\rangle = -\left\langle (-A)^{\frac{1}{2}}\psi, (-A^*)^{\frac{1}{2}}D\mu(t,\psi)\right\rangle$.

Bearing (26) in mind, (36) can be written in the same form as (34) with $\psi \in \mathcal{D}(A^{\frac{1}{2}})$. Using the Riesz representation theorem, we can identify the first order derivative of $\mu(t,\psi)$ with respect to $\psi \in \mathcal{H}$ denoted by $D\mu(t,\psi)$ with an element of $\mathcal{H}$. This yields

$$D\mu(t,\psi)(\phi_1) := \langle D\mu(t,\psi), \phi_1\rangle, \quad \phi_1 \in \mathcal{H}.$$

By the same arguments as in [1, Theorem 9.4, Remark 9.5] and (35), one shows that $X(t,\psi)$ is twice differentiable with respect to $\psi \in \mathcal{H}$ and the second derivative $D^2\mu(t,\psi)$ with a linear operator in $\mathcal{H}$. Hence taking the derivative in both sides of (33) yields

$$\langle D\mu(t,\psi), \phi_1\rangle = \mathbb{E}\left(\varphi'(X(t,\psi)).\xi_1(t)\right), \tag{37}$$

where $\xi_1(t)$ is the mild solution of the following problem

$$d\xi_1(t) = [A\xi_1(t) + F'(X(t,\psi)).\xi_1(t)]dt, \quad \xi_1(0) = \phi_1. \tag{38}$$

Differentiating again (37) yields

$$\left\langle D^2\mu(t,\psi)\phi_1, \phi_2\right\rangle = \mathbb{E}\left(\varphi''(X(t,\psi)).(\xi_1(t),\xi_2(t))\right) + \mathbb{E}\left(\varphi'(X(t,\psi)).\eta_{1,2}(t)\right), \tag{39}$$

where $\xi_1(t)$ and $\xi_2(t)$ satisfy (38) with initial values ϕ_1 and ϕ_2 respectively, and $\eta_{1,2}$ satisfies

$$d\eta_{1,2}(t) = \left[A\eta_{1,2}(t) + F'(X(t,\psi)).\eta_{1,2}(t) + F''(X(t,\psi))(\xi_1(t),\xi_2(t))\right]dt, \quad \eta_{1,2}(0) = 0. \tag{40}$$

Proposition 3.2 (*Weak Error Representation Formula*). *Let Assumptions 2.1, 2.3 and 2.2 be fulfilled. For any $\varphi \in C_b^2(\mathcal{H},\mathbb{R})$ the following weak error representation formula holds*

$$\begin{aligned}\mathbb{E}[\varphi(X^h(T)) - \varphi(X(T))] &= \mathbb{E}[\mu(T,X^h(0)) - \mu(T,X(0))] + \mathbb{E}\int_0^T \langle A_hX^h(s), D\mu(T-s,X^h(s))\rangle ds \\ &+ \mathbb{E}\int_0^T \left\langle (-A)^{\frac{1}{2}}X^h(s), (-A^*)^{\frac{1}{2}}D\mu(T-s,X^h(s))\right\rangle ds \\ &+ \mathbb{E}\int_0^T \left\langle P_hF\left(X^h(s)\right) - F(X^h(s)), D\mu(T-s,X^h(s))\right\rangle ds \\ &+ \frac{1}{2}\mathbb{E}\int_0^T Tr\left[D^2\mu(T-s,X^h(s))\left(P_hQP_h - Q\right)\right]ds.\end{aligned}$$

Proof. Let us introduce the following shift process

$$v(t,X_0) := \mu(T-t,X_0), \quad 0 \le t \le T, \tag{41}$$

where μ is defined by (33) with $\psi = X_0$. Simple computations yields

$$v(0,X(0)) = \mu(T,X(0)) = \mathbb{E}(\varphi(X(T))), \quad v(T,X^h(T)) = \mu(0,X^h(T)) = \mathbb{E}\left(g(X^h(T))\right).$$

Consequently, we have the following decomposition of the space weak error

$$\begin{aligned}\mathbb{E}\left(\varphi\left(X^h(T)\right) - \varphi(X(T))\right) &= \mathbb{E}\left(v(T,X^h(T)) - v(0,X(0))\right) \\ &= \mathbb{E}\left(v(0,X^h(0)) - v(0,X(0))\right) + \mathbb{E}\left(v(T,X^h(T)) - v(X^h(0),0)\right).\end{aligned} \tag{42}$$

Applying the Itô formula [1, Theorem 4.32] to $G(t,x) = v(t, X^h(t))$ in the interval $[0, T]$ yields

$$\begin{aligned} v(T, X^h(T)) - v(0, X^h(0)) &= \int_0^T Dv(s, X^h(s))P_h dW(s) \\ &+ \int_0^T \left[\frac{\partial v}{\partial t}(s, X^h(s)) + \left\langle Dv(s, X^h(s)), A_h X^h(s) + P_h F(X^h(s))\right\rangle\right] ds \\ &+ \frac{1}{2} Tr \left[D^2 v(s, X^h(s)) \left(P_h Q^{\frac{1}{2}}\right)\left(P_h Q^{\frac{1}{2}}\right)^*\right] ds. \end{aligned} \tag{43}$$

From the fact $v(s, X^h(s))$ satisfies (36), it follows that

$$\begin{aligned} \frac{\partial v}{\partial t}(s, X^h(s)) &= -\frac{\partial u}{\partial t}(T-s, X^h(s)) = \left\langle (-A)^{\frac{1}{2}} X^h(s), (-A^*)^{\frac{1}{2}} D\mu(T-s, X^h(s))\right\rangle \\ &- \left\langle F(X^h(s)), D\mu(T-s, X^h(s))\right\rangle - \frac{1}{2} Tr\left[D^2\mu(T-s, X^h(s)) Q^{\frac{1}{2}} \left(Q^{\frac{1}{2}}\right)^*\right]. \end{aligned} \tag{44}$$

Substituting (44) in (43) yields

$$\begin{aligned} v(T, X^h(T)) - v(0, X^h(0)) &= \int_0^T D\mu(s, X^h(s))P_h dW(s) + \int_0^T \left\langle A_h X^h(s), D\mu(T-s, X^h(s))\right\rangle ds \\ &+ \int_0^T \left\langle (-A)^{\frac{1}{2}} X^h(s), (-A^*)^{\frac{1}{2}} D\mu(T-s, X^h(s))\right\rangle ds \\ &+ \int_0^T \left\langle P_h F\left(X^h(s)\right) - F(X^h(s)), D\mu(T-s, X^h(s))\right\rangle ds \\ &+ \frac{1}{2}\int_0^T Tr\left[D^2\mu(T-s, X^h(s))\left(P_h Q^{\frac{1}{2}} \left(Q^{\frac{1}{2}}\right)^* P_h^* - Q^{\frac{1}{2}}\left(Q^{\frac{1}{2}}\right)^*\right)\right] ds. \end{aligned} \tag{45}$$

Taking the expectation in both sides of (45) and using the fact that the expectation of the Itô integral vanishes, we arrive at

$$\begin{aligned} \mathbb{E}\left[v(T, X^h(T)) - v(0, X^h(0))\right] &= \mathbb{E}\int_0^T \left\langle A_h X^h(s), D\mu(T-s, X^h(s))\right\rangle ds \\ &+ \mathbb{E}\int_0^T \left\langle (-A)^{\frac{1}{2}} X^h(s), (-A^*)^{\frac{1}{2}} D\mu(T-s, X^h(s))\right\rangle ds \\ &+ \mathbb{E}\int_0^T \left\langle P_h F\left(X^h(s)\right) - F(X^h(s)), D\mu(T-s, X^h(s))\right\rangle ds \\ &+ \frac{1}{2}\mathbb{E}\int_0^T Tr\left[D^2\mu(T-s, X^h(s))\left(P_h Q P_h - Q\right)\right] ds. \end{aligned} \tag{46}$$

Substituting (46) in (42) completes the proof of Proposition 3.2. ■

4. Weak error estimate

The main results of this section is formulated in the following theorem.

Theorem 4.1. *Let $X(t)$ and $X^h(t)$ be the mild solution of* (1) and (21) *respectively. Let Assumptions 2.3, 2.2 and 3.1 be fulfilled. Then the following error estimate holds*

$$|\mathbb{E}[\varphi(X(T)) - \varphi(X^h(T))]| \leq Ch^{2\beta-\epsilon},$$

where φ is any test function satisfying Assumption 2.2 and C is a constant independent of h.

In order to prove Theorem 4.1, we need some preliminaries results.

4.1. Preliminaries estimates

The proof of Theorem 4.1 relies heavily on the regularity estimates of the solution to the Kolmogorov equation (34), which we provide in the next lemma. Such regularity results were also obtained for instance in [18]. But here we emphasize that we are working under weaker assumptions on the nonlinearity F.

Lemma 4.1. *Let Assumptions 2.1, 2.3, 2.2 and 3.1 be fulfilled. For any $\gamma \in [0, 1)$ and $\gamma_1, \gamma_2 \in [0, 1)$ such that $0 \leq \gamma_1+\gamma_2 < 1$, there exist constants C_γ, C_{γ_1} and C_{γ_1,γ_2} such that*

$$\|(-A_h^*)^{\gamma} D\mu^h(t, \psi)\| \leq C_\gamma t^{-\gamma}, \quad \|(-A_h^*)^{\gamma_2} D^2\mu^h(t, \psi)(-A_h)^{\gamma_1}\|_{\mathcal{L}(\mathcal{H})} \leq C_{\gamma_1,\gamma_2}\left(1 + t^{-\gamma_1+\gamma_2}\right), \tag{47}$$

$$\|(-A_h)^{\gamma_1} D\mu^h(t, \psi)\| \leq C_{\gamma_1} t^{-\gamma_1}, \quad \|(-A_h)^{\gamma_2} D^2\mu^h(t, \psi)(-A_h)^{\gamma_1}\|_{\mathcal{L}(\mathcal{H})} \leq C_{\gamma_1,\gamma_2}\left(1 + t^{-\gamma_1+\gamma_2}\right), \tag{48}$$

for any $t, t_1, t_2 \in [0, T]$, where $\psi \in V_h$ and $\mu^h(t, \psi)$ is given by (33).

