WEAK CONVERGENCE FOR A SPATIAL APPROXIMATION OF THE NONLINEAR STOCHASTIC HEAT EQUATION

ADAM ANDERSSON AND STIG LARSSON

Abstract. We find the weak rate of convergence of approximate solutions of the nonlinear stochastic heat equation, when discretized in space by a standard finite element method. Both multiplicative and additive noise is considered under different assumptions.

This extends an earlier result of Debussche in which time discretization is considered for the stochastic heat equation perturbed by white noise. It is known that this equation only has a solution in one space dimension. In order to get results for higher dimensions, colored noise is considered here, besides the white noise case where considerably weaker assumptions on the noise term is needed. Integration by parts in the Malliavin sense is used in the proof. The rate of weak convergence is, as expected, essentially twice the rate of strong convergence.

1. Introduction and main result

Let $\mathcal{D} \subset \mathbf{R}^d$ be a bounded, convex and polygonal domain. We consider, for $T > 0$, the stochastic heat equation with Dirichlet boundary condition, written in abstract form as a stochastic evolution equation in $H = L_2(\mathcal{D})$:

$$\mathrm{d}X(t) + [AX(t) - f(X(t))]\,\mathrm{d}t = g(X(t))\,\mathrm{d}W(t),\ t \in (0,T]; \quad X(0) = X_0. \tag{1.1}$$

This equation is driven by a cylindrical Q-Wiener process $(W(t))_{t\in[0,T]}$ in a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbf{P})$. The covariance operator Q is selfadjoint and positive semidefinite, not necessarily of finite trace. For technical reasons we consider a deterministic initial value $X_0 \in H$.

The leading linear operator A is, for simplicity, taken to be $-\Delta$ with domain $\mathrm{dom}(A) = H^2(\mathcal{D}) \cap H_0^1(\mathcal{D})$, where $\Delta = \sum_{k=1}^d \partial^2/\partial x_k^2$ is the Laplace operator. It is well known that $-A$ generates an analytic semigroup of bounded linear operators on H. We denote it by $(E(t))_{t\geq 0}$. The spaces $\dot{H}^\beta = \mathrm{dom}(A^{\frac{\beta}{2}})$, defined by fractional powers of A, are used to measure the spatial regularity. We denote the norm and inner product in $H = L_2(\mathcal{D})$ by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$.

Let U, V be separable Hilbert spaces and let $\mathcal{L}(U,V)$ denote the Banach space of all linear bounded operators. We denote by $\mathcal{L}_1(U,V) \subset \mathcal{L}_2(U,V) \subset \mathcal{L}(U,V)$ the subspaces consisting of trace class operators and Hilbert-Schmidt operators, respectively. We use the abbreviations $\mathcal{L}(U) = \mathcal{L}(U,U)$, $\mathcal{L} = \mathcal{L}(H)$ when $H = L_2(\mathcal{D})$, and similarly for $\mathcal{L}_p$, $p = 1, 2$. Central in the theory of stochastic integration is the space $U_0 = Q^{1/2}(H)$. We write $\mathcal{L}_2^0 = \mathcal{L}_2(U_0, H)$. By $\mathcal{C}_{\mathrm{b}}^k(U,V)$ we denote the space of not necessarily bounded functions from a Banach space U to a Banach space V that have continuous and bounded Fréchet derivatives of orders $1, \ldots, k$. For more precise definitions, see Section 2 below.

We use a "regularity parameter" β such that $\|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0} = \|A^{\frac{\beta-1}{2}}Q^{\frac12}\|_{\mathcal{L}_2} < \infty$. If $Q = I$, then $\|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0} = \|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2} < \infty$, if and only if $d = 1$ and $\beta < \frac12$, see (2.9). We consider two sets of assumptions according to the type of noise term.

A. *Additive noise in multiple dimensions.* Assume that $f \in \mathcal{C}_{\mathrm{b}}^2(H,H)$, $g(x) = I$ for all $x \in H$, and $\|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0} = \|A^{\frac{\beta-1}{2}}Q^{\frac12}\|_{\mathcal{L}_2} < \infty$ for some $\beta \in [\frac12, 1]$.

1991 *Mathematics Subject Classification.* 65M60, 60H15, 60H35, 65C30.

Key words and phrases. non-linear heat equation, SPDE, finite element, error estimate, weak convergence, multiplicative noise, Malliavin calculus.

B. *Multiplicative noise in one dimension.* Assume that $f \in \mathcal{C}_{\mathrm{b}}^2(H,H)$, $g(x) = B + Cx + \tilde{g}(x)$, where $B \in \mathcal{L}$, $C \in \mathcal{L}(H,\mathcal{L})$, and $\tilde{g} \in \mathcal{C}_{\mathrm{b}}^2(\dot{H}^{-\frac{1}{2}},\mathcal{L})$. Moreover, assume that $d = 1$, $Q = I$, and select any $\beta \in (0, \frac{1}{2})$.

Under either of these assumptions we have a unique mild solution to (1.1) satisfying the stochastic fixed point equation

$$X(t) = E(t)X_0 + \int_0^t E(t-s)f(X(s))\,\mathrm{d}s + \int_0^t E(t-s)g(X(s))\,\mathrm{d}W(s), \quad t \in [0,T]. \tag{1.2}$$

One can also show that the solution has spatial regularity of order β, i.e., it is of the form $X\colon [0,T] \times \Omega \to \dot{H}^{\beta}$, $\mathbf{P}$-almost surely, see Theorem 2.3 below and the discussion preceding it.

In this paper we consider space discretization of equation (1.1) by means of a standard finite element method. Let $(S_h)_{h\in(0,1)}$ be the family of spaces of continuous piecewise linear functions corresponding to a quasi-uniform family of triangulations of $\mathcal{D}$ with $S_h \subset H_0^1(\mathcal{D})$. The parameter h specifies the maximal diameter in the triangulation. Let $P_h\colon H \to S_h$ denote the orthogonal projection. We define the discrete Laplacian as the operator $A_h\colon S_h \to S_h$ satisfying the variational equality

$$\langle A_h\psi, \chi\rangle = \langle \nabla\psi, \nabla\chi\rangle, \quad \forall \psi, \chi \in S_h. \tag{1.3}$$

The finite element approximation of the elliptic problem $Au = f$ is the unique solution of the equation $A_h u_h = P_h f$. It is known that $\|u_h - u\| = \|A_h^{-1}P_h f - A^{-1}f\| = \mathcal{O}(h^2)$ as $h \to 0$, if $f \in L_2(\mathcal{D})$. The semigroup generated by $-A_h$ is denoted $(E_h(t))_{t\geq 0}$. The spatially semidiscrete analogue of (1.1) is to find a process $(X_h(t))_{t\in(0,T]}$ with values in S_h such that

$$\mathrm{d}X_h(t) + [A_h X_h(t) - P_h f(X(t))]\,\mathrm{d}t = P_h g(X_h(t))\,\mathrm{d}W(t),\ t \in (0,T]; \quad X_h(0) = P_h X_0, \tag{1.4}$$

or in mild form

$$\begin{aligned} X_h(t) = E_h(t)P_h X_0 &+ \int_0^t E_h(t-s)P_h f(X_h(s))\,\mathrm{d}s \\ &+ \int_0^t E_h(t-s)P_h g(X_h(s))\,\mathrm{d}W(s), \quad t \in [0,T]. \end{aligned} \tag{1.5}$$

The existence of a unique mild solution can be proved in a similar way as for (1.2). It is also known that we have strong convergence of order β under Assumptions **A** or **B**, see (2.26). Our goal is to prove weak convergence in the form

$$\mathbf{E}[G(X(T)) - G(X_h(T))] = \mathcal{O}(h^{2\beta-\epsilon}),$$

for any $\epsilon > 0$ and any testfunction $G \in \mathcal{C}_{\mathrm{b}}^2$.

For an exhaustive list of references for approximations of stochastic partial differential equations, see, e.g., [5]. We mention some works related to the situation studied here. Weak convergence of numerical schemes for linear equations with additive noise is treated in [6], [14], [13], and [19]. In the first paper full discretization of the stochastic heat equation is considered for colored noise in multiple dimension, i.e., our Assumption **A** with $f = 0$. Papers [14] and [13] deal with semidiscretization in space and full discretization, respectively, for the linear stochastic heat, Cahn-Hilliard, and wave equations, also with additive colored noise. The fourth paper provides an extension to impulsive noise.

The only results on weak convergence for nonlinear equations are those of [9], [10], [4], [5], [1] and [24]. In the work [9], discretization in time with implicit Euler and Crank-Nicolson schemes is considered for semilinear parabolic equations with additive noise. Paper [10] treats the wave equation with additive white noise, discretized by a leap-frog scheme. This case is a bit different from the others, due to the lack of analyticity of the semigroup for the wave equation in contrast to the heat equation. In [4] semidiscretization in time for the nonlinear stochastic Schrödinger equation with multiplicative white noise is considered.

The papers [4], [6], [14], [13], and [19] express the weak error by means of a Kolmogorov equation after removing the linear term $AX(t)$ by a transformation of variables. This transformation does not work for the nonlinear heat equation. This difficulty is handled in [5] by means of an integration by parts from the Malliavin calculus. This paper proves weak convergence of temporal semidiscretizations for the nonlinear heat equation with multiplicative noise in one space dimension, i.e, our Assumption **B**. Under the same assumptions, except for an extra boundedness condition on the nonlinearity, in [1] the method of [5] is exploited to prove weak convergence for the invariant measure of temporally discrete approximations. In [24] the same proof technique is used to study time discretization for the heat equation with additive noise in multiple dimensions, i.e., our Assumption **A**.

In the present paper we extend the results of [24] and [5] to spatial discretization. Our Assumptions **A** and **B** coincide with the ones in these two papers, respectively. Therefore we may quote some moment estimates from these papers. One difficulty that arises in connection with the spatial discretization is that the projector P_h does not commute with the projector onto eigenspaces of A.

In all these works the rate of weak convergence is, up to an arbitrary $\epsilon > 0$, twice that of strong convergence. The Malliavin calculus is a useful tool in the study of weak convergence of semilinear equations. It has been utilized in [9], [5], and [15] in completely different ways. It plays a central role in the proof of our Theorem 1.1, following the method of [5]. In the papers [1] and [24] the technique of [5] is also used.

The result of this paper actually concerns the convergence of the law $\mathcal{L}(X_h(T)) = \mathbf{P} \circ (X_h(T))^{-1}$ of the random variables $(X_h(T))_{h\in(0,1)}$, as the mesh size parameter $h \to 0$. We say that the law of $X_h(T)$ converges weakly to that of $X(T)$, if $\mathbf{E}[G(X_h(T))] \to \mathbf{E}[G(X(T))]$ as $h \to 0$, for all test functions $G \in \mathcal{C}_{\mathrm{b}}(H, \mathbf{R})$, the space of all bounded continuous functions on H. This convergence follows from the strong convergence $\mathbf{E}[\|X_h(T) - X(T)\|^2] = \mathcal{O}(h^{\beta})$, see [16] and the discussion below, and the weak rate obtained is thus β under mild assumptions. For $G \in \mathcal{C}^2_{\mathrm{b}}(H, \mathbf{R})$, we obtain in this paper the rate of weak convergence $2\beta - \epsilon$, for an arbitrary $\epsilon > 0$.

Theorem 1.1. *Assume either Assumption* ***A*** *or Assumption* ***B*** *and let X and X_h be the solutions of the equations* (1.2) *and* (1.5)*, respectively. Then, for every test function $G \in \mathcal{C}^2_{\mathrm{b}}(H, \mathbf{R})$ and $\gamma \in [0, \beta)$, we have the convergence*

$$|\mathbf{E}[G(X(T)) - G(X_h(T))]| = \mathcal{O}(h^{2\gamma}), \quad \text{as } h \to 0.$$

The weak error is interesting by various reasons. It measures the error made by sampling from an approximate probability law of $X(T)$, rather than the deviation from the trajectory of an exact solution, as for the strong error. The result tells us that the weak error, when approximating the quantity $\mathbf{E}[G(X(T))]$ by $\mathbf{E}[G(X_h(T))]$, is decreasing fast as $h \to 0$ for smooth G.

Section 2 is devoted to preliminaries. In Subsection 2.1 compact operators and tensor products are introduced. We need Schatten classes more general than the trace class and Hilbert-Schmidt operators. In Subsection 2.2 some notation for Fréchet derivatives is fixed. The semigroup framework and basic material on the finite element method are presented in Subsection 2.3. In Subsection 2.4 the Malliavin calculus and stochastic integration is introduced. Subsection 2.5 is about the stochastic equations (1.2) and (1.5). In Section 3 two moment estimates for the Malliavin derivative of $X_h(t)$ are proved. Section 4 contains regularity results for the Kolmogorov equation, adapting results from [5] and [24] to our setting. The proof of Theorem 1.1 is given in Section 5.

2. PRELIMINARIES

2.1. **Compact operators and tensor products.** Given two separable real Hilbert spaces $(U, \langle\cdot,\cdot\rangle_U)$ and $(V, \langle\cdot,\cdot\rangle_V)$, let $\mathcal{L}(U, V)$ denote the Banach space of all bounded and linear operators $U \to V$ endowed with the uniform norm. We write $\mathcal{L}(U) = \mathcal{L}(U, U)$. Let $(\sigma_i)_{i\in\mathcal{I}}$ be the collection of singular values of a compact operator $T \in \mathcal{L}(U)$. These are the eigenvalues of the operator $|T| = (TT^*)^{1/2}$. The index set $\mathcal{I}$ is finite or countable. Let, for $1 \le p < \infty$, the

Schatten class $\mathcal{L}_p = \mathcal{L}_p(U)$ be all $T \in \mathcal{L}(U)$ for which

$$\|T\|_{\mathcal{L}_p} = \Big(\sum_{i\in\mathcal{I}} \sigma_i^p\Big)^{\frac{1}{p}} < \infty. \tag{2.1}$$

We set by definition $\mathcal{L}_\infty = \mathcal{L}$. The Schatten classes are Banach spaces equipped with the norms (2.1). The class $\mathcal{L}_1$ is the space of trace class operators. Take an arbitrary ON-basis $(e_n)_{n\in\mathbf{N}} \subset U$. We define the trace of an operator $T \in \mathcal{L}_1(U)$ as the quantity

$$\mathrm{Tr}(T) = \sum_{i\in\mathbf{N}} \langle Te_i, e_i\rangle_U.$$

It is independent of the particular choice of ON-basis. If $T \in \mathcal{L}_1$ and $T \geq 0$, then $\mathrm{Tr}(T) = \|T\|_{\mathcal{L}_1}$. In general, the relation

$$|\,\mathrm{Tr}(T)| \leq \|T\|_{\mathcal{L}_1} \tag{2.2}$$

holds for $T \in \mathcal{L}_1$. It follows directly from the definition that $\mathrm{Tr}(T) = \mathrm{Tr}(T^*)$ for $T \in \mathcal{L}_1$. Moreover,

$$\mathrm{Tr}(ST) = \mathrm{Tr}(TS), \tag{2.3}$$

whenever $S \in \mathcal{L}(U, V)$ and $T \in \mathcal{L}(V, U)$ satisfies $ST \in \mathcal{L}_1(V)$ and $TS \in \mathcal{L}_1(U)$.

More generally, the class $\mathcal{L}_2(U, V)$ is the space of Hilbert-Schmidt operators from U to V. It is defined as the Hilbert space with the scalar product and norm

$$\langle S, T\rangle_{\mathcal{L}_2(U,V)} = \sum_{i\in\mathbf{N}} \langle Se_i, Te_i\rangle_V = \mathrm{Tr}(T^*S) = \mathrm{Tr}(ST^*), \tag{2.4}$$

$$\|T\|_{\mathcal{L}_2(U,V)} = \Big(\sum_{i\in\mathbf{N}} \|Te_i\|_V^2\Big)^{\frac{1}{2}} = \sqrt{\mathrm{Tr}(TT^*)}. \tag{2.5}$$

The choice of ON-basis $(e_n)_{n\in\mathbf{N}} \subset U$ is arbitrary. For $U = V$ the class $\mathcal{L}_2 = \mathcal{L}_2(U)$ is alone to enjoy this property. For $\mathcal{L}_p$ with $p \neq 2$, only an eigenbasis of $|T|$ can be used.

The following Hölder type inequality for Schatten classes holds:

$$\|ST\|_{\mathcal{L}_r} \leq \|S\|_{\mathcal{L}_p}\|T\|_{\mathcal{L}_q}, \tag{2.6}$$

for $r^{-1} = p^{-1} + q^{-1}$, $p, q, r \in [1, \infty]$. The border case

$$\|ST\|_{\mathcal{L}_r} \leq \|S\|_{\mathcal{L}}\|T\|_{\mathcal{L}_r} \tag{2.7}$$

is included, meaning that $\mathcal{L}_r(U)$ is an ideal of the Banach algebra $\mathcal{L}(U)$. Also

$$|\langle S, T\rangle_{\mathcal{L}_2}| = |\,\mathrm{Tr}(ST^*)| \leq \|ST^*\|_{\mathcal{L}_1} \leq \|S\|_{\mathcal{L}}\|T\|_{\mathcal{L}_1}. \tag{2.8}$$

For more about the Schatten classes see [7].

The tensor product space $U \otimes V$ of two Hilbert spaces U and V is a Hilbert space together with a bilinear mapping $U \times V \to U \otimes V, (u, v) \mapsto u \otimes v$ with dense range and with the inner product

$$\langle u_1 \otimes v_1, u_2 \otimes v_2\rangle_{U\otimes V} = \langle u_1, u_2\rangle_U \langle v_1, v_2\rangle_V, \quad u_1, u_2 \in U,\ v_1, v_2 \in V.$$

If $(u_n)_{n\in\mathbf{N}} \subset U$ and $(v_n)_{n\in\mathbf{N}} \subset V$ are ON-bases, then $(u_m \otimes v_n)_{m,n\in\mathbf{N}} \subset U \otimes V$ is an ON-basis. The space $U \otimes V$ can be realized in several isomorphic ways. If the tensor product $u \otimes v$ realizes a rank one operator $(u \otimes v)\phi = \langle v, \phi\rangle_V u$, for $\phi \in V$, then $U \otimes V \cong \mathcal{L}_2(V, U)$. If U and V are spaces of functions of independent variables $x \in \mathcal{D}_1$ and $y \in \mathcal{D}_2$, then $(u \otimes v)(x, y) = u(x)v(y)$ is also a realization of $U \otimes V$. For instance, if $U = L_2(\mathcal{D})$ and $V = L_2(\Omega)$, where $\mathcal{D}$ is our spatial domain and Ω the sample space, then $U \otimes V = L_2(\Omega \times \mathcal{D}) \cong L_2(\Omega, L_2(\mathcal{D}))$, i.e., $L_2(\mathcal{D})$-valued square integrable random variables. For a detailed introduction to tensor products, see [11, Appendix E].

2.2. **Fréchet derivatives.** Let $(U, \|\cdot\|_U)$ and $(V, \|\cdot\|_V)$ be Banach spaces. By $\mathcal{C}_{\mathrm{b}}^m(U,V)$ we denote the space of not necessarily bounded mappings $g\colon U \to V$ having m continuous and bounded Fréchet derivatives $Dg, D^2g, \dots, D^m g$. We endow it with the seminorm $|\cdot|_{\mathcal{C}_{\mathrm{b}}^m(U,V)}$, determined as the smallest constant $C \geq 0$ such that

$$\sup_{x\in U} \|D^m g(x)\cdot(\phi_1,\dots,\phi_m)\|_V \leq C\|\phi_1\|_U \cdots \|\phi_m\|_U, \quad \forall \phi_1,\dots,\phi_m \in U.$$

It will be convenient to write $\mathcal{C}_{\mathrm{b}}^m = \mathcal{C}_{\mathrm{b}}^m(U,V)$. From the context it will be clear what we mean.