Proof. Let us start with the first estimate of (48). The mild solution of (38) is given by

$$\xi_1^h(t) = S_h(t)\phi_1 + \int_0^t S_h(t-s)P_hF'(X^h(s, \psi))\xi_1^h(s)ds. \tag{49}$$

Taking the norm in both sides of (49), inserting an appropriate power of $-A_h$ yields

$$\begin{aligned}
&\|\xi_1^h(t)\| \\
&\leq \|S_h(t)(-A_h)^{\gamma}(-A_h)^{-\gamma}\phi_1\| + \int_0^t \|S_h(t-s)(-A_h)^{\frac{\delta}{2}}(-A_h)^{-\frac{\delta}{2}}P_hF'(X^h(s, \psi))\xi_1^h(s)\|ds \\
&\leq \|S_h(t)(-A_h)^{\gamma}(-A_h)^{-\gamma}\phi_1\| + \int_0^t \|(-A_h)^{\frac{\delta}{2}}S_h(t-s)(-A_h)^{-\frac{\delta}{2}}P_hF'(X^h(s, \psi))\xi_1^h(s)\|ds \\
&\leq \|S_h(t)(-A_h)^{\gamma}(-A_h)^{-\gamma}\phi_1\| + \int_0^t \|(-A_h)^{\frac{\delta}{2}}S_h(t-s)\|_{\mathcal{L}(\mathcal{H})}\|(-A_h)^{-\frac{\delta}{2}}P_hF'(X^h(s, \psi))\xi_1^h(s)\|ds.
\end{aligned}$$

Using the smoothing properties of the semigroup, [29, (3.12)] and Assumption 2.3 yields

$$\begin{aligned}
\|\xi_1^h(t)\| &\leq \|(-A_h)^{\gamma}S_h(t)\|_{\mathcal{L}(\mathcal{H})}\|(-A_h)^{-\gamma}\phi_1\| + C\int_0^t (t-s)^{-\frac{\delta}{2}}\|(-A_h)^{-\frac{\delta}{2}}P_hF'(X^h(s, \psi))\xi_1^h(s)\|ds \\
&\leq Ct^{-\gamma}\|(-A_h)^{-\gamma}\phi_1\| + C\int_0^t (t-s)^{-\frac{\delta}{2}}\|(-A)^{-\frac{\delta}{2}}F'(X^h(s, \psi))\xi_1^h(s)\|ds \\
&\leq Ct^{-\gamma}\|(-A_h)^{-\gamma}\phi_1\| + C\int_0^t (t-s)^{-\frac{\delta}{2}}\|\xi_1^h(s)\|ds.
\end{aligned}$$

Applying the generalized Gronwall inequality (see e.g., [33, Lemma 6.3] or [28, Lemma 7.1.1]) to the preceding inequality yields

$$\|\xi_1^h(t)\| \leq Ct^{-\gamma}\|(-A_h)^{-\gamma}\phi_1\|.$$

Using (33), Assumption 3.1 and Proposition 2.1 , it follows from the preceding inequality that

$$\begin{aligned}
|D\mu^h(t, \psi).\phi_1| &\leq \mathbb{E}\left(\|\varphi'(X^h(t, \psi))\|.\|\xi_1^h(t)\|\right) \leq C\mathbb{E}\|X^h(t)\|^{m-1}t^{-\gamma}\|(-A_h)^{-\gamma}\phi_1\| \\
&\leq Ct^{-\gamma}\|(-A_h)^{-\gamma}\phi_1\|.
\end{aligned}$$

Using Cauchy–Schwarz inequality and the preceding inequality, it follows that

$$\left|\langle(-A^*)^{\gamma}D\mu^h(t, \psi), \phi_1\rangle\right| = \left|\langle D\mu^h(t, \psi), (-A)^{\gamma}\phi_1\rangle\right| \leq Ct^{-\gamma}\|(-A_h)^{-\gamma}(-A_h)^{\gamma}\phi_1\| = Ct^{-\gamma}\|\phi_1\|. \tag{50}$$

Since we can identify $D\mu^h(t, \psi)$ with an element in $\mathcal{H}$, it follows from (50) that

$$\|(-A_h^*)^{\gamma}D\mu^h(t, \psi)\| \leq Ct^{-\gamma}, \quad \psi \in \mathcal{H}.$$

This completes the proof of the first estimate of (48). Using Lemma 3.1 yields

$$\|(-A_h)^{\gamma}D\mu^h(t, \psi)\| \leq \|(-A_h)^{\gamma}(-A_h^*)^{-\gamma}\|_{\mathcal{L}(\mathcal{H})}\|(-A_h^*)^{\gamma}D\mu^h(t, \psi)\| \leq Ct^{-\gamma}, \ \ \psi \in \mathcal{H}.$$

This proves the second estimate of (48). It remains to prove (47). Note that the mild solution of the process satisfying (40) is given by

$$\eta_{1,2}^h(t) = \int_0^t S_h(t-s)\left[P_hF'(X^h(s, \psi)).\eta_{1,2}^h(s) + P_hF''(X^h(s, \psi))(\xi_1^h(s), \xi_2^h(s))\right]ds. \tag{51}$$

Taking the norm in both sides of (51), inserting an appropriate power of A_h, using the smoothing properties of the semigroup and [29, (3.12)] (or [17, (70)]) yields

$$\begin{aligned}
\|\eta^h_{1,2}(t)\| &\leq \int_0^t \|S_h(t-s)(-A_h)^{\frac{\delta}{2}}(-A_h)^{-\frac{\delta}{2}}P_hF'(X^h(s,\psi))\eta^h_{1,2}(s)\|ds \\
&+ \int_0^t \|S_h(t-s)(-A_h)^{\eta}(-A_h)^{-\eta}P_hF''(X^h(s,\psi))(\xi^h_1(s),\xi^h_2(s))\|ds \\
&\leq \int_0^t \|(-A_h)^{\frac{\delta}{2}}S_h(t-s)\|_{\mathcal{L}(\mathcal{H})}\|(-A_h)^{-\frac{\delta}{2}}P_hF'(X^h(s,\psi))\eta^h_{1,2}(s)\|ds \\
&+ \int_0^t \|(-A_h)^{\eta}S_h(t-s)\|\mathcal{L}(\mathcal{H})(-A_h)^{-\eta}P_hF''(X^h(s,\psi))(\xi^h_1(s),\xi^h_2(s))\parallel ds \\
&\leq C\int_0^t (t-s)^{-\frac{\delta}{2}}\|(-A)^{-\frac{\delta}{2}}F'(X^h(s,\psi))\eta^h_{1,2}(s)\|ds \\
&+ C\int_0^t (t-s)^{-\eta}\|(-A)^{-\eta}F''(X^h(s,\psi))(\xi^h_1(s),\xi^h_2(s))\|ds.
\end{aligned}$$

Using (48) and Assumption 2.3, it follows from the above inequality that

$$\begin{aligned}
\|\eta^h_{1,2}(t)\| &\leq C\int_0^t (t-s)^{-\frac{\delta}{2}}\|\eta^h_{1,2}(s)\|ds + C\int_0^t (t-s)^{-\eta}\|\xi^h_1(s)\|\|\xi^h_2(s)\|ds \\
&\leq C\int_0^t (t-s)^{-\frac{\delta}{2}}\|\eta^h_{1,2}(s)\|ds + C\int_0^t (t-s)^{-\eta}s^{-\gamma_1-\gamma_2}\|(-A_h)^{-\gamma_1}\phi_1\|\|(-A_h)^{\gamma_2}\phi\|ds \\
&\leq Ct^{1-\gamma_1-\gamma_2-\eta}\|(-A_h)^{-\gamma_1}\phi_1\|\|(-A_h)^{\gamma_2}\phi\| + C\int_0^t (t-s)^{-\frac{\delta}{2}}\|\eta^h_{1,2}(s)\|ds, \qquad (52)
\end{aligned}$$

where at the last step we used the estimate

$$\begin{aligned}
\int_0^t (t-s)^{-\eta}s^{-\gamma_1-\gamma_2}ds &= \int_0^{\frac{t}{2}} (t-s)^{-\eta}s^{-\gamma_1-\gamma_2}ds + \int_{\frac{t}{2}}^t (t-s)^{-\eta}s^{-\gamma_1-\gamma_2}ds \\
&\leq Ct^{-\eta}\int_0^{\frac{t}{2}} s^{-\gamma_1-\gamma_2}ds + Ct^{-\gamma_1-\gamma_2}\int_0^t (t-s)^{-\eta}ds \leq Ct^{1-\gamma_1-\gamma_2-\eta}.
\end{aligned}$$

Applying the generalized Gronwall inequality ([33, Lemma 6.3] or [28, Lemma 7.1.1]) to (52) yields

$$\|\eta^h_{1,2}(t)\| \leq Ct^{1-\gamma_1-\gamma_2-\eta}\|(-A_h)^{-\gamma_1}\phi_1\|\|(-A_h)^{\gamma_2}\phi\|. \qquad (53)$$

From (39), it follows by using (50), (53), Assumption 3.1 and Proposition 2.1, it follows that

$$\begin{aligned}
\left|\left\langle D^2\mu^h(t,\psi)\phi_1,\phi_2\right\rangle\right| &\leq \mathbb{E}\left[\|\varphi''(X^h(t,\psi))\|\|\xi^h_1(t)\|\|\xi^h_2(t)\|\right] + \mathbb{E}\left[\|\varphi'(X^h(t,\psi))\|\|\eta_{1,2}(t)\|\right] \\
&\leq \mathbb{E}\|X^h(t,\psi)\|^{m-2}\|\xi^h_1(t)\|\|\xi^h_2(t)\| + \mathbb{E}\|X^h(t,\psi)\|^{m-1}\|\eta_{1,2}(t)\| \\
&\leq C\left(t^{-\gamma_1-\gamma_2} + t^{1-\gamma_1-\gamma_2-\eta}\right)\|(-A_h)^{-\gamma_1}\phi_1\|\|(-A_h)^{\gamma_2}\phi\| \\
&\leq Ct^{-\gamma_1-\gamma_2}\|(-A_h)^{-\gamma_1}\phi_1\|\|(-A_h)^{\gamma_2}\phi\|,
\end{aligned}$$

where at the last step we used the fact $1-\eta > 0$. Using the preceding inequality, it follows that

$$\begin{aligned}
\left|\left\langle (-A^*_h)^{\gamma_2}D^2\mu^h(t,\psi)(-A_h)^{\gamma_1}\phi_1,\phi_2\right\rangle\right| &= \left|\left\langle D^2\mu^h(t,\psi)(-A_h)^{\gamma_1}\phi_1,(-A_h)^{\gamma_2}\phi_2\right\rangle\right| \\
&\leq Ct^{-\gamma_1-\gamma_2}\|\phi_1\|\|\phi_2\|. \qquad (54)
\end{aligned}$$

Taking the supremum over $\phi_1,\phi_2 \in \mathcal{H}$ such that $\|\phi_1\|,\ \|\phi_2\| \leq 1$, it follows from (54) that

$$\|(-A^*_h)^{\gamma_2}D^2\mu^h(t,\psi)(-A_h)^{\gamma_1}\|_{\mathcal{L}(\mathcal{H})} \leq Ct^{-\gamma_1-\gamma_2}.$$

Using Lemma 3.1, it follows from the above inequality that

$$\begin{aligned}
\|(-A_h)^{\gamma_2}D^2\mu^h(t,\psi)(-A_h)^{\gamma_1}\|_{\mathcal{L}(\mathcal{H})} &\leq \|(-A_h)^{\gamma_2}(-A^*_h)^{-\gamma_2}\|_{\mathcal{L}(\mathcal{H})}\|(-A^*_h)^{\gamma_2}D^2\mu^h(t,\psi)(-A_h)^{\gamma_1}\|_{\mathcal{L}(\mathcal{H})} \\
&\leq Ct^{-\gamma_1-\gamma_2}.
\end{aligned}$$

The proof of the lemma is thus completed. ■

The following lemma will be useful in the rest of this paper. Its proof in the case of linear self-adjoint operator A and nonlinear function $F \in C^2_b(\mathcal{H},\mathcal{H})$ can be found in [11, lemma 3.1]. Here, we are working under weaker assumptions on F, namely Assumption 2.3 and with not neccesarily self-adjoint operator A_h.

Lemma 4.2. *Let Assumptions 2.1, 2.2 and 2.3 be fulfilled. Then the Malliavin derivative of $X^h(t)$ satisfies the following regularity estimate*

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

Proof. The proof goes along the sames lines as that of [11, Lemma 3.1]. Since here we are working under weaker assumptions on the nonlinearity F, let us provide details of the proof. For $u\in U_0$, differentiating both sides of (23) and employing (6) yields

$$\mathcal{D}_s^uX^h(t)=S_h(t-s)P_hu+\int_s^t S_h(t-r)P_hDF(X^h(r)).\mathcal{D}_s^uX^h(r)dr. \tag{55}$$

Taking the norm in both sides of (55), inserting an appropriate power of A_h, using the smoothing properties of the semigroup $S_h(t)$, [17, (70)] (or [10, (2.4)]) and Assumption 2.3 yields

$$\begin{aligned}
\mathbb{E}\|\mathcal{D}_s^uX^h(t)\|^2 &\leq \left\|S_h(t-s)(-A_h)^{\frac{1-\beta}{2}}(-A_h)^{\frac{\beta-1}{2}}P_hu\right\|^2\\
&+\int_s^t\mathbb{E}\|S_h(t-r)(-A_h)^{\frac{\delta}{2}}(-A_h)^{-\frac{\delta}{2}}P_hDF(X^h(r)).\mathcal{D}_s^uX^h(r)\|^2dr\\
&\leq \left\|(-A_h)^{\frac{1-\beta}{2}}S_h(t-s)\right\|^2_{\mathcal{L}(H)}\left\|(-A_h)^{\frac{\beta-1}{2}}P_hu\right\|^2\\
&+\int_s^t\mathbb{E}\left[\|(-A_h)^{\frac{\delta}{2}}S_h(t-r)\|^2_{\mathcal{L}(H)}\|(-A_h)^{-\frac{\delta}{2}}P_hF'(X^h(r))\mathcal{D}_s^uX^h(r)\|^2\right]dr\\
&\leq C(t-s)^{\beta-1}\|(-A)^{\frac{\beta-1}{2}}u\|^2+C\int_s^t(t-r)^{-\delta}\mathbb{E}\|(-A)^{-\frac{\delta}{2}}F'(X^h(r))\mathcal{D}_s^uX^h(r)\|^2dr\\
&\leq C(t-s)^{\beta-1}\|(-A)^{\frac{\beta-1}{2}}u\|^2+C\int_s^t(t-r)^{-\delta}\mathbb{E}\|\mathcal{D}_s^uX^h(r)\|^2dr.
\end{aligned}\tag{56}$$

Applying the generalized Gronwall inequality ([33, Lemma 6.3] or [28, Lemma 7.1.1]) to (56) yields

$$\mathbb{E}\|\mathcal{D}_s^uX^h(t)\|^2\leq C(t-s)^{\beta-1}\|(-A)^{\frac{\beta-1}{2}}u\|^2. \tag{57}$$

From (55), we obtain

$$(-A_h)^{\frac{\beta-1}{2}}\mathcal{D}_s^uX^h(t)=(-A_h)^{\frac{\beta-1}{2}}S_h(t-s)P_hu+\int_s^t(-A_h)^{\frac{\beta-1}{2}}S_h(t-r)P_hDF(X^h(r)).\mathcal{D}_s^uX^h(r)dr. \tag{58}$$

Taking the norm in both sides of (58), using elementary inequalities, the boundedness of $(-A_h)^{\frac{\beta-1}{2}}$, the smoothing properties of $S_h(t)$, Assumption 2.3, taking the expectation in both sides, using [17, (70)] (or [10, (2.4)]) and (57) yields

$$\begin{aligned}
&\mathbb{E}\left\|(-A_h)^{\frac{\beta-1}{2}}\mathcal{D}_s^uX^h(t)\right\|^2\\
&\leq 2\|S_h(t-s)\|^2_{\mathcal{L}(\mathcal{H})}\left\|(-A_h)^{\frac{\beta-1}{2}}P_hu\right\|^2\\
&+2\int_s^t\|(-A_h)^{\frac{\beta-1}{2}}\|^2_{\mathcal{L}(\mathcal{H})}\mathbb{E}\|S_h(t-r)(-A_h)^{\frac{\delta}{2}}(-A_h)^{-\frac{\delta}{2}}P_hDF(X^h(r))\mathcal{D}_s^uX^h(r)\|^2dr\\
&\leq C\|(-A)^{\frac{\beta-1}{2}}u\|^2+\int_s^t\|(-A_h)^{\frac{\delta}{2}}S_h(t-r)\|^2_{\mathcal{L}(\mathcal{H})}\mathbb{E}\|(-A_h)^{-\frac{\delta}{2}}P_hF'(X^h(r))\mathcal{D}_s^uX^h(r)\|^2dr\\
&\leq C\|(-A)^{\frac{\beta-1}{2}}u\|^2+\int_s^t(t-r)^{-\delta}\mathbb{E}\|(-A)^{-\frac{\delta}{2}}F'(X^h(r))\mathcal{D}_s^uX^h(r)\|^2dr\\
&\leq C\|(-A)^{\frac{\beta-1}{2}}u\|^2+\int_s^t(t-r)^{-\delta}\mathbb{E}\|\mathcal{D}_s^uX^h(r)\|^2dr\leq C\|(-A)^{\frac{\beta-1}{2}}u\|^2\\
&+\int_s^t(t-r)^{-\delta}(r-s)^{\beta-1}\|(-A)^{\frac{\beta-1}{2}}u\|^2dr\\
&\leq C\|(-A)^{\frac{\beta-1}{2}}u\|^2+C(t-s)^{\beta-\delta}\|(-A)^{\frac{\beta-1}{2}}u\|^2\leq C\|(-A)^{\frac{\beta-1}{2}}u\|^2,
\end{aligned}\tag{59}$$

where at the last step we used the estimate

$$\begin{aligned}\int_s^t (t-r)^{-\delta}(r-s)^{\beta-1}dr &= \int_s^{\frac{t+s}{2}} (t-r)^{-\delta}(r-s)^{\beta-1}dr + \int_{\frac{t+s}{2}}^t (t-r)^{-\delta}(r-s)^{\beta-1}dr\\ &\leq C(t-s)^{-\delta}\int_s^{\frac{t+s}{2}} (r-s)^{\beta-1}dr + C(t-s)^{\beta-1}\int_{\frac{t+s}{2}}^t (r-s)^{-\delta}dr\\ &\leq C(t-s)^{\beta-\delta} \leq C.\end{aligned}$$

Note that the estimate in (59) is uniform with respect to $u \in U_0$. Let $(u_i)_{i\in\mathbb{N}}$ be an orthonormal basis of U_0. Using the definition of $\mathcal{L}_2^0$ and (59) yields

$$\mathbb{E}\|(-A_h)^{\frac{\beta-1}{2}}\mathcal{D}_s^u X^h(t)\|^2_{\mathcal{L}_2^0} = \mathbb{E}\sum_{i\in\mathbb{N}}\mathbb{E}\|(-A_h)^{\frac{\beta-1}{2}}\mathcal{D}_s^{u_i}X^h(t)\|^2 \leq C\sum_{i\in\mathbb{N}}\|(-A)^{\frac{\beta-1}{2}}u_i\|^2 = C\|(-A)^{\frac{\beta-1}{2}}u\|^2_{\mathcal{L}_2^0} < \infty.$$

This completes the proof of the lemma. ■

Lemma 4.3 ([6, Lemma 11]). *Under Assumption 2.2, the following holds*

$$\left\|(-A_h)^{\frac{\beta-1}{2}}P_hQ^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})} \leq C.$$

With the above preparation we are now ready to prove the main result of this section.