Let us consider the important case when U is a Hilbert space and $V = \mathbf{R}$. The Fréchet derivative $Dg(x)$ of a function $g\colon U \to \mathbf{R}$ is a bounded linear functional on U for fixed $x \in H$ and it can thus be identified by its gradient using the Riesz representation theorem, i.e., $Dg(x)\cdot\phi = \langle Dg(x), \phi\rangle$. In the same way the second derivative enjoys a representation as a bounded linear operator by the identity $D^2g(x)\cdot(\phi,\psi) = \langle D^2g(x)\phi, \psi\rangle$. We will use both representations and it will lead to no confusion.

2.3. **The functional analytic framework.** We will now introduce the semigroup framework on which our analysis of equations (1.2) and (1.5) relies. Recall from Section 1 that $A = -\Delta$ with $\mathrm{dom}(A) = H^2(\mathcal{D}) \cap H_0^1(\mathcal{D})$ and $H = L_2(\mathcal{D})$ with $\mathcal{D} \subset \mathbf{R}^d$ a convex polygonal domain. We denote $\|\cdot\| = \|\cdot\|_H$ and $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_H$. The operator A is closed, selfadjoint and positive definite.

There is an orthonormal eigenbasis $(\varphi_i)_{i\in\mathbf{N}} \subset H$ with corresponding eigenvalues $0 < \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_i \to \infty$, as $i \to \infty$, for which $A\varphi_i = \lambda_i\varphi_i$, $i \in \mathbf{N}$. The asymptotics $\lambda_i \sim i^{2/d}$, as $i \to \infty$, is well known. When the space dimension $d = 1$, as in Assumption **B**, we have

$$\mathrm{Tr}(A^{-\frac{1}{2}\gamma}) = \|A^{-\frac{1}{2}\gamma}\|_{\mathcal{L}_1} = \|A^{-\frac{1}{4}\gamma}\|_{\mathcal{L}_2}^2 < \infty, \quad \forall\gamma > 1, \text{ if } d = 1. \tag{2.9}$$

This means that $\beta \in (0,1)$ under Assumption **B**.

We define norms of fractional orders by

$$\|v\|_{\dot{H}^\beta} = \|A^{\frac{\beta}{2}} v\| = \Big(\sum_{i\in\mathbf{N}} \lambda_i^\beta \langle v, \varphi_i\rangle^2\Big)^{\frac{1}{2}}, \quad \beta \in \mathbf{R}.$$

The spaces $\dot{H}^\beta$ are then, for $\beta \geq 0$, defined as $\mathrm{dom}(A^{\frac{\beta}{2}})$ and for $\beta < 0$ as the closure of H with respect to the $\dot{H}^\beta$-norm. The space $\dot{H}^{-\gamma}$ of negative order can be identified with the dual space of $\dot{H}^\gamma$. Clearly $\dot{H}^0 = H$, and it is also well known that $\dot{H}^1 = H_0^1(\mathcal{D})$ and $\dot{H}^2 = H^2(\mathcal{D}) \cap H_0^1(\mathcal{D})$, see [22, Ch. 3].

Let $(S_h)_{h\in(0,1)}$ denote a family of standard finite element spaces of continuous piecewise linear functions corresponding to a quasi-uniform family of triangulations, for which h denotes the largest diameter in the triangulation. Then $S_h \subset \dot{H}^1$. By P_h we denote the orthogonal projector of H onto S_h. Let $A_h\colon S_h \to S_h$ be the unique operator satisfying

$$\langle A_h\psi, \chi\rangle = \langle \nabla\psi, \nabla\chi\rangle, \quad \forall \psi,\chi \in S_h.$$

This is the discrete Laplacian. By definition

$$\|A_h^{\frac{1}{2}}\varphi_h\| = \|\nabla\varphi_h\| = \|A^{\frac{1}{2}}\varphi_h\| = \|\varphi_h\|_{\dot{H}^1}, \quad \varphi_h \in S_h. \tag{2.10}$$

Therefore, P_h can be extended to $\dot{H}^{-1}$, so that for all $\varphi \in \dot{H}^{-1}$,

$$\|A_h^{-\frac{1}{2}} P_h\varphi\| = \sup_{\psi\in S_h} \frac{\langle\varphi,\psi\rangle}{\|A_h^{\frac{1}{2}}\psi\|} = \sup_{\psi\in S_h} \frac{\langle\varphi,\psi\rangle}{\|A^{\frac{1}{2}}\psi\|} \leq \sup_{\psi\in \dot{H}^1} \frac{\langle\varphi,\psi\rangle}{\|A^{\frac{1}{2}}\psi\|} = \|A^{-\frac{1}{2}}\varphi\|. \tag{2.11}$$

Moreover,

$$\|A_h^{\frac{1}{2}} P_h\varphi\| \leq C\|A^{\frac{1}{2}}\varphi\|, \quad \varphi \in \dot{H}^1, \text{ uniformly in } h. \tag{2.12}$$

This follows from (2.10) and the well-known fact that P_h is bounded with respect to $\|\cdot\|_{\dot{H}^1} = \|A^{\frac{1}{2}}\cdot\|$, when we use a quasi-uniform mesh family. Interpolation between this and (2.11) yields

$$\|A_h^\gamma P_h\varphi\| \leq C\|A^\gamma\varphi\|, \quad \varphi \in \dot{H}^\gamma, \ \gamma \in [-\tfrac{1}{2}, \tfrac{1}{2}]. \tag{2.13}$$

Furthermore, (2.12) means that $\|A_h^{\frac12} P_h A^{-\frac12}\|_{\mathcal{L}} < \infty$. Hence,

$$\|A^{-\frac12} A_h^{\frac12} P_h\|_{\mathcal{L}} = \|(A^{-\frac12} A_h^{\frac12} P_h)^*\|_{\mathcal{L}} = \|A_h^{\frac12} P_h A^{-\frac12}\|_{\mathcal{L}} \le C,$$

so that $\|A^{\frac12} A_h^{-\frac12} P_h \varphi\| \le C\|\varphi\|$ or

$$\|A^{-\frac12}\varphi_h\| \le C\|A_h^{-\frac12}\varphi_h\|, \quad \varphi_h \in S_h.$$

Interpolating between this and (2.10) yields

$$\|A^{\gamma}\varphi_h\| \le C\|A_h^{\gamma}\varphi_h\|, \quad \varphi_h \in S_h,\ \gamma \in [-\tfrac12, \tfrac12].$$

Using also (2.13) yields the norm equivalence

$$c\|A_h^{\gamma}\varphi_h\| \le \|A^{\gamma}\varphi_h\| \le C\|A_h^{\gamma}\varphi_h\|, \quad \varphi_h \in S_h,\ \gamma \in [-\tfrac12, \tfrac12]. \tag{2.14}$$

The interpolations above are valid since $(\dot H^{\beta})_{\beta\in[-1,1]}$ and $(\dot H_h^{\beta})_{\beta\in[-1,1]}$ are real interpolation spaces, where $\dot H_h^{\beta} = S_h$ with norm $\|v_h\|_{\dot H_h^{\beta}} = \|A_h^{\frac{\beta}{2}} v_h\|$. For positive order this is standard, see for instance [20]. For negative order, let $\beta \in [0,1]$ and notice that

$$[\dot H^0, \dot H^{-1}]_{\beta,2} = [(\dot H^0)^*, (\dot H^1)^*]_{\beta,2} = [\dot H^0, \dot H^1]_{\beta,2}^* = (\dot H^{\beta})^* = \dot H^{-\beta}.$$

We define the Ritz projector $R_h\colon \dot H^1 \to S_h$ to be the orthogonal projection with respect to the $\dot H^1$-scalar product. Since $\mathcal{D}$ is convex and polygonal it is well known that

$$\|A^{\frac{s}{2}}(I - R_h)A^{-\frac{r}{2}}\|_{\mathcal{L}} \le Ch^{r-s}, \quad 0 \le s \le 1 \le r \le 2. \tag{2.15}$$

For P_h the following error estimate holds

$$\|A^{\frac{s}{2}}(I - P_h)A^{-\frac{r}{2}}\|_{\mathcal{L}} \le Ch^{r-s}, \quad 0 \le s \le 1, \quad 0 \le s \le r \le 2. \tag{2.16}$$

For more about the finite element method, see [2] for elliptic equations and [22] for parabolic.

Denote by N_h the dimension of S_h. There is an orthonormal eigenbasis $(\varphi_i^h)_{i=1}^{N_h} \subset S_h$ corresponding to A_h with eigenvalues $0 < \lambda_1^h \le \lambda_2^h \le \cdots \le \lambda_{N_h}^h$. The operators $-A$ and $-A_h$ generate analytic semigroups $(E(t))_{t\ge0}$ and $(E_h(t))_{t\ge0}$, respectively. They are spectrally given by

$$E(t)v = \sum_{i\in\mathbf{N}} e^{-\lambda_i t}\langle v, \varphi_i\rangle \varphi_i, \quad v \in H,\ t \ge 0, \tag{2.17}$$

and

$$E_h(t)v_h = \sum_{i=1}^{N_h} e^{-\lambda_i^h t}\langle v_h, \varphi_i^h\rangle \varphi_i^h, \quad v_h \in S_h,\ t \ge 0.$$

The semigroup $(E_h(t))_{t\ge0}$ solves the parabolic equation $\dot u_h + A_h u_h = 0$, $t \ge 0$, with $u_h(0) = P_h v$, in the sense that $u_h(t) = E_h(t)P_h v$.

Important for our analysis is the estimate

$$\|A^{\gamma}E(t)\|_{\mathcal{L}} + \|A_h^{\gamma}E_h(t)P_h\|_{\mathcal{L}} \le C_{\gamma} t^{-\gamma}, \quad \gamma \ge 0,\ t > 0, \quad \text{uniformly in } h. \tag{2.18}$$

It is standard and is enjoyed by all analytic semigroups.

Let P_m denote the spectral projection onto the space spanned by the m first eigenvectors $(\varphi_i)_{i=1}^m$ of A. An easy calculation shows that

$$\|(I - P_m)A^{-r}\|_{\mathcal{L}} \le \lambda_m^{-r}, \quad r \ge 0. \tag{2.19}$$

In our analysis we will use the notation $a \lesssim b$, to mean that there exists a constant $C > 0$ such that $a \le Cb$. The constant will never depend on the mesh size h.