4.2. Proof of Theorem 4.1

Using Proposition 3.2 yields

$$\begin{aligned}|\mathbb{E}[\varphi(X^h(T))-\varphi(X(T))]| \;&\leq\; |\mathbb{E}[\mu(T,X^h(0))-\mu(T,X(0))]|\\ &+\left|\mathbb{E}\int_0^T\left\langle A_hX^h(s), D\mu(T-s,X^h(s))\right\rangle ds\right.\\ &+\left.\int_0^T\mathbb{E}\left\langle(-A)^{\frac{1}{2}}X^h(s),(-A^*)^{\frac{1}{2}}D\mu(T-s,X^h(s))\right\rangle ds\right|\\ &+\left|\mathbb{E}\int_0^T\left\langle P_hF\left(X^h(s)\right)-F(X^h(s)),D\mu(T-s,X^h(s))\right\rangle ds\right|\\ &+\left|\frac{1}{2}\mathbb{E}\int_0^T Tr\left[D^2\mu(T-s,X^h(s))\,(P_hQP_h-Q)\right]ds\right|\\ &=: |I_1|+|I_2|+|I_3|+|I_4|.\end{aligned} \tag{60}$$

In the following subsections, we estimate $I_1,\ldots,I_4$ separately.

4.2.1. Estimate of I_1

Bearing in mind that $X_0^h = P_hX_0$, $X(0)=X_0$ and using the chain rule it follows that

$$\begin{aligned}I_1 &= \mathbb{E}\left[\int_0^1\langle D\mu(T,X_0+t(P_hX_0-X_0)),P_hX_0-X_0\rangle\,dt\right]\\ &= \mathbb{E}\left[\int_0^1\left\langle(-A)^{-1+\frac{\epsilon}{2}}(-A)^{1-\frac{\epsilon}{2}}D\mu(T,X_0+t(P_hX_0-X_0)),P_hX_0-X_0\right\rangle dt\right]\\ &= \mathbb{E}\left[\int_0^1\left\langle(-A)^{1-\frac{\epsilon}{2}}D\mu(T,X_0+t(P_hX_0-X_0)),\left((-A)^{-1+\frac{\epsilon}{2}}\right)^*(P_h-\mathbf{I})X_0\right\rangle dt\right].\end{aligned}$$

Using Lemma 4.1, Proposition 2.1(v) and (30) yields

$$\begin{aligned}|I_1| &\leq \mathbb{E}\left[\int_0^1\left\|(-A)^{1-\frac{\epsilon}{2}}D\mu(T,X_0+t(P_hX_0-X_0))\right\|\left\|\left((-A)^{-1+\frac{\epsilon}{2}}\right)^*(P_h-\mathbf{I})X_0\right\|dt\right]\\ &\leq CT^{-1+\frac{\epsilon}{2}}\int_0^1\left\|\left((-A)^{-1+\frac{\epsilon}{2}}\right)^*(P_h-\mathbf{I})\right\|_{\mathcal{L}(\mathcal{H})}dt\\ &\leq CT^{-1+\frac{\epsilon}{2}}\int_0^1\left\|\left((P_h-\mathbf{I})(-A)^{-1+\frac{\epsilon}{2}}\right)^*\right\|_{\mathcal{L}(\mathcal{H})}dt \leq Ch^{2-\epsilon}.\end{aligned} \tag{61}$$

4.2.2. Estimate of I_2

Let us recall that $A = A_1 + A_2$, where A_1 is the self-adjoint part and A_2 the non self-adjoint part. Hence, the expression of I_2 can be decomposed as follows

$$
\begin{aligned}
I_2 &= \mathbb{E}\left[\int_0^T \left\langle (A_{1,h} - A_1)X^h(t), D\mu(T-t, X^h(t))\right\rangle dt\right] \\
&\quad + \mathbb{E}\left[\int_0^T \left\langle (A_{2,h} - A_2)X^h(t), D\mu(T-t, X^h(t))\right\rangle dt\right] =: I_{21} + I_{22},
\end{aligned} \tag{62}
$$

where we also used (26). Note that from (19), (27) and (20), one can easily check that $R_1^h = A_{1,h}^{-1}P_hA_1$. Using the later relation, the fact that $P_hA_1X^h(t) \in V_h$, $X^h(t) \in V_h$ and $P_hD\mu(T-t, X^h(t)) \in V_h$, it follows by using the definition of P_h (19) that

$$
\begin{aligned}
\left\langle (A_{1,h} - A_1)X^h(t), D\mu(T-t, X^h(t))\right\rangle &= \left\langle (P_hA_{1,h} - A_1P_h)X^h(t), D\mu(T-t, X^h(t))\right\rangle \\
&= \left\langle X^h(t), (A_{1,h}P_h - P_hA_1)D\mu(T-t, X^h(t))\right\rangle \\
&= \left\langle X^h(t), A_{1,h}P_h\left(\mathbf{I} - A_{1,h}^{-1}P_hA_1\right)D\mu(T-t, X^h(t))\right\rangle \\
&= \left\langle X^h(t), A_{1,h}P_h\left(\mathbf{I} - R_1^h\right)D\mu(T-t, X^h(t))\right\rangle.
\end{aligned} \tag{63}
$$

Using the mild form of $X^h(t)$ (23) yields the following decomposition

$$
\begin{aligned}
I_{21} &= \mathbb{E}\left[\int_0^T \left\langle S_h(t)P_hX_0, A_{1,h}P_h\left(\mathbf{I} - R_1^h\right)D\mu(T-t, X^h(t))\right\rangle dt\right] \\
&\quad + \mathbb{E}\left[\int_0^T \left\langle \int_0^t S_h(t-s)P_hF(X^h(s))ds, A_{1,h}P_h\left(\mathbf{I} - R_1^h\right)D\mu(T-t, X^h(t))\right\rangle dt\right] \\
&\quad + \mathbb{E}\left[\int_0^T \left\langle \int_0^t S_h(t-s)P_hdW(s), A_{1,h}P_h\left(\mathbf{I} - R_1^h\right)D\mu(T-t, X^h(t))\right\rangle dt\right] \\
&=: I_{21}^{(1)} + I_{21}^{(2)} + I_{21}^{(3)}.
\end{aligned} \tag{64}
$$

Let us start with the estimate of $I_{21}^{(1)}$. Using the equivalence of norms [29] $\|(-A_1)^\alpha v\| \approx \|(-A)^\alpha v\|$, $v \in \mathcal{D}((-A)^\alpha)$, $\alpha \in [0,1]$, $\|(-A_{1,h})^\alpha v\| \approx \|(-A_h)^\alpha v\|$, $v \in \mathcal{D}((-A)^\alpha) \cap V_h$, $\alpha \in \left[0, \frac{1}{2}\right]$, [6, Lemma 1], Lemma 4.1 and (32), it follows that

$$
\begin{aligned}
&|I_{21}^{(1)}| \\
&= \left|\mathbb{E}\left[\int_0^T \left\langle A_{1,h}S_h(t)P_hX_0, P_h\left(\mathbf{I} - R_1^h\right)(-A_1)^{-1+\frac{\epsilon}{2}}(-A_1)^{1-\frac{\epsilon}{2}}D\mu(T-t, X^h(t))\right\rangle dt\right]\right| \\
&= \left|\mathbb{E}\left[\int_0^T \left\langle (-A_{1,h})^{1-\frac{\epsilon}{2}}S_h(t)P_hX_0, (-A_{1,h})^{\frac{\epsilon}{2}}P_h\left(\mathbf{I} - R_1^h\right)(-A_1)^{-1+\frac{\epsilon}{2}}(-A_1)^{1-\frac{\epsilon}{2}}D\mu(T-t, X^h(t))\right\rangle dt\right]\right| \\
&\leq C\mathbb{E}\int_0^T \|(-A_{1,h})^{1-\frac{\epsilon}{2}}S_h(t)P_hX_0\| \left\|(-A_{1,h})^{\epsilon}P_h\left(\mathbf{I} - R_1^h\right)(-A_1)^{-1+\frac{\epsilon}{2}}(-A_1)^{1-\frac{\epsilon}{2}}D\mu(T-t, X^h(t))\right\| dt \\
&\leq C\mathbb{E}\int_0^T \|(-A_h)^{1-\frac{\epsilon}{2}}S_h(t)P_hX_0\| \left\|(-A_1)^{\frac{\epsilon}{2}}\left(\mathbf{I} - R_1^h\right)(-A_1)^{-1+\frac{\epsilon}{2}}(-A_1)^{1-\frac{\epsilon}{2}}D\mu(T-t, X^h(t))\right\| dt \\
&\leq C\mathbb{E}\int_0^T t^{-1+\frac{\epsilon}{2}}\|(-A_1)^{\frac{\epsilon}{2}}\left(\mathbf{I} - R_1^h\right)(-A_1)^{-1+\epsilon}\|_{\mathcal{L}(\mathcal{H})}\left\|(-A_1)^{1-\frac{\epsilon}{2}}D\mu(T-t, X^h(t))\right\| dt \\
&\leq C\mathbb{E}\int_0^T t^{-1+\frac{\epsilon}{2}}\|(-A_1)^{\frac{\epsilon}{2}}\left(\mathbf{I} - R_1^h\right)(-A_1)^{-1+\frac{\epsilon}{2}}\|_{\mathcal{L}(\mathcal{H})}\left\|(-A)^{1-\epsilon}D\mu(T-t, X^h(t))\right\| dt \\
&\leq Ch^{2-\epsilon}\int_0^T t^{-1+\frac{\epsilon}{2}}(T-t)^{-1+\frac{\epsilon}{2}}dt \leq Ch^{2-\epsilon}.
\end{aligned} \tag{65}
$$