We will frequently use the following Gronwall lemma:

Lemma 2.1 (Generalized Gronwall lemma)**.** *Let $\varphi(t) \ge 0$ be a continuous function on $[0,T]$. If, for some $A, B \ge 0$ and $\alpha, \beta \in [0,1)$, the inequality*

$$\varphi(t) \le At^{-\alpha} + B\int_0^t (t-s)^{-\beta}\varphi(s)\,\mathrm{d}s$$

holds, then there is $C = C(B, T, \alpha, \beta)$ *such that*

$$\varphi(t) \le CAt^{-\alpha}, \quad t \in (0, T].$$

2.4. **The stochastic integral and Malliavin calculus.** Since we use the Malliavin calculus in the proof of our main result, we outline a framework for the stochastic integral in which this calculus has a natural role. This is an alternative to the more classical procedure, presented in [3]. Our presentation of the Wiener integral relies on [23], and the Malliavin calculus on [8] and [18], where a natural extension of the framework of [21] to Hilbert space valued stochastic integrals using tensor products is presented.

The covariance operator $Q \in \mathcal{L}(H)$ is self adjoint and positive semidefinite. Let $Q^{1/2}$ denote the unique positive square root. Let $Q^{-1/2}$ be its inverse, restricted to $(\ker Q)^{\perp}$. Define the Hilbert space $U_0 = Q^{1/2}(H)$, equipped with the scalar product $\langle u, v\rangle_{U_0} = \langle Q^{-1/2}u, Q^{-1/2}v\rangle$. If $\mathrm{Tr}(Q) < \infty$, then the triple $i\colon U_0 \hookrightarrow H$ is an abstract Wiener space, where i is the inclusion mapping $i\colon x \mapsto x$. This triple induces a Gaussian probability measure on H with mean 0 and covariance Q. It is referred to as an abstract Wiener measure. The space U_0 is called the Cameron-Martin space in this context.

Let $I\colon L_2([0, T], U_0) \to L_2(\Omega)$ be an isonormal process: For every $\phi \in L_2([0, T], U_0)$ the random variable $I(\phi)$ is centered Gaussian and I has the covariance structure

$$\mathbf{E}[I(\phi)I(\psi)] = \langle \phi, \psi\rangle_{L_2([0,T],U_0)}, \quad \phi, \psi \in L_2([0, T], U_0).$$

The existence of I follows by an application of the Kolmogorov Extension Theorem.

Define, for $u \in U_0$, the cylindrical Q-Wiener process $W\colon [0, T] \times U_0 \to L_2(\Omega)$ by

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

For $u \in U_0$ the process $W(t)u$, $t \in [0, T]$ is a Brownian motion and given $u, v \in U_0$

$$\mathbf{E}[W(t)uW(s)v] = \min(s, t)\langle u, v\rangle_{U_0}.$$

The space of Hilbert-Schmidt operators $\mathcal{L}_2^0 = \mathcal{L}_2(U_0, H)$ can be identified with $H \otimes U_0$, and $h \otimes u \in \mathcal{L}_2^0$ for $h \in H$, $u \in U_0$ being the operator $(h \otimes u)v = \langle u, v\rangle_{U_0} h$, $v \in U_0$.

We now define the H-valued Wiener integral for the simplest possible integrands. Let $\Phi = \chi_{[a,b]} \otimes (h \otimes u) \in L_2([0, T], \mathcal{L}_2^0)$, for $a, b \in [0, T]$, $h \in H$ and $u \in U_0$. Then the Wiener integral of Φ is defined as the H-valued random variable

$$\int_0^T \Phi(s)\,\mathrm{d}W(s) = I(\chi_{[a,b]} \otimes u) \otimes h = \big(W(b)u - W(a)u\big) \otimes h \in L_2(\Omega, H).$$

It is not difficult to show that for such integrands the following property, known as Wiener's isometry, holds:

$$\mathbf{E}\Big[\Big\| \int_0^T \Phi(t)\,\mathrm{d}W(t)\Big\|_H^2\Big] = \int_0^T \|\Phi(t)\|_{\mathcal{L}_2^0}^2\,\mathrm{d}t.$$

The integral extends directly to linear combinations of such integrands by linearity of I. By the Wiener isometry, completeness of $L_2([0, T], \mathcal{L}_2^0)$ and classical approximation results for $L_2([0, T])$-functions and for compact operators, it extends to all $\Phi \in L_2([0, T], \mathcal{L}_2^0)$.

Let $\mathcal{C}_{\mathrm{p}}^{\infty}(\mathbf{R}^n)$ denote the space of all $\mathcal{C}^{\infty}$-functions over $\mathbf{R}^n$ with polynomial growth. We define the family of smooth cylindrical random variables

$$\mathcal{S} = \{X = f(I(\phi_1), \dots, I(\phi_N)) : f \in \mathcal{C}_{\mathrm{p}}^{\infty}(\mathbf{R}^N),\ \phi_1, \dots, \phi_N \in L_2([0, T], U_0),\ N \ge 1\}$$

and the corresponding family with values in H as

$$\mathcal{S}(H) = \Big\{F = \sum_{i=1}^{M} X_i \otimes h_i : X_1, \dots, X_M \in \mathcal{S},\ h_1, \dots, h_M \in H,\ M \ge 1\Big\}.$$

The Malliavin derivative of a random variable in $\mathcal{S}$ with representation $X = f(I(\phi_1), \dots, I(\phi_N))$ is defined as the $L_2([0, T], U_0)$-valued random variable $DX = \sum_{i=1}^{N} \partial_i f(I(\phi_1), \dots, I(\phi_N)) \otimes \phi_i$. Clearly this is a U_0-valued stochastic process. We write $D_t X = \sum_{i=1}^{N} \partial_i f(I(\phi_1), \dots, I(\phi_N)) \otimes$

$\phi_i(t)$ for $t \in [0,T]$. The Malliavin derivative of a random variable $F \in \mathcal{S}(H)$ with the representation $F = \sum_{i=1}^M f_i(I(\phi_1),\dots,I(\phi_N)) \otimes h_i$ is given by

$$D_t F = \sum_{i=1}^M \sum_{j=1}^N \partial_j f_i(I(\phi_1),\dots,I(\phi_N)) \otimes (h_i \otimes \phi_j(t)).$$

Thus $(D_tF)_{t\in[0,T]}$ is an $\mathcal{L}_2^0$-valued stochastic process. By $D_t^u F$ we denote the derivative of F in the direction $u \in U_0$ at time t, i.e., $D_t^u F = D_t F u$, where

$$D_t F u = \sum_{i=1}^M \sum_{j=1}^N \langle u, \phi_j(t)\rangle_{U_0}\, \partial_j f_i(I(\phi_1),\dots,I(\phi_N)) \otimes h_i.$$

At the very heart of Malliavin calculus is the following integration by parts formula. It tells that, for all $F \in \mathcal{S}(H)$ and $\Phi \in L_2([0,T],\mathcal{L}_2^0)$,

$$\mathbf{E}\langle DF, \Phi\rangle_{L_2([0,T],\mathcal{L}_2^0)} = \mathbf{E}\Big\langle F, \int_0^T \Phi(t)\,\mathrm{d}W(t)\Big\rangle_H. \tag{2.20}$$

Thus the Wiener integral is the adjoint of $D\colon \mathcal{S}(H) \subset L_2(\Omega,H) \to L_2(\Omega\times[0,T],\mathcal{L}_2^0)$ for deterministic integrands. Formula (2.20) follows from the corresponding formula for real-valued smooth stochastic variables. The derivative operator D is known to be closable. We define the Watanabe Sobolev space $\mathbf{D}^{1,2}(H)$ as the closure of $\mathcal{S}(H)$ with respect to the norm

$$\|F\|_{\mathbf{D}^{1,2}(H)} = \Big(\mathbf{E}\big[\|F\|_H^2\big] + \mathbf{E}\Big[\int_0^T \|D_tF\|_{\mathcal{L}_2^0}^2\,\mathrm{d}t\Big]\Big)^{\frac{1}{2}}.$$

Denote by $\mathrm{dom}(\delta)$ the elements $\Phi \in L_2(\Omega\times[0,T],\mathcal{L}_2^0)$ for which $\mathbf{E}[\langle DF,\Phi\rangle_{L_2([0,T],\mathcal{L}_2^0)}]$ defines a bounded linear functional for $F \in \mathbf{D}^{1,2}(H)$. For any such Φ the functional $l_\Phi(F) = \mathbf{E}[\langle DF,\Phi\rangle_{L_2([0,T],\mathcal{L}_2^0)}]$ can be extended by continuity to all $F \in L_2(\Omega,H)$. The Riesz representation theorem guarantees the existence of an adjoint operator to D, namely $\delta\colon \mathrm{dom}(\delta) \subset L_2(\Omega\times[0,T],\mathcal{L}_2^0) \to L_2(\Omega,H)$ that satisfies

$$\mathbf{E}[\langle DF,\Phi\rangle_{L_2([0,T],\mathcal{L}_2^0)}] = \mathbf{E}[\langle F,\delta(\Phi)\rangle_H], \quad \forall F \in \mathbf{D}^{1,2}(H). \tag{2.21}$$

This is a natural extension of (2.20) to a much larger class of integrands. In [8, Lemme 2.10] it is proved that for any predictable process $\Phi \in L_2(\Omega\times[0,T],\mathcal{L}_2^0)$ the action of δ on Φ coincides with that of the Itô integral, i.e.,

$$\delta(\Phi) = \int_0^T \Phi(t)\,\mathrm{d}W(t).$$

Instead of relying on Itô theory we take this as the definition of the Itô integral. We remark that $\mathrm{dom}(\delta)$ contains processes that are not predictable and thus δ is an extension of the Itô integral to such integrands. In this context δ is called the Skorohod integral.

The following lemma [5, Lemma 2.1] has a central role in the proof of our main result.