For the estimate of $I_{21}^{(2)}$, using Assumption 2.3 the equivalence of norms [29] $\|(-A_1)^\alpha v\| \approx \|(-A)^\alpha v\|$, $v \in \mathcal{D}((-A)^\alpha)$, $\alpha \in [0,1]$, $\|(-A_{1,h})^\alpha v\| \approx \|(-A_h)^\alpha v\|$, $v \in \mathcal{D}((-A)^\alpha) \cap V_h$, $\alpha \in \left[0, \frac{1}{2}\right]$, [6, Lemma 1], Lemma 4.1 and (32). This yields

$$
\begin{aligned}
|I_{21}^{(2)}| &= \left|\mathbb{E}\left[\int_0^T \left\langle \int_0^t A_{1,h}S_h(t-s)P_hF(X^h(s))ds, P_h(\mathbf{I} - R_1^h)D\mu(T-t, X^h(t))\right\rangle dt\right]\right| \\
&= \Bigg|\mathbb{E}\Bigg[\int_0^T \Bigg\langle \int_0^t (-A_{1,h})^{1-\frac{\epsilon}{2}}S_h(t-s)P_hF(X^h(s))ds, \\
&\qquad (-A_{1,h})^{\frac{\epsilon}{2}}P_h(\mathbf{I} - R_1^h)(-A_1)^{-1+\frac{\epsilon}{2}}(-A_1)^{1-\frac{\epsilon}{2}}D\mu(T-t, X^h(t))\Bigg\rangle dt\Bigg]\Bigg|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \mathbb{E}\int_0^T\int_0^t \|(-A_{1,h})^{1-\frac{\epsilon}{2}}S_h(t-s)P_hF(X^h(s))\|(-A_{1,h})^{\frac{\epsilon}{2}}P_h(\mathbf{I}-R_1^h)(-A_1)^{-1+\frac{\epsilon}{2}}\\
&\quad(-A_1)^{1-\frac{\epsilon}{2}}D\mu(T-t,X^h(t))\| dsdt\\
&\leq \mathbb{E}\int_0^T\int_0^t \|(-A_h)^{1-\frac{\epsilon}{2}}S_h(t-s)P_hF(X^h(s))\|\|(-A_1)^{\frac{\epsilon}{2}}(\mathbf{I}-R_1^h)(-A_1)^{-1+\frac{\epsilon}{2}}\|_{\mathcal{L}(\mathcal{H})}\\
&\quad\times\|(-A_1)^{1-\frac{\epsilon}{2}}D\mu(T-t,X^h(t))\|dsdt\\
&\leq Ch^{2-\epsilon}\int_0^T\int_0^t (t-s)^{-1+\frac{\epsilon}{2}}(T-t)^{-1+\frac{\epsilon}{2}}dsdt\leq Ch^{2-\epsilon}.
\end{aligned}
\tag{66}
$$

Let us now estimate $I_{21}^{(3)}$. Using the integration by parts formula in the Malliavin sense (7) and the chain rules yields

$$
\begin{aligned}
I_{21}^{(3)} &= \mathbb{E}\left[\int_0^T\left\langle A_{1,h}\int_0^t S_h(t-s)P_hdW(s), P_h(\mathbf{I}-R_1^h)D\mu(T-t,X^h(t))\right\rangle dt\right]\\
&= \mathbb{E}\left[\int_0^T\int_0^t \left\langle A_{1,h}S_h(t-s)P_h, P_h(\mathbf{I}-R_1^h)D^2\mu(T-t,X^h(t))\mathcal{D}_s(X^h(t))\right\rangle_{\mathcal{L}_2^0}dsdt\right].
\end{aligned}
\tag{67}
$$

Inserting appropriate powers of A and $A_{1,h}$ in (67), using Cauchy's inequality, Lemmas 4.1, 3.1, 4.2, the equivalence of norms [29] $\|(-A_1)^\alpha v\|\approx\|(-A)^\alpha v\|$, $v\in\mathcal{D}((-A)^\alpha)$, $\alpha\in[0,1]$, $\|(-A_{1,h})^\alpha v\|\approx\|(-A_h)^\alpha v\|$, $v\in\mathcal{D}((-A)^\alpha)\cap V_h$, $\alpha\in\left[0,\frac{1}{2}\right]$, [6, Lemma 1] and (32), it follows that

$$
\begin{aligned}
|I_{21}^{(3)}| &= \Bigg|\mathbb{E}\Bigg[\int_0^T\int_0^t\Big\langle(-A_{1,h})^{\frac{1+\beta-\epsilon}{2}}S_h(t-s)(-A_h)^{\frac{1-\beta}{2}}(-A_h)^{\frac{\beta-1}{2}}P_h,\\
&\quad(-A_{1,h})^{\frac{1-\beta+\epsilon}{2}}P_h(\mathbf{I}-R_1^h)(-A_1)^{\frac{-\beta-1+\epsilon}{2}}(-A_1)^{\frac{\beta+1-\epsilon}{2}}D^2\mu(T-t,X^h(t))\mathcal{D}_s(X^h(t))\Big\rangle_{\mathcal{L}_2^0}\Bigg]dsdt\Bigg|\\
&\leq \mathbb{E}\int_0^T\int_0^t\left\|(-A_{1,h})^{\frac{1+\beta-\epsilon}{2}}S_h(t-s)(-A_h)^{\frac{1-\beta}{2}}(-A_h)^{\frac{\beta-1}{2}}P_h\right\|_{\mathcal{L}_2^0}\\
&\times\left\|(-A_{1,h})^{\frac{1-\beta+\epsilon}{2}}P_h(\mathbf{I}-R_1^h)(-A_1)^{\frac{-\beta-1+\epsilon}{2}}(-A_1)^{\frac{\beta+1-\epsilon}{2}}D^2\mu(T-t,X^h(t))\mathcal{D}_sX^h(t)\right\|_{\mathcal{L}_2^0}dsdt\\
&\leq \mathbb{E}\int_0^T\int_0^t\left\|(-A_{1,h})^{\frac{1+\beta-\epsilon}{2}}(-A_h)^{\frac{1-\beta}{2}}S_h(t-s)\right\|_{\mathcal{L}(\mathcal{H})}\left\|(-A_h)^{\frac{\beta-1}{2}}P_h\right\|_{\mathcal{L}_2^0}\\
&\quad\times\left\|(-A_1)^{\frac{1-\beta+\epsilon}{2}}(\mathbf{I}-R_1^h)(-A_1)^{\frac{-\beta-1+\epsilon}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\\
&\quad\times\left\|(-A_1)^{\frac{\beta+1-\epsilon}{2}}D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\left\|(-A)^{\frac{\beta-1}{2}}\mathcal{D}_sX^h(t)\right\|_{\mathcal{L}_2^0}dsdt\\
&\leq Ch^{2\beta-\epsilon}\int_0^T\int_0^t(t-s)^{-1+\epsilon}(T-t)^{-\epsilon}dsdt\leq Ch^{2\beta-\epsilon}.
\end{aligned}
\tag{68}
$$

Substituting (68), (66) and (65) in (64) yields

$$
|I_{21}|\leq Ch^{2\beta-2\epsilon}. \tag{69}
$$

Let us now estimate I_{22}. First of all, since $X^h(t)\in V_h$ and $A_{2,h}R_2^h=P_hA_2$, it holds that

$$
A_{2,h}X^h(t)-A_2X^h(t)=A_{2,h}R_2^hX^h(t)-A_2X^h(t)=(P_h-\mathbf{I})A_2X^h(t). \tag{70}
$$

Substituting (70) in the expression of I_2 in (60) and using the mild form of $X^h(t)$ (see (23)) yields the following decomposition

$$
\begin{aligned}
I_{22} &= \mathbb{E}\left[\int_0^T\left\langle(P_h-\mathbf{I})A_2X^h(t), D\mu(T-t,X^h(t))\right\rangle dt\right]\\
&= \mathbb{E}\left[\int_0^T\left\langle A_2X^h(t), (P_h-\mathbf{I})D\mu(T-t,X^h(t))\right\rangle dt\right]\\
&= \mathbb{E}\left[\int_0^T\left\langle A_2S_h(t)P_hX_0, (P_h-\mathbf{I})D\mu(T-t,X^h(t))\right\rangle dt\right]\\
&+ \mathbb{E}\left[\int_0^T\left\langle A_2\int_0^t S_h(t-s)P_hF(X^h(s))ds, (P_h-\mathbf{I})D\mu(T-t,X^h(t))\right\rangle dt\right]\\
&+ \mathbb{E}\left[\int_0^T\left\langle A_2\int_0^t S_h(t-s)P_hdW(s), (P_h-\mathbf{I})D\mu(T-t,X^h(t))\right\rangle dt\right]\\
&=: I_{22}^{(1)}+I_{22}^{(2)}+I_{22}^{(3)}.
\end{aligned}
\tag{71}
$$

Now let us start with the estimate of $I_{22}^{(1)}$. Using the estimate $\|A_2 v\| \leq C\|(-A)^{\frac{1}{2}}v\|$, $v \in \mathcal{D}((-A)^{\frac{1}{2}})$ (cf. [29]), the equivalence of norms $\|(-A_h)^{\frac{1}{2}}v\| \approx \|(-A)^{\frac{1}{2}}v\|$, $v \in V_h$, (30) and Lemma 4.1 yields

$$\begin{aligned}
|I_{22}^{(1)}| &= \left|\mathbb{E}\left[\int_0^T \left\langle A_2 S_h(t)P_hX_0, (P_h-\mathbf{I})(-A)^{-\beta+\frac{\epsilon}{2}}(-A)^{\beta-\frac{\epsilon}{2}}D\mu(T-t,X^h(t))\right\rangle dt\right]\right| \\
&\leq \mathbb{E}\int_0^T \|A_2S_h(t)P_hX_0\| \left\|(P_h-\mathbf{I})(-A)^{-\beta+\frac{\epsilon}{2}}(-A)^{\beta-\frac{\epsilon}{2}}D\mu(T-t,X^h(t))\right\| dt \\
&\leq C\int_0^T \|(-A)^{\frac{1}{2}}S_h(t)P_hX_0\|\|(P_h-\mathbf{I})(-A)^{-\beta+\frac{\epsilon}{2}}\|_{\mathcal{L}(\mathcal{H})}(T-t)^{-\beta+\frac{\epsilon}{2}}dt \\
&\leq C\int_0^T \|(-A_h)^{\frac{1}{2}}S_h(t-s)P_hX_0\|\|(P_h-\mathbf{I})(-A)^{-\beta+\frac{\epsilon}{2}}\|_{\mathcal{L}(\mathcal{H})}(T-t)^{-\beta+\frac{\epsilon}{2}}dt \\
&\leq Ch^{2\beta-\epsilon}\int_0^T t^{-\frac{1}{2}}(T-t)^{-\beta+\frac{\epsilon}{2}}dt \leq Ch^{2\beta-\epsilon}.
\end{aligned} \tag{72}$$