Lemma 2.2. *For any random variable $F \in \mathbf{D}^{1,2}(H)$ and any predictable process $\Phi \in L_2([0,T]\times\Omega,\mathcal{L}_2^0)$ the following integration by parts formula is valid.*

$$\mathbf{E}\Big[\Big\langle \int_0^T \Phi(t)\,\mathrm{d}W(t), F\Big\rangle_H\Big] = \mathbf{E}\Big[\int_0^T \langle \Phi(t), D_tF\rangle_{\mathcal{L}_2^0}\,\mathrm{d}t\Big].$$

Proof. This is just a restatement of (2.21) for Φ predictable. □

A corollary of Lemma 2.2 is the Itô isometry. It reads

$$\mathbf{E}\Big[\Big\|\int_0^T \Phi(t)\,\mathrm{d}W(t)\Big\|_H^2\Big] = \mathbf{E}\Big[\int_0^T \|\Phi(t)\|_{\mathcal{L}_2^0}^2\,\mathrm{d}t\Big], \quad \forall \Phi \in L_2([0,T]\times\Omega,\mathcal{L}_2^0),\ \text{predictable}. \tag{2.22}$$

The Malliavin derivative acts on its adjoint by $D_s^u\delta(\Phi) = \delta(D_s^u\Phi) + \Phi(s)u$ or in terms of the Itô integral $\delta(\chi_{[0,t]}\Phi) = \int_0^t \Phi(r)\,\mathrm{d}W(r)$ with a predictable $\Phi \in L_2([0,T]\times\Omega,\mathcal{L}_2^0)$ satisfying

$\Phi(t) \in \mathbf{D}^{1,2}(\mathcal{L}_2^0)$ for all $t \in [0,T]$:

$$D_s^u \int_0^t \Phi(r)\,\mathrm{d}W(r) = \int_0^t D_s^u \Phi(r)\,\mathrm{d}W(r) + \Phi(s)u, \quad 0 \le s \le t \le T. \tag{2.23}$$

If $s > t$, then $D_s^u \int_0^t \Phi(r)\,\mathrm{d}W(r) = 0$, since the integral is $\mathcal{F}_t$-measurable. The class of $F \in \mathbf{D}^{1,2}(H)$ that are $\mathcal{F}_0$-measurable coincides with the class of constant deterministic random variables. Let V be another separable real Hilbert space and $\sigma \in \mathcal{C}_{\mathrm{b}}^1(H,V)$. Then $\sigma(F) \in \mathbf{D}^{1,2}(V)$ and

$$D_t^u(\sigma(F)) = D\sigma(F) \cdot D_t^u F, \quad u \in U_0,\ F \in \mathbf{D}^{1,2}(H), \tag{2.24}$$

$$D_t(\sigma(F)) = D\sigma(F) D_t F, \qquad F \in \mathbf{D}^{1,2}(H). \tag{2.25}$$

2.5. **Existence and uniqueness.** Existence and uniqueness of a solution to (1.2), under Assumption **A** with $\beta = 1$, is stated as [3, Theorem 7.4]. This is the case when $\mathrm{Tr}(Q) < \infty$. The extension to $\beta \in [\frac{1}{2}, 1)$ is straight-forward. For Assumption **B** existence and uniqueness is given as [3, Theorem 7.6]. By using the methods of [12] and [17] one can show that the regularity in space is of order β, i.e., the solution X is of the form $[0,T] \times \Omega \to \dot{H}^\beta$, $\mathbf{P}$-a.s.. Recall here that β is some number $\beta \in [\frac{1}{2}, 1]$ under Assumption **A** and any number $\beta \in (0, \frac{1}{2})$ under Assumption **B**. The family $(X_h)_{h\in(0,1)}$ of solution processes of the discrete equation (1.5), corresponding to the family of triangulations, is treated analogously and clearly $X_h(t) \in S_h \subset \dot{H}^1$, $\mathbf{P}$-a.s.. The estimate $\mathbf{E}\|A^{\frac{\gamma}{2}} X_h(t)\|^2 \le C(1 + \|X_0\|^2)$, $\gamma \in [0,1]$ is uniform in h, only for $\gamma \in [0,\beta]$. The strong convergence

$$\big(\mathbf{E}\|X(T) - X_h(T)\|^2\big)^{\frac{1}{2}} \le Ch^\beta, \tag{2.26}$$

is proved in [16] under the assumption of trace class noise. The proof is easier under Assumptions **A** and **B**. We formulate a qualitative bound for the solution processes in the following theorem.

Theorem 2.3. *Under either Assumption **A** or Assumption **B** there exists unique predictable solutions $X \in C([0,T], L_2(\Omega, H))$ and $X_h \in C([0,T], L_2(\Omega, S_h))$ to equation (1.2) and* (1.5) *respectively. We refer to these solutions as the unique mild solutions of (1.1) and* (1.5)*. There exists a constant C, such that the following moment estimate holds*

$$\sup_{t\in[0,T]} \mathbf{E}\|X(t)\|^2 + \sup_{t\in[0,T]} \mathbf{E}\|X_h(t)\|^2 \le C(1 + \|X_0\|^2). \tag{2.27}$$

3. Estimates of the Malliavin derivative of the solution

We consider the Malliavin derivative of the discrete solution process and prove some estimates needed later. Differentiating the equation (1.5) formally in direction $u \in U_0$, using (2.23), (2.24), and the fact that we have a deterministic initial value, yields

$$\begin{aligned} D_s^u X_h(t) &= E_h(t-s)P_h g(X_h(s))u + \int_s^t E_h(t-r)P_h Df(X_h(r)) \cdot D_s^u X_h(r)\,\mathrm{d}r \\ &\quad + \int_s^t E_h(t-r)P_h\big(Dg(X_h(r)) \cdot D_s^u X_h(r)\big)\,\mathrm{d}W(r), \quad 0 \le s \le t \le T. \end{aligned} \tag{3.1}$$

This equation is treated much like (1.5) itself. It has a unique solution.

Before we proceed to the estimate of the Malliavin derivative we notice that, by the linear growth of f and g, implied by their bounded first derivative, and the moment estimate (2.27) for X and X_h yields

$$\sup_{t\in[0,T]} \mathbf{E}\|f(Y(t))\|^2 + \sup_{t\in[0,T]} \mathbf{E}\|A^{\frac{\beta-1}{2}} g(Y(t))\|_{\mathcal{L}_2^0}^2 \lesssim 1 + \|X_0\|^2, \quad Y = X \text{ or } X_h. \tag{3.2}$$

Lemma 3.1. *Consider equation (1.5) under Assumption **A**. Then the Malliavin derivative of X_h, given as the solution $D_s X_h$ to equation (3.1), satisfies for some constant $C = C(T) > 0$ the bound:*

$$\mathbf{E}\big[\|A_h^{\frac{\beta-1}{2}} D_s X_h(t)\|_{\mathcal{L}_2^0}^2\big] \le C, \quad 0 \le s \le t \le T.$$

Proof. We make use of equation (3.1) with $g(x) = I$, $Dg(x) = 0$, for the proof and recall that $\beta - 1 \in [-\frac{1}{2}, 0]$. Fix $u \in U_0$. Thanks to the Cauchy-Schwarz inequality we get that

$$\mathbf{E}\|D_s^u X_h(t)\|^2 \lesssim \|E_h(t-s) A_h^{\frac{1-\beta}{2}} A_h^{\frac{\beta-1}{2}} P_h u\|^2 + \int_s^t \mathbf{E}\|E_h(t-r) P_h Df(X_h(r)) \cdot D_s^u X_h(r)\|^2 \,\mathrm{d}r.$$

In view of (2.13) and the boundedness of Df and $E_h(t)$ we have

$$\mathbf{E}\|D_s^u X_h(t)\|^2 \lesssim \|A_h^{\frac{1-\beta}{2}} E_h(t-s) P_h\|_{\mathcal{L}}^2 \,\|A^{\frac{\beta-1}{2}} u\|^2 + \int_s^t |f|_{\mathcal{C}_{\mathrm{b}}^1}^2 \,\mathbf{E}\|D_s^u X_h(r)\|^2 \,\mathrm{d}r. \tag{3.3}$$

The analyticity of the semigroup (2.18) yields

$$\mathbf{E}\|D_s^u X_h(t)\|^2 \lesssim (t-s)^{\beta-1} \|A^{\frac{\beta-1}{2}} u\|^2 + \int_s^t \mathbf{E}\|D_s^u X_h(r)\|^2 \,\mathrm{d}r$$

and applying Gronwall's Lemma 2.1, for fixed $s \in [0, t)$, gives

$$\mathbf{E}\|D_s^u X_h(t)\|^2 \lesssim (t-s)^{\beta-1} \|A^{\frac{\beta-1}{2}} u\|^2. \tag{3.4}$$

Proceeding as in the proof of (3.3), we obtain also

$$\mathbf{E}\|A_h^{\frac{\beta-1}{2}} D_s^u X_h(t)\|^2 \lesssim \|A^{\frac{\beta-1}{2}} u\|^2 + \int_s^t \mathbf{E}\|D_s^u X_h(r)\|^2 \,\mathrm{d}r.$$

Estimate (3.4) is applicable. Thus

$$\int_s^t \mathbf{E}\|D_s^u X_h(r)\|^2 \,\mathrm{d}r \lesssim \int_s^t (r-s)^{\beta-1} \,\mathrm{d}r \;\|A^{\frac{\beta-1}{2}} u\|^2 \lesssim (t-s)^{\beta} \|A^{\frac{\beta-1}{2}} u\|^2,$$

and hence

$$\mathbf{E}\|A_h^{\frac{\beta-1}{2}} D_s^u X_h(t)\|^2 \lesssim \|A^{\frac{\beta-1}{2}} u\|^2. \tag{3.5}$$

Notice that this is uniform with respect to $u \in U_0$. We take an ON-basis $(u_i)_{i\in\mathbf{N}} \subset U_0$ and compute the $\mathcal{L}_2^0$-norm according to (2.4). Using Tonelli's Theorem and (3.5) we get that

$$\begin{aligned}\mathbf{E}\|A_h^{\frac{\beta-1}{2}} D_s X_h(t)\|_{\mathcal{L}_2^0}^2 &= \mathbf{E}\sum_{i\in\mathbf{N}} \|A_h^{\frac{\beta-1}{2}} D_s^{u_i} X_h(t)\|^2 = \sum_{i\in\mathbf{N}} \mathbf{E}\|A_h^{\frac{\beta-1}{2}} D_s^{u_i} X_h(t)\|^2 \\ &\lesssim \sum_{i\in\mathbf{N}} \|A^{\frac{\beta-1}{2}} u_i\|^2 = \|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0}^2.\end{aligned}$$