Let us move to the estimate of I_{22}. Using the estimate $\|A_2 v\| \leq C\|(-A)^{\frac{1}{2}}v\|$, $v \in \mathcal{D}((-A)^{\frac{1}{2}})$ (cf. [29]), the equivalence of norms $\|(-A_h)^{\frac{1}{2}}v\| \approx \|(-A)^{\frac{1}{2}}v\|$, $v \in V_h$, (30), Lemma 4.1, Theorem 2.1 and Assumption 2.3 yields

$$\begin{aligned}
&|I_{22}^{(2)}| \\
&= \left|\mathbb{E}\left[\int_0^T \left\langle A_2\int_0^t S_h(t-s)P_hF(X^h(s))ds, (P_h-\mathbf{I})(-A)^{-\beta+\frac{\epsilon}{2}}(-A)^{\beta-\frac{\epsilon}{2}}D\mu(T-t,X^h(t))\right\rangle dt\right]\right| \\
&\leq \mathbb{E}\int_0^T \left\|\int_0^t A_2S_h(t-s)P_hF(X^h(s))ds\right\| \left\|(P_h-\mathbf{I})(-A)^{-\beta+\frac{\epsilon}{2}}(-A)^{\beta-\frac{\epsilon}{2}}D\mu(T-t,X^h(t))\right\| dt \\
&\leq \mathbb{E}\int_0^T \left(\int_0^t \|(-A)^{\frac{1}{2}}S_h(t-s)P_hF(X^h(s))\|ds\right)\|(P_h-\mathbf{I})(-A)^{-\beta+\frac{\epsilon}{2}}\|_{\mathcal{L}(\mathcal{H})} \\
&\quad\times \left\|(-A)^{\beta-\frac{\epsilon}{2}}D\mu(T-t,X^h(t))\right\| dt \\
&\leq Ch^{2\beta-\epsilon}\int_0^T \mathbb{E}\left(\int_0^t \|(-A_h)^{\frac{1}{2}}S_h(t-s)\|_{\mathcal{L}(\mathcal{H})}\|X^h(s)\|ds\right)(T-t)^{-\beta-\frac{\epsilon}{2}}dt \\
&\leq Ch^{2\beta-2\epsilon}\int_0^T\int_0^t (t-s)^{-\frac{1}{2}}(T-t)^{-\beta-\frac{\epsilon}{2}}dsdt \leq Ch^{2\beta-\epsilon}.
\end{aligned} \tag{73}$$

Let us now estimate $I_{22}^{(3)}$. Using the integration by parts formula in the Malliavin sense (see (7)) and the chain rules yields

$$\begin{aligned}
I_{22}^{(3)} &= \mathbb{E}\left[\int_0^T \left\langle A_2\int_0^t S_h(t-s)P_hdW(s), (P_h-\mathbf{I})D\mu(T-t,X^h(t))\right\rangle dt\right] \\
&= \mathbb{E}\left[\int_0^T\int_0^t \left\langle A_2S_h(t-s)P_h, (P_h-\mathbf{I})D^2\mu(T-t,X^h(t))\mathcal{D}_s(X^h(t))\right\rangle_{\mathcal{L}_2^0} dsdt\right].
\end{aligned}$$

Note that using the equivalence of norms $\|(-A_h)^{\frac{1}{2}}v\| \approx \|(-A)^{\frac{1}{2}}v\|$, $v \in \mathcal{D}((-A)^{\frac{1}{2}})$ (cf. [29, (2.12)]) and the interpolation inequality yields

$$C\|(-A)^{\alpha}v\| \leq \|(-A_h)^{\alpha}v\| \leq C'\|(-A)^{\alpha}v\|, \quad v \in \mathcal{D}((-A)^{\alpha})\cap V_h, \quad \alpha \in \left[0, \frac{1}{2}\right]. \tag{74}$$

Inserting appropriate powers of A and A_h in the expression of $I_{22}^{(3)}$ and using Cauchy–Schwarz's inequality yields

$$\begin{aligned}
|I_{22}^{(3)}| &= \left|\mathbb{E}\left[\int_0^T\int_0^t \left\langle A_2S_h(t-s)P_h, (P_h-\mathbf{I})(-A)^{\frac{-1-\beta+\epsilon}{2}}\right.\right.\right. \\
&\qquad \left.\left.\left.(-A)^{\frac{\beta+1-\epsilon}{2}}D^2\mu(T-t,X^h(t))A^{\frac{1-\beta}{2}}(-A)^{\frac{\beta-1}{2}}\mathcal{D}_s(X^h(t))\right\rangle_{\mathcal{L}_2^0}\right]dsdt\right| \\
&\leq \mathbb{E}\int_0^T\int_0^t \left\|A_2S_h(t-s)(-A_h)^{\frac{1-\beta}{2}}(-A_h)^{\frac{\beta-1}{2}}P_h\right\|_{\mathcal{L}_2^0}\left\|(P_h-\mathbf{I})(-A)^{\frac{-1-\beta+\epsilon}{2}}\right. \\
&\qquad \left.(-A)^{\frac{\beta+1-\epsilon}{2}}D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}(-A)^{\frac{\beta-1}{2}}\mathcal{D}_sX^h(t)\right\|_{\mathcal{L}_2^0}dsdt \\
&\leq \mathbb{E}\int_0^T\int_0^t \left\|A_2(-A_h)^{\frac{1-\beta}{2}}S_h(t-s)\right\|_{\mathcal{L}(\mathcal{H})}\left\|(-A_h)^{\frac{\beta-1}{2}}P_h\right\|_{\mathcal{L}_2^0}
\end{aligned}$$

$$\times \left\|(P_h-\mathbf{I})(-A)^{\frac{-1-\beta+\epsilon}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\left\|(-A)^{\frac{\beta+1-\epsilon}{2}}D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathcal{H})}$$
$$\times \left\|(-A)^{\frac{\beta-1}{2}}\mathcal{D}_sX^h(t)\right\|_{\mathcal{L}_2^0}dsdt. \tag{75}$$

Using Lemmas 4.1, 3.1, (74) and (30) yields

$$\begin{aligned}
&\leq Ch^{1+\beta}\int_0^T\int_0^t\left\|(-A_h)^{1-\frac{\beta}{2}}S_h(t-s)\right\|_{\mathcal{L}(\mathcal{H})}(T-t)^{-\epsilon}\left\|(-A)^{\frac{\beta-1}{2}}\mathcal{D}_sX^h(t)\right\|_{\mathcal{L}_2^0}dsdt\\
&\leq Ch^{1+\beta}\int_0^T\int_0^t(t-s)^{-1+\frac{\beta}{2}}(T-t)^{-\frac{\epsilon}{2}}\left\|(-A)^{\frac{\beta-1}{2}}(-A_h)^{\frac{-\beta+1}{2}}(-A_h)^{\frac{\beta-1}{2}}\mathcal{D}_sX^h(t)\right\|_{\mathcal{L}_2^0}dsdt\\
&\leq Ch^{1+\beta}\int_0^T\int_0^t(t-s)^{-1+\frac{\beta}{2}}(T-t)^{-\frac{\epsilon}{2}}\left\|(-A_h)^{\frac{\beta-1}{2}}\mathcal{D}_sX^h(t)\right\|_{\mathcal{L}_2^0}dsdt\\
&\leq Ch^{1+\beta}\int_0^T\int_0^t\left\|(-A_h)^{1-\frac{\beta}{2}}S_h(t-s)\right\|_{\mathcal{L}(\mathcal{H})}(T-t)^{-\frac{\epsilon}{2}}dsdt\leq Ch^{2\beta-\epsilon}.
\end{aligned}\tag{76}$$

Substituting (76), (73) and (72) in (71) yields

$$|I_{22}|\leq Ch^{2\beta}. \tag{77}$$

Substituting (77) and (69) in (62) yields

$$I_2\leq Ch^{2\beta}. \tag{78}$$

4.2.3. Estimate of I_3 and I_4

Let us start by estimating I_3. Inserting powers of A in the expression of I_3 (see (60)) yields

$$\begin{aligned}
I_3&=\mathbb{E}\left[\int_0^T\left\langle(P_h-\mathbf{I})F(X^h(t)),D\mu(T-t,X^h(t))\right\rangle dt\right]\\
&=\mathbb{E}\left[\int_0^T\left\langle F(X^h(t)),(P_h-\mathbf{I})D\mu(T-t,X^h(t))\right\rangle dt\right]\\
&=\mathbb{E}\left[\int_0^T\left\langle F(X^h(t)),(P_h-\mathbf{I})(-A)^{-\beta+\frac{\epsilon}{2}}(-A)^{\beta-\frac{\epsilon}{2}}D\mu(T-t,X^h(t))\right\rangle dt\right].
\end{aligned}\tag{79}$$

Using (30), Lemma 4.1, Assumption 2.3 and Theorem 2.1 yields

$$\begin{aligned}
|I_3|&\leq\mathbb{E}\left[\int_0^T\|F(X^h(t))\|\,\|(P_h-\mathbf{I})(-A)^{-\beta+\frac{\epsilon}{2}}(-A)^{\beta-\frac{\epsilon}{2}}D\mu(T-t,X^h(t))\|dt\right]\\
&\leq C\mathbb{E}\left[\int_0^T\|X^h(t)\|\,\|(P_h-\mathbf{I})(-A)^{-\beta+\frac{\epsilon}{2}}\|_{\mathcal{L}(\mathcal{H})}\left\|(-A)^{\beta-\frac{\epsilon}{2}}D\mu(T-t,X^h(t))\right\|dt\right]\\
&\leq Ch^{2\beta-\epsilon}\int_0^T(T-t)^{-\beta+\frac{\epsilon}{2}}dt\leq Ch^{2\beta-\epsilon}.
\end{aligned}\tag{80}$$

Now let us move to the estimate of I_4. We decompose I_4 as follows

$$\begin{aligned}
I_4&=\frac{1}{2}\mathbb{E}\int_0^T Tr\left[D^2\mu(T-t,X^h(t))(P_hQP_h-QP_h+QP_h-Q)\right]dt\\
&=\frac{1}{2}\mathbb{E}\int_0^T Tr\left[D^2\mu(T-t,X^h(t))(P_h-\mathbf{I})QP_h\right]dt\\
&+\frac{1}{2}\mathbb{E}\int_0^T Tr\left[D^2\mu(T-t,X^h(t))Q(P_h-\mathbf{I})\right]dt=:I_{41}+I_{42}.
\end{aligned}\tag{81}$$

Using Proposition 2.1 and inserting an appropriate power of A yields

$$\begin{aligned}
|I_{41}|&=\frac{1}{2}\left|\mathbb{E}\int_0^T Tr\left[Q^{\frac{1}{2}}P_hD^2\mu(T-t,X^h(t))(P_h-\mathbf{I})Q^{\frac{1}{2}}\right]dt\right|\\
&\leq C\mathbb{E}\int_0^T\left\|Q^{\frac{1}{2}}P_hD^2\mu(T-t,X^h(t))(P_h-\mathbf{I})Q^{\frac{1}{2}}\right\|_{\mathcal{L}_1(\mathcal{H})}dt\\
&\leq C\mathbb{E}\int_0^T\left\|Q^{\frac{1}{2}}P_h\left((-A)^{\frac{\beta-1}{2}}\right)^*\left((-A)^{\frac{1-\beta}{2}}\right)^*D^2\mu(T-t,X^h(t))\left((-A)^{\frac{1+\beta-\epsilon}{2}}\right)^*\right.\\
&\quad\left.\left((-A)^{\frac{-\beta-1+\epsilon}{2}}\right)^*(P_h-\mathbf{I})\left((-A)^{\frac{1-\beta}{2}}\right)^*\left((-A)^{\frac{\beta-1}{2}}\right)^*Q^{\frac{1}{2}}\right\|_{\mathcal{L}_1(\mathcal{H})}dt
\end{aligned}$$

$$\begin{aligned}
&\leq C\mathbb{E}\int_0^T \left\|Q^{\frac{1}{2}}\left((-A)^{\frac{\beta-1}{2}}P_h\right)^*\left((-A)^{\frac{1-\beta}{2}}\right)^* D^2\mu(T-t,X^h(t))\left((-A)^{\frac{1+\beta-\epsilon}{2}}\right)^*\right\|_{\mathcal{L}_2(\mathcal{H})}\\
&\quad\times\left\|\left((-A)^{\frac{-\beta-1+\epsilon}{2}}\right)^*(P_h-\mathbf{I})((-A)^{\frac{1-\beta}{2}})^*((-A)^{\frac{\beta-1}{2}})^*Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}dt\\
&\leq C\mathbb{E}\int_0^T \left\|\left((-A)^{\frac{\beta-1}{2}}P_hQ^{\frac{1}{2}}\right)^*((-A)^{\frac{1-\beta}{2}})^*D^2\mu(T-t,X^h(t))\left((-A)^{\frac{1+\beta-\epsilon}{2}}\right)^*\right\|_{\mathcal{L}_2(\mathcal{H})}\\
&\quad\times\left\|\left((-A)^{\frac{-\beta-1+\epsilon}{2}}\right)^*(P_h-\mathbf{I})^*\left((-A)^{\frac{1-\beta}{2}}\right)^*\left((-A)^{\frac{\beta-1}{2}}\right)^*Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}dt\\
&=: C\mathbb{E}\int_0^T |I_{41}^{(1)}(t)|\times|I_{41}^{(2)}(t)|dt.
\end{aligned} \tag{82}$$

Using Proposition 2.1, Lemmas 4.1, 4.3 and 3.1 yields

$$\begin{aligned}
|I_{41}^{(1)}(t)| &= \left\|\left((-A)^{\frac{\beta-1}{2}}P_hQ^{\frac{1}{2}}\right)^*\left((-A)^{\frac{1-\beta}{2}}\right)^*D^2\mu(T-t,X^h(t))\left((-A)^{\frac{1+\beta-\epsilon}{2}}\right)^*\right\|_{\mathcal{L}_2(\mathcal{H})}\\
&= \left\|\left((-A)^{\frac{1+\beta-\epsilon}{2}}\right)^*D^2\mu(T-t,X^h(t))\left((-A)^{\frac{1-\beta}{2}}\right)^*\left((-A)^{\frac{\beta-1}{2}}P_hQ^{\frac{1}{2}}\right)^*\right\|_{\mathcal{L}_2(\mathcal{H})}\\
&\leq \left\|\left((-A)^{\frac{1+\beta}{2}}\right)^*D^2\mu(T-t,X^h(t))\left((-A)^{\frac{1-\beta}{2}}\right)^*\right\|_{\mathcal{L}(\mathcal{H})}\left\|\left((-A)^{\frac{\beta-1}{2}}P_hQ^{\frac{1}{2}}\right)^*\right\|_{\mathcal{L}_2(\mathcal{H})}\\
&\leq \left\|\left((-A)^{\frac{1+\beta-\epsilon}{2}}\right)^*(-A)^{\frac{-\beta-1}{2}}(-A)^{\frac{1+\beta}{2}}D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}(-A)^{\frac{\beta-1}{2}}\right.\\
&\quad\left.\left((-A)^{\frac{1-\beta}{2}}\right)^*\right\|_{\mathcal{L}(\mathcal{H})}\left\|(-A)^{\frac{\beta-1}{2}}P_hQ^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}\\
&\leq \left\|\left((-A)^{\frac{1+\beta-\epsilon}{2}}\right)^*(-A)^{\frac{-\beta-1+\epsilon}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\left\|(-A)^{\frac{1+\beta-\epsilon}{2}}D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\\
&\quad\times\left\|(-A)^{\frac{\beta-1}{2}}\left((-A)^{\frac{1-\beta}{2}}\right)^*\right\|_{\mathcal{L}(\mathcal{H})}\\
&\leq C(T-t)^{1-\frac{\epsilon}{2}}.
\end{aligned} \tag{83}$$

Using (30), Proposition 2.1, Lemmas 4.3 and 3.1 yields

$$\begin{aligned}
|I_{41}^{(2)}(t)| &= \left\|\left((-A)^{\frac{-\beta-1+\epsilon}{2}}\right)^*(P_h-\mathbf{I})^*\left((-A)^{\frac{1-\beta}{2}}\right)^*\left((-A)^{\frac{\beta-1}{2}}\right)^*Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}\\
&= \left\|\left(A^{\frac{1-\beta}{2}}(P_h-\mathbf{I})(-A)^{\frac{-1-\beta+\epsilon}{2}}\right)^*\left((-A)^{\frac{\beta-1}{2}}\right)^*Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}\\
&\leq \left\|\left((-A)^{\frac{1-\beta}{2}}(P_h-\mathbf{I})(-A)^{\frac{-1-\beta+\epsilon}{2}}\right)^*\right\|_{\mathcal{L}(\mathcal{H})}\left\|\left((-A)^{\frac{\beta-1}{2}}\right)^*Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}\\
&\leq \left\|A^{\frac{1-\beta}{2}}(P_h-\mathbf{I})(-A)^{\frac{-1-\beta+\epsilon}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\left\|\left((-A)^{\frac{\beta-1}{2}}\right)^*(-A)^{\frac{1-\beta}{2}}(-A)^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}\\
&\leq Ch^{2\beta-\epsilon}\left\|\left((-A)^{\frac{\beta-1}{2}}\right)^*(-A)^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\left\|(-A)^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}\leq Ch^{2\beta-\epsilon}.
\end{aligned} \tag{84}$$

Substituting (84) and (83) in (82) yields

$$|I_{41}|\leq Ch^{2\beta-\epsilon}\int_0^T(T-t)^{-1+\frac{\epsilon}{2}}dt\leq Ch^{2\beta-\epsilon}. \tag{85}$$

Let us move to the estimate of I_{42}. Using Proposition 2.1 yields

$$\begin{aligned}
I_{42} &= \frac{1}{2}\mathbb{E}\int_0^T Tr\left[D^2\mu(T-t,X^h(t))Q^{\frac{1}{2}}Q^{\frac{1}{2}}(P_h-\mathbf{I})\right]dt\\
&= \frac{1}{2}\mathbb{E}\int_0^T Tr\left[Q^{\frac{1}{2}}(P_h-\mathbf{I})D^2\mu(T-t,X^h(t))Q^{\frac{1}{2}}\right]dt\\
&\leq \frac{1}{2}\mathbb{E}\int_0^T\left\|Q^{\frac{1}{2}}(P_h-\mathbf{I})D^2\mu(T-t,X^h(t))Q^{\frac{1}{2}}\right\|_{\mathcal{L}_1(\mathcal{H})}dt.
\end{aligned} \tag{86}$$