□

For the white noise case we will need the following lemma that is a space discrete analogue of [5, Lemma 4.3]. Recall that in this case $Q = I$, $U_0 = H$, $\mathcal{L}_2^0 = \mathcal{L}_2$

Lemma 3.2. *Consider equation (1.5) under Assumption **B**. Then, for $\gamma \in [0, \frac{1}{2})$, the Malliavin derivative satisfies the following estimate:*

$$\mathbf{E}\|A_h^{\frac{\gamma}{2}} D_s X_h(t)\|_{\mathcal{L}}^2 \le C(t-s)^{-\gamma}, \quad 0 \le s \le t \le T.$$

Proof. Let $u \in H$, and take norms in (3.1) using the Cauchy-Schwarz inequality and the Itô isometry (2.22) to get

$$\begin{aligned}\mathbf{E}\|A_h^{\frac{\gamma}{2}} D_s^u X_h(t)\|^2 \lesssim{}& \mathbf{E}\|A_h^{\frac{\gamma}{2}} E_h(t-s) P_h g(X_h(s)) u\|^2 \\ &+ \int_s^t \mathbf{E}\|A_h^{\frac{\gamma}{2}} E_h(t-s) P_h Df(X_h(s)) \cdot D_s^u X_h(r)\|^2 \,\mathrm{d}r \\ &+ \int_s^t \mathbf{E}\|A_h^{\frac{1}{4}+\epsilon} A_h^{\frac{\gamma}{2}} E_h(t-s) A_h^{-\frac{1}{4}-\epsilon} P_h Dg(X_h(s)) \cdot D_s^u X_h(r)\|_{\mathcal{L}_2}^2 \,\mathrm{d}r.\end{aligned}$$

For $\epsilon > 0$ small enough we have by (2.7) and (2.18)

$$\begin{aligned}\mathbf{E}\|A_h^{\frac{\gamma}{2}} D_s^u X_h(t)\|^2 &\lesssim (t-s)^{-\gamma} \sup_{s\in[0,T]} \mathbf{E}\|g(X_h(s))\|_{\mathcal{L}}^2 \|u\|^2 \\ &\quad + \int_s^t (t-r)^{-\gamma}\, |f|^2_{\mathcal{C}^1_{\mathrm{b}}}\, \mathbf{E}\|A_h^{\frac{\gamma}{2}} D_s^u X_h(r)\|^2 \,\mathrm{d}r \\ &\quad + \int_s^t (t-r)^{-\gamma-\frac12-2\epsilon} \|A_h^{-\frac14-\epsilon} P_h\|^2_{\mathcal{L}_2}\, |g|^2_{\mathcal{C}^1_{\mathrm{b}}}\, \mathbf{E}\|A_h^{\frac{\gamma}{2}} D_s^u X_h(r)\|^2 \,\mathrm{d}r.\end{aligned}$$

By (2.9), (2.7) and (2.13) we have

$$\|A_h^{-\frac14-\epsilon} P_h\|_{\mathcal{L}_2} \lesssim \|A_h^{-\frac14-\epsilon} P_h A^{\frac14+\epsilon}\|_{\mathcal{L}} \|A^{-\frac14-\epsilon}\|_{\mathcal{L}_2} \lesssim \|A^{-\frac14-\epsilon}\|_{\mathcal{L}_2} < \infty$$

and by Gronwall's Lemma 2.1 and (3.2)

$$\mathbf{E}\|A_h^{\frac{\gamma}{2}} D_s^u X_h(t)\|^2 \lesssim (t-s)^{-\gamma}(1+\|X_0\|^2)\|u\|^2.$$

□

4. Regularity results for the Kolmogorov equation

In [5], [1] and [24], in the case of discretization in time, the proofs of the weak convergence is performed for finite-dimensional spectral Galerkin approximations. The use of the Itô formula and the Kolmogorov equation is in this way justified. The estimates are uniform in the dimension $m \in \mathbf{N}$ of the approximation space and they thus hold in the limit. The approximation is not made explicit in the proof. For the discretization in space some more care need to be taken. This is due to the fact that the operators A and A_h do not commute.

Recall that P_m is the projection onto the subspace of $H_m \subset H$ spanned by the first $m \in \mathbf{N}$ eigenvectors $(\varphi_i)_{i=1}^m$ of A. Let $A_m = P_m A P_m = A P_m = P_m A$. By $(E_m(t))_{t\ge0}$ we denote the semigroup generated by $-A_m$, i.e., it is given by the m first terms in the spectral representation (2.17) of $(E(t))_{t\ge0}$.

We denote by X_m^x the solution of equation

$$X_m^x(t) = E_m(t)P_m x + \int_0^t E_m(t-s)P_m f(X_m^x(s))\,\mathrm{d}s + \int_0^t E_m(t-s)P_m g(X_m^x(s))\,\mathrm{d}W(s),\ t\in[0,T].$$

Define the function $u_m(t,x) = \mathbf{E}[G(X_m^x(t))]$ for $t \in [0,T]$ and $x \in H$. Note that $u(t, P_m x) = u(t,x)$ for $x \in H$. It is well known, see e.g. [3, Theorem 9.16], that $u_m\colon [0,T]\times H \to \mathbf{R}$ is a solution to the Kolmogorov equation

$$\begin{aligned} \dot u_m(t,x) + L_m u_m(t,x) &= 0, && (t,x)\in(0,T]\times H,\\ u_m(0,x) &= G(P_m x), && x \in H,\end{aligned}$$

where the Markov generator L_m is given by

$$(L_m v)(x) = \big\langle A_m x - P_m f(x), Dv(x)\big\rangle - \frac12 \operatorname{Tr}\big(P_m g(x) Q g^*(x) P_m D^2 v(x)\big),\quad x\in H.$$

The proof of Theorem 1.1 relies heavily on estimates of the derivatives Du_m and D^2u_m of u_m of the form: for some $\alpha > 0$ we have

$$\sup_{x\in H} \|A^\lambda D u_m(t,x)\| \le C t^{-\lambda} |G|_{C^1_{\mathrm{b}}},\quad t\in(0,T],\ \lambda\in[0,\alpha), \tag{4.1}$$

$$\sup_{x\in H} \|A^\lambda D^2 u_m(t,x) A^\rho\|_{\mathcal{L}} \le C t^{-(\rho+\lambda)} |G|_{C^2_{\mathrm{b}}},\quad t\in(0,T],\ \lambda,\rho\in[0,\alpha),\ \lambda+\rho<1. \tag{4.2}$$

In the case of colored noise it turns out that we need $\alpha \ge (1+\beta)/2$ to obtain convergence of the right rate. So far, to our knowledge, there is no satisfactory result in this direction for multiplicative noise. But for additive colored noise, case **A**, the situation is much easier and the estimates hold for $\alpha = 1$, see Lemma 3.3 in [24]. For the white noise case the estimates are stated as Lemma 4.4 and Lemma 4.5 in [5] with $\alpha = \frac12$. Thus, in case **A** we have $\beta \in [\frac12, 1]$, (4.1) and (4.2) with $\alpha = 1$, and in case **B** we have $\beta \in [0, \frac12)$ and (4.1) and (4.2) with $\alpha = \frac12$.

Since we have the operator A in (4.1) and (4.2) instead of the more natural choice A_m we outline their proofs. We will use that $Du(t,x)\cdot\phi = \mathbf{E}[DG(X_m^x(t))\cdot\eta_m^{\phi,x}(t)]$, where

$$\begin{aligned}\eta_m^{\phi,x}(t) &= E_m(t)P_m\phi + \int_0^t E_m(t-s)P_m Df(X_m^x(s))\cdot\eta_m^{\phi,x}(s)\,\mathrm{d}s\\ &\quad + \int_0^t E_m(t-s)P_m\big(Dg(X_m^x(s))\cdot\eta_m^{\phi,x}(s)\big)\,\mathrm{d}W(s).\end{aligned}$$

In the proofs of Lemma 3.3 in [24] for case **A** with $\alpha = 1$ and Lemma 4.4 in [5] for the case **B** with $\alpha = \frac{1}{2}$ it is proved that

$$\Big(\sup_{x\in H}\mathbf{E}\|\eta_m^{\phi,x}\|^2\Big)^{\frac{1}{2}} \lesssim t^{-\lambda}\|A_m^{-\lambda}P_m\phi\|, \quad t\in(0,T],\ \lambda\in[0,\alpha). \tag{4.3}$$

Therefore

$$\begin{aligned}\langle A^\lambda Du_m(t,x),\psi\rangle &= \langle Du_m(t,x), A^\lambda\psi\rangle = \mathbf{E}[DG(X_m^x(t))\cdot\eta_m^{A^\lambda\psi,x}(t)] \le |G|_{\mathcal{C}_{\mathrm{b}}^1}\big(\mathbf{E}\|\eta_m^{A^\lambda\psi,x}(t)\|^2\big)^{\frac{1}{2}}\\ &\lesssim |G|_{\mathcal{C}_{\mathrm{b}}^1}t^{-\lambda}\|A_m^{-\lambda}P_mA^\lambda\psi\| = |G|_{\mathcal{C}_{\mathrm{b}}^1}t^{-\lambda}\|P_m\psi\| \le |G|_{\mathcal{C}_{\mathrm{b}}^1}t^{-\lambda}\|\psi\|,\end{aligned}$$

implying (4.1).

For (4.2) we notice that

$$D^2u_m(t,x)\cdot(\phi,\psi) = \mathbf{E}[D^2G(X_m^x(t))\cdot(\eta_m^{\phi,x}(t),\eta_m^{\psi,x}(t)) + DG(X_m^x(t))\cdot\zeta_m^{\phi,\psi,x}(t)], \tag{4.4}$$

where

$$\begin{aligned}\zeta_m^{\phi,\psi,x}(t) &= \int_0^t E_m(t-s)P_m\big(D^2f(X_m^x(s))\cdot(\eta_m^{\phi,x}(s),\eta_m^{\psi,x}(s)) + Df(X_m^x(s))\cdot\zeta_m^{\phi,\psi,x}(s)\big)\,\mathrm{d}s\\ &\quad + \int_0^t E_m(t-s)P_m\big(D^2g(X_m^x(s))\cdot(\eta_m^{\phi,x}(s),\eta_m^{\psi,x}(s)) + Dg(X_m^x(s))\cdot\zeta_m^{\phi,\psi,x}(s)\big)\,\mathrm{d}W(s).\end{aligned}$$

In the proof of Lemma 3.3 in [24] for case **A** with $\alpha = 1$ and Lemma 4.5 in [5] for the case **B** with $\alpha = \frac{1}{2}$ it is shown that

$$\Big(\sup_{t\in[0,T]}\sup_{x\in H}\mathbf{E}\|\zeta_m^{\phi,\psi,x}(t)\|^2\Big)^{\frac{1}{2}} \lesssim \|A_m^{-\rho}P_m\phi\|\|A_m^{-\lambda}P_m\psi\|, \quad \lambda,\rho\in[0,\alpha),\ \lambda+\rho<1. \tag{4.5}$$

Since $D^2u_m\cdot(\phi,\psi) = \langle D^2u_m\phi,\psi\rangle$ and by (4.4) and the Cauchy-Schwarz inequality

$$\begin{aligned}\langle A^\lambda D^2u_m(t,x)A^\rho\phi,\psi\rangle &= \langle D^2u_m(t,x)A^\rho\phi, A^\lambda\psi\rangle\\ &= \mathbf{E}\big[D^2G(X_m^x(t))\cdot(\eta_m^{A^\lambda\psi,x}(t),\eta_m^{A^\rho\phi,x}(t)) + DG(X_m^x(t))\cdot\zeta_m^{A^\lambda\psi,A^\rho\phi,x}(t)\big]\\ &\le |G|_{\mathcal{C}_{\mathrm{b}}^2}\big(\mathbf{E}\|\eta_m^{A^\lambda\psi,x}(t)\|^2\big)^{\frac{1}{2}}\big(\mathbf{E}\|\eta_m^{A^\rho\psi,x}(t)\|^2\big)^{\frac{1}{2}} + |G|_{\mathcal{C}_{\mathrm{b}}^1}\big(\mathbf{E}\|\zeta_m^{A^\lambda\psi,A^\rho\phi,x}(t)\|^2\big)^{\frac{1}{2}}\end{aligned}$$

Applying (4.3) and (4.5) yields

$$\begin{aligned}\langle A^\lambda D^2u_m(t,x)A^\rho\phi,\psi\rangle &\lesssim (|G|_{\mathcal{C}_{\mathrm{b}}^2}t^{-\lambda-\rho} + |G|_{\mathcal{C}_{\mathrm{b}}^1})\|A_m^{-\lambda}P_mA^\lambda\phi\|\|A_m^{-\rho}P_mA^\rho\psi\|\\ &\lesssim t^{-\lambda-\rho}\|\phi\|\|\psi\|.\end{aligned}$$

Thus (4.2) is valid.

5. Proof of Theorem 1.1

The error will split into several terms, some of which are common to Assumptions **A** and **B** and some are not. We will first present the proof under Assumption **A**. When doing so we write it as if the noise were multiplicative, i.e., with the operator g included. This will ease the presentation of the white noise case **B**.

5.1. **The case of colored noise.** For an $\mathcal{F}_T$-measurable, H_m-valued random variable ξ, the law of iterated expectation and Proposition 1.12 in [3] yields

$$\mathbf{E}[G(\xi)] = \mathbf{E}[\mathbf{E}[G(\xi)|\mathcal{F}_T]] = \mathbf{E}[\mathbf{E}[G(X^\xi(0))|\mathcal{F}_T]] = \mathbf{E}[u_m(0,\xi)]. \tag{5.1}$$

Thus, the weak error splits into four terms:

$$\begin{aligned}
&\mathbf{E}[G(X(T)) - G(X_h(T))] \\
&\quad = \mathbf{E}[G(X(T)) - G(X_m(T))] + \mathbf{E}[G(X_m(t)) - G(P_m X_h(T))] + \mathbf{E}[G(P_m X_h(T)) - G(X_h(T))] \\
&\quad = \mathbf{E}[G(X(T)) - G(X_m(T))] + u_m(T, X_0) - u_m(T, X_h(0)) \\
&\qquad + \mathbf{E}[u_m(T, X_h(0)) - u_m(0, X_h(T))] + \mathbf{E}[G(P_m X_h(T)) - G(X_h(T))] \\
&\quad = e_1^m(T) + e_2^m(T) + e_3^m(T) + e_4^m(T).
\end{aligned}$$

Our intention is to let $m \to \infty$ and see that the remaining terms is of the right order. The first one is easy to treat since when we let $m \to \infty$ the term $e_1^m(T) \to 0$ by the low order of weak convergence implied by the strong convergence. The second term $e_2^m(T)$ is still easy but needs computations. These holds under both of our assumptions. Using the Cauchy-Schwarz inequality, estimate (4.1) with $0 \le \lambda = \beta - \epsilon < \alpha = 1$, and the error estimate (2.16), we obtain for small $\epsilon > 0$

$$\begin{aligned}
e_2^m(T) = u_m(T, X_0) - u_m(T, P_h X_0) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\lambda} u_m(T, P_h X_0 + \lambda(I - P_h)X_0)\,\mathrm{d}\lambda \\
&= \int_0^1 \Big\langle A^{\beta-\epsilon} D u_m(T, P_h X_0 + \lambda(I - P_h)X_0), P_m A^{-\beta+\epsilon}(I - P_h)X_0 \Big\rangle\,\mathrm{d}\lambda \\
&\le \int_0^1 \|A^{\beta-\epsilon} D u_m(T, P_h X_0 + \lambda(I - P_h)X_0)\|_{\mathcal{L}} \|P_m\|_{\mathcal{L}} \|A^{\epsilon-\beta}(I - P_h)\|_{\mathcal{L}} \|X_0\|\,\mathrm{d}\lambda \\
&\lesssim h^{2\beta-2\epsilon} T^{-\beta} |G|_{C_{\mathrm{b}}^1} \|X_0\| \lesssim h^{2\beta-2\epsilon}, \quad \text{uniformly in } m.
\end{aligned}$$

Here we used that A and P_m commute and that

$$\|A^{\epsilon-\beta}(I - P_h)\|_{\mathcal{L}} \le \|(A^{\epsilon-\beta}(I - P_h))^*\|_{\mathcal{L}} \le \|(I - P_h)A^{\epsilon-\beta}\|_{\mathcal{L}}.$$

Notice here that we could have got a sharp result with $\epsilon = 0$ under Assumption **A**, in the case $\beta < 1$. However, $e_2(T)$ does not allow a sharp rate.

We now turn to the third error term $e_3^m(T)$. For this we need the Markov generator L_h of the finite element solution X_h. It is given by

$$(L_h v)(x) = \big\langle A_h x - P_h f(x), Dv(x) \big\rangle - \frac{1}{2} \operatorname{Tr}\big(P_h g(x) Q g^*(x) P_h D^2 v(x)\big), \quad x \in S_h.$$

Itô's formula and the Kolmogorov equation gives that

$$\begin{aligned}
e_3^m(T) &= -\mathbf{E}[u_m(T - t, X_h(t)) - u_m(T - 0, X_h(0))]\Big|_{t=T} \\
&= -\mathbf{E}\Big[\int_0^T \dot{u}_m(T - t, X_h(t)) + L_h u_m(T - t, X_h(t))\,\mathrm{d}t\Big] \\
&= \mathbf{E}\int_0^T (L_m - L_h) u_m(T - t, X_h(t))\,\mathrm{d}t.
\end{aligned}$$

The error $e_3^m(T)$ now naturally divides into three terms:

$$\begin{aligned}
|e_3^m(T)| \le &\Big|\mathbf{E}\int_0^T \Big\langle (A_m - A_h) X_h(t), D u_m(T - t, X_h(t)) \Big\rangle\,\mathrm{d}t\Big| \\
&+ \Big|\mathbf{E}\int_0^T \Big\langle (P_m - P_h) f(X_h(t)), D u_m(T - t, X_h(t)) \Big\rangle\,\mathrm{d}t\Big| \\
&+ \Big|\frac{1}{2}\mathbf{E}\int_0^T \operatorname{Tr}\Big\{\Big[P_m g(X_h(t)) Q g^*(X_h(t)) P_m - P_h g(X_h(t)) Q g^*(X_h(t)) P_h\Big] \\
&\qquad \times D^2 u_m(T - t, X_h(t))\Big\}\,\mathrm{d}t\Big| \\
= &\; I + J + K.
\end{aligned}$$

The Ritz projection R_h can be expressed in the form $R_h = A_h^{-1} P_h A$. Observing this we can write

$$\begin{aligned}\langle (A_m - A_h)X_h, Du_m\rangle &= \langle (A_m P_h - P_h A_h)X_h, Du_m\rangle = \langle X_h, (P_h A_m - A_h P_h) Du_m\rangle \\ &= \langle X_h, A_h P_h (A_h^{-1} P_h A_m - I) Du_m\rangle = \langle X_h, A_h P_h (A_h^{-1} P_h A P_m - I) Du_m\rangle \\ &= \langle X_h, A_h P_h (R_h - I) P_m Du_m\rangle + \langle X_h, A_h P_h (P_m - I) Du_m\rangle.\end{aligned}$$

This enables us to rewrite the term I so that we can apply the error estimates (2.15) and (2.19) for R_h and P_m respectively. We substitute for X_h the mild equation (1.2) and treat the terms separately and estimate

$$\begin{aligned} I &\le \Big| \mathbf{E} \int_0^T \Big\langle E_h(t) P_h X_0, A_h P_h (R_h - I) P_m Du_m(T-t, X_h(t)) \Big\rangle \, \mathrm{d}t \Big| \\ &\quad + \Big| \mathbf{E} \int_0^T \Big\langle \int_0^t E_h(t-s) P_h f(X_h(s)) \, \mathrm{d}s, A_h P_h (R_h - I) P_m Du_m(T-t, X_h(t)) \Big\rangle \, \mathrm{d}t \Big| \\ &\quad + \Big| \mathbf{E} \int_0^T \Big\langle \int_0^t E_h(t-s) P_h g(X_h(s)) \, \mathrm{d}W(s), A_h P_h (R_h - I) P_m Du_m(T-t, X_h(t)) \Big\rangle \, \mathrm{d}t \Big| \\ &\quad + \Big| \mathbf{E} \int_0^T \Big\langle A_h X_h, (P_m - I) Du_m(T-t, X_h(t)) \Big\rangle \, \mathrm{d}t \Big| \\ &= I_1^h + I_2^h + I_3^h + I^m. \end{aligned}$$