Inserting appropriate powers of A, using again Proposition 2.1 and Assumption 2.2 yields

$$\begin{aligned}
|I_{42}| &\leq C\mathbb{E}\int_0^T\left\|Q^{\frac{1}{2}}\left((-A)^{\frac{\beta-1}{2}}\right)^*\left((-A)^{\frac{1-\beta}{2}}\right)^*(P_h-\mathbf{I})D^2\mu(T-t,X^h(t))Q^{\frac{1}{2}}\right\|_{\mathcal{L}_1(\mathcal{H})}dt\\
&\leq C\mathbb{E}\int_0^T\left\|Q^{\frac{1}{2}}\left((-A)^{\frac{\beta-1}{2}}\right)^*\right\|_{\mathcal{L}_2(\mathcal{H})}\left\|\left((-A)^{\frac{1-\beta}{2}}\right)^*(P_h-\mathbf{I})D^2\mu(T-t,X^h(t))Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}dt
\end{aligned}$$

$$
\begin{aligned}
&\leq C\mathbb{E}\int_0^T \left\|\left((-A)^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}\right)^*\right\|_{\mathcal{L}_2(\mathcal{H})}\left\|\left((-A)^{\frac{1-\beta}{2}}\right)^*(P_h-\mathbf{I})D^2\mu(T-t,X^h(t))Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}dt\\
&\leq C\mathbb{E}\int_0^T \left\|\left((-A)^{\frac{1-\beta}{2}}\right)^*(P_h-\mathbf{I})D^2\mu(T-t,X^h(t))Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}dt\\
&\leq C\mathbb{E}\int_0^T \left\|\left((-A)^{\frac{1-\beta}{2}}\right)^*(P_h-\mathbf{I})D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}(-A)^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}dt\\
&\leq C\mathbb{E}\int_0^T \Bigg\|\left((-A)^{\frac{1-\beta}{2}}\right)^*(-A)^{\frac{\beta-1}{2}}(-A)^{\frac{1-\beta}{2}}(P_h-\mathbf{I})\\
&\quad D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}\Bigg\|_{\mathcal{L}(\mathcal{H})}\left\|(-A)^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}\right\|_{\mathcal{L}_2(\mathcal{H})}dt\\
&\leq C\mathbb{E}\int_0^T \left\|\left((-A)^{\frac{1-\beta}{2}}\right)^*(-A)^{\frac{\beta-1}{2}}(-A)^{\frac{1-\beta}{2}}(P_h-\mathbf{I})D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathcal{H})}dt\\
&=: C\mathbb{E}\int_0^T |I_{42}^{(1)}(t)|dt. \qquad (87)
\end{aligned}
$$

Using Lemmas 4.3, 4.1 and (30) yields

$$
\begin{aligned}
|I_{42}^{(1)}(t)| &\leq \left\|\left((-A)^{\frac{1-\beta}{2}}\right)^*(-A)^{\frac{\beta-1}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\left\|(-A)^{\frac{1-\beta}{2}}(P_h-\mathbf{I})D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\\
&\leq C\left\|(-A)^{\frac{1-\beta}{2}}(P_h-\mathbf{I})D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\\
&\leq C\left\|(-A)^{\frac{1-\beta}{2}}(P_h-\mathbf{I})(-A)^{\frac{-\beta-1+\epsilon}{2}}(-A)^{\frac{1+\beta-\epsilon}{2}}D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\\
&\leq C\left\|(-A)^{\frac{1-\beta}{2}}(P_h-\mathbf{I})(-A)^{\frac{-\beta-1+\epsilon}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\left\|(-A)^{\frac{1+\beta-\epsilon}{2}}D^2\mu(T-t,X^h(t))(-A)^{\frac{1-\beta}{2}}\right\|_{\mathcal{L}(\mathcal{H})}\\
&\leq Ch^{2\beta-\epsilon}(T-t)^{-1+\frac{\epsilon}{2}}. \qquad (88)
\end{aligned}
$$

Substituting (88) in (87) yields

$$
|I_{42}| \leq Ch^{2\beta-\epsilon}\int_0^T (T-t)^{-1+\frac{\epsilon}{2}}dt \leq Ch^{2\beta-\epsilon}. \qquad (89)
$$

Substituting (89) and (85) in (72) yields

$$
|I_4| \leq Ch^{2\beta-\epsilon}. \qquad (90)
$$

Substituting (90), (80), (69) and (61) in (60) yields

$$
|\mathbb{E}[\varphi(X^h(T))-\varphi(X(T))]| \leq Ch^{2\beta-\epsilon}.
$$

This completes the proof of Theorem 4.1.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

The work of J. D. Mukam was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 317210226 – SFB 1283. We would like to thank the referees for their careful reading and insight comments which allow to improve this paper.

References

[1] Prato D, Zabczyk GJ. Stochastic equations in infinite dimensions. Vol 152. Cambridge, United Kingdom: Cambrige University Press; 2014.
[2] Prévôt C, Röckner M. A concise course on stochastic partial differential equations. Lecture notes in mathematics, vol. 1905, Berlin: Springer; 2007.
[3] Jentzen A, Kloeden PE, Winkel G. Efficient simulation of nonlinear parabolic SPDEs with additive noise. Ann Appl Probab 2011;21(3):908–50.

[4] Kovács S. Larsson M, Lindgren F. Strong convergence of the finite element method with truncated noise for semilinear parabolic stochastic equations with additive noise. Numer Algorithms 2010;53. 309–220.
[5] Kruse R. Optimal error estimates of Galerkin finite element methods for stochastic partial differential equations with multiplicative noise. IMA J Numer Anal 2014;34(1):217–51.
[6] Mukam JD, Tambue A. Strong convergence analysis of the stochastic exponential rosenbrock scheme for the finite element discretization of semilinear SPDEs driven by multiplicative and additive noise. J Sci Comput 2018;74(2):937–78.
[7] Lord GJ, Tambue A. Stochastic exponential integrators for the finite element discretization of SPDEs for multiplicative & additive noise. IMA J Numer Anal 2013;2:515–43.
[8] Mukam JD, Tambue A. A note on exponential Rosenbrock–Euler method for the finite element discretization of a semilinear parabolic partial differential equation. Comput Math Appl 2018;76(7):1719–38.
[9] Tambue A. Efficient numerical schemes for porous media flow (Ph.D. thesis), Department of Mathematics, Heriot–Watt University; 2010.
[10] Geissert M, Kovács M, Larsson S. Rate of weak convergence of the finite element method for the stochastic heat equation with additive noise. BIT Numer Math 2009;49:343–56.
[11] Andersson A, Larsson S. Weak convergence for a spatial approximation of the nonlinear stochastic heat equation. Math Comp 2016;85:1335–58.
[12] Cui J, Hong J. Strong and weak convergence rates of a spatial approximation for stochastic partial differential equations with one-sided Lipschitz coefficients. Siam J Numer Anal 2019;57(4):1815–41.
[13] Debussche A, Printems J. Weak order for the discretization of the stochastic heat equation. Math Comp 2009;78:845–63.
[14] Debussche A. Weak approximation of stochastic partial differential equations: the nonlinear case. Math Comp 2011;80:89–117.
[15] Bréhier C-E, Debussche A. Kolmogorov equations and weak order analysis for SPDEs with nonlinear diffusion coefficient. J Math Pures Appl 2018;119:193–253.
[16] Zhang Z, Rozovskii B, Karniadakis GM. Strong and weak convergence order of finite element methods for stochastic PDEs with spatial white noise. Numer Math 2016;134:61–89.
[17] Tambue A, Ngnotchouye JMT. Weak convergence for a stochastic exponential integrator and finite element discretization of stochastic partial differential equation with multiplicative & additive noise. Appl Numer Math 2016;108:57–86.
[18] Wang X, Gan S. Weak convergence analysis of the linear implicit Euler method for semilinear stochastic partial differential equations with additive noise. J Math Anal Appl 2013;398:151–69.
[19] Wang X. Weak error estimates of the exponential Euler scheme for semi-linear SPDEs without Malliavin calculus. Discrete Contin Dyn Syst 2016;36(1):481–97.
[20] Kovács M, Larsson S, Lindgren F. Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise. II. Fully discrete schemes. BIT 2013;53:497–525.
[21] Meng C, Gan S, Wang X. Weak convergence rates for an explicit full-discretization of stochastic Allen-Cahn equation with additive noise. J Sci Comput 2021;86:34. http://dx.doi.org/10.1007/s10915-020-01378-8.
[22] Andersson A, Kruse R, Larsson S. Duality in refined Sobolev-Malliavin spaces and weak approximation of SPDE. Stoch PDE: Anal Comput 2016;4:113–49.
[23] Clément E, Kohatsu-Higa A, Lamberton D. A duality appraoch for the weak approximation of stochastic differential equations. Ann Appl Probab 2006;16:1124–54.
[24] Kruse R. Strong and weak approximation of stochastic evolution equations. Lecture notes in mathematics, vol. 2093, Berlin: Springer; 2014.
[25] Chow P-L. Stochastic partial differential equations. Appl. math. nonlinear sci. ser., Chapman Hall/CRC; 2007.
[26] Fujita F, Suzuki T. Evolution problems (Part 1). In: Ciarlet PG, Lions JL, editors. Handbook of numerical analysis, Vol. 2. Amsterdam, The Netherlands: North-Holland; 1991, p. 789–928.
[27] Wang X. An exponential integrator scheme for time discretization of nonlinear stochastic wave squation. J Sci Comput 2015;64:234–63.
[28] Henry D. Geometric theory of semilinear parabolic equations. Lecture notes in mathematics, vol. 840, Berlin: Springer; 1981.
[29] Larsson S. Nonsmooth data error estimates with applications to the study of the long-time behavior of finite element solutions of semilinear parabolic problems. Preprint 1992-36, Department of Mathematics, Chalmers University of Technology; 1992.
[30] Nambu T. Characterization of the domain of fractional powers of a class of elliptic differential operators with feedback boundary conditions. J Differential Equations 1997;136:294–324.
[31] Helfrich H-P. Error estimates for semidiscrete Galerkin type approximations to semilinear evolution equations with nonsmooth initial data. Numer Math 1987;51:559–69.
[32] Thomée V. Galerkin finite element methods for parabolic problems. Berlin: Springer-Verlag; 1997.
[33] Elliot CM, Larsson SS. Error estimates with smooth and nonsmooth data for a finite element method for the Cahn-Hilliard equation. Math Comp 1992;58:603–30.