We now estimate I_1^h. Let $\epsilon > 0$ be small. Using (2.15), (2.13), (2.18), and (4.1) yields

$$\begin{aligned} I_1^h &= \Big| \mathbf{E} \int_0^T \Big\langle A_h^{1-\epsilon} E_h(t) P_h X_0, A_h^{\epsilon} P_h (R_h - I) A^{-\beta+\epsilon} P_m A^{\beta-\epsilon} Du_m(T-t, X_h(t)) \Big\rangle \, \mathrm{d}t \Big| \\ &\le \mathbf{E} \int_0^T \| A_h^{1-\epsilon} E_h(t) P_h \|_{\mathcal{L}} \|X_0\| \| A_h^{\epsilon} P_h (R_h - I) A^{-\beta+\epsilon} \|_{\mathcal{L}} \|P_m\|_{\mathcal{L}} \\ &\qquad \times \sup_{x \in H} \| A^{\beta-\epsilon} Du_m(T-t, x) \| \, \mathrm{d}t \\ &\lesssim h^{2\beta-4\epsilon} \int_0^T t^{-1+\epsilon} (T-t)^{-\beta+\epsilon} \, \mathrm{d}t \, |G|_{\mathcal{C}_{\mathrm{b}}^1} \|X_0\| \lesssim h^{2\beta-4\epsilon}. \end{aligned}$$

The term I_2^h is easily estimated as follows:

$$\begin{aligned} I_2^h &= \Big| \mathbf{E} \int_0^T \Big\langle \int_0^t A_h^{1-\epsilon} E_h(t-s) P_h f(X_h(s)) \, \mathrm{d}s, \\ &\qquad A_h^{\epsilon} P_h (R_h - I) A^{-\beta+\epsilon} P_m A^{\beta-\epsilon} Du_m(T-t, X_h(t)) \Big\rangle \, \mathrm{d}t \Big| \\ &\le \int_0^T \int_0^t \| A_h^{1-\epsilon} E_h(t-s) P_h \|_{\mathcal{L}} \big(\mathbf{E} \| f(X_h(s)) \|^2 \big)^{\frac{1}{2}} \\ &\qquad \times \| A_h^{\epsilon} P_h (R_h - I) A^{-\beta+\epsilon} \|_{\mathcal{L}} \|P_m\|_{\mathcal{L}} \sup_{x \in H} \| A^{\beta-\epsilon} Du_m(T-t, x) \| \, \mathrm{d}s \, \mathrm{d}t. \end{aligned}$$

Using (2.15), (2.13), (2.18) and (3.2) yields

$$I_2^h \lesssim h^{2\beta-4\epsilon} \int_0^T \int_0^t (T-t)^{-\beta+\epsilon} (t-s)^{-1+\epsilon} \, \mathrm{d}s \, \mathrm{d}t \lesssim h^{2\beta-4\epsilon}.$$

For I_3 we use the Malliavin integration by parts formula from Lemma 2.2 together with the chain rule (2.25) to obtain the error representation

$$\begin{aligned} I_3^h &= \Big| \mathbf{E} \int_0^T \Big\langle \int_0^t E_h(t-s) P_h g(X_h(s)) \, \mathrm{d}W(s), \ A_h P_h (R_h - I) P_m Du_m(T-t, X_h(t)) \Big\rangle \, \mathrm{d}t \Big| \\ &= \Big| \mathbf{E} \int_0^T \int_0^t \Big\langle E_h(t-s) P_h g(X_h(s)), \\ &\qquad A_h P_h (R_h - I) P_m D^2 u_m(T-t, X_h(t)) P_m D_s X_h(t) \Big\rangle_{\mathcal{L}_2^0} \, \mathrm{d}s \, \mathrm{d}t \Big|. \end{aligned}$$

Here we treat Assumptions **A** and **B** separately and start with **A**; **B** is postponed to the next subsection. Distributing powers of A and A_h carefully and setting $g(x) = I$, we write

$$\begin{aligned}\big\langle E_h P_h, A_h P_h (R_h - I) P_m D^2 u_m \, P_m D_s X_h \big\rangle_{\mathcal{L}_2^0} &= \big\langle A_h^{\frac{1+\beta}{2}-\epsilon} E_h A_h^{\frac{1-\beta}{2}} A_h^{\frac{\beta-1}{2}} P_h, \\ &A_h^{\frac{1-\beta}{2}+\epsilon} P_h (R_h - I) A^{-\frac{1+\beta}{2}+\epsilon} P_m A^{\frac{1+\beta}{2}-\epsilon} D^2 u_m \, A^{\frac{1-\beta}{2}} P_m A^{\frac{\beta-1}{2}} D_s X_h \big\rangle_{\mathcal{L}_2^0}.\end{aligned}$$

Using the Cauchy-Schwarz inequality for $\mathcal{L}_2^0$ and (2.7) yields

$$\begin{aligned}I_3^h \leq \mathbf{E} \int_0^T \int_0^t &\|A_h^{1-\epsilon} E_h(t-s) P_h\|_{\mathcal{L}} \|A_h^{\frac{\beta-1}{2}} P_h\|_{\mathcal{L}_2^0} \|A_h^{\frac{1-\beta}{2}+\epsilon} P_h (R_h - I) A^{-\frac{1+\beta}{2}+\epsilon}\|_{\mathcal{L}} \|P_m\|_{\mathcal{L}} \\ &\times \sup_{x \in H} \|A^{\frac{1+\beta}{2}-\epsilon} D^2 u_m (T-t, x) A^{\frac{1-\beta}{2}}\|_{\mathcal{L}} \|A^{\frac{\beta-1}{2}} D_s X_h(t)\|_{\mathcal{L}_2^0} \,\mathrm{d}s \,\mathrm{d}t.\end{aligned}$$

We use (2.13) to get $\|A_h^{\frac{\beta-1}{2}} P_h\|_{\mathcal{L}_2^0} \lesssim \|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0}$. The norm equivalence (2.14) and the fact that $D_s^u X_h(t) \in S_h$, $\mathbf{P}$-a.s., for every $u \in U_0$ yields

$$\|A^{\frac{\beta-1}{2}} D_s X_h(t)\|_{\mathcal{L}_2^0} \lesssim \|A_h^{\frac{\beta-1}{2}} D_s X_h(t)\|_{\mathcal{L}_2^0}.$$

The analyticity of the semigroup (2.18), the error estimate (2.15) together with (2.13), the gradient estimate (4.2), Tonelli's theorem and the Cauchy-Schwarz inequality now imply that

$$I_3^h \lesssim h^{2\beta-4\epsilon} |G|_{C_{\mathrm{b}}^2} \|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0} \int_0^T \int_0^t \big(\mathbf{E} \|A_h^{\frac{\beta-1}{2}} D_s X_h(t)\|_{\mathcal{L}_2^0}^2\big)^{\frac{1}{2}} (T-t)^{-1+\epsilon} (t-s)^{-1+\epsilon} \,\mathrm{d}s \,\mathrm{d}t.$$

Applying Lemma 3.1 we finally get

$$I_3^h \lesssim h^{2\beta-4\epsilon} \int_0^T \int_0^t (T-t)^{-1+\epsilon} (t-s)^{-1+\epsilon} \,\mathrm{d}s \,\mathrm{d}t \lesssim h^{2\beta-4\epsilon}.$$

Using (2.19), (4.1), the Cauchy-Schwarz inequality and (2.27) yields

$$\begin{aligned}I^m &\leq \mathbf{E} \int_0^T \|A_h X_h(t)\|_{\mathcal{L}} \|(P_m - I) A^{-\frac{1}{2}+\epsilon}\| \sup_{x \in H} \|A^{\frac{1}{2}-\epsilon} D u_m (T-t, x)\|_{\mathcal{L}} \,\mathrm{d}t \\ &\lesssim \lambda_m^{-\frac{1}{2}+\epsilon} \|A_h P_h\|_{\mathcal{L}} \Big(\sup_{t \in [0,T]} \mathbf{E} \|X_h(t)\|^2 \Big)^{\frac{1}{2}} \int_0^T (T-t)^{-\frac{1}{2}+\epsilon} \,\mathrm{d}t.\end{aligned}$$

Letting $m \to \infty$ for fixed h yields $I^m \to 0$ and $\lim_{m\to\infty} I \lesssim h^{2\beta-4\epsilon}$.

The term J is considered next. Writing $P_m - P_h = (P_m - I) + (I - P_h)$ we get the natural decomposition $J \leq J^m + J^h$. Using the Cauchy-Schwarz inequality, the error estimate (2.16), (4.1), and (3.2) yields for $i \in \{h, m\}$

$$\begin{aligned}J^i &= \Big| \mathbf{E} \int_0^T \Big\langle (I - P_i) D u_m (T-t, P_m X_h(t)), f(X_h(t)) \Big\rangle \,\mathrm{d}s \Big| \\ &\leq \int_0^T \|(I - P_i) A^{-\beta+\epsilon}\|_{\mathcal{L}} \sup_{x \in H^m} \|A^{\beta-\epsilon} D u_m (T-t, x)\| \big(\mathbf{E} \|f(X_h(t))\|^2 \big)^{\frac{1}{2}} \,\mathrm{d}t \\ &\lesssim \|(I - P_i) A^{-\beta+\epsilon}\|_{\mathcal{L}} |G|_{\mathcal{C}_{\mathrm{b}}^1} \int_0^T (T-t)^{-\beta+\epsilon} \,\mathrm{d}t.\end{aligned}$$

We have

$$J^h \lesssim h^{2\beta-2\epsilon}, \quad \text{and} \quad J^m \lesssim \lambda_m^{-\beta+\epsilon}.$$

For K we write

$$\begin{aligned}&P_m g Q g^* P_m - P_h g Q g^* P_h \\ &\quad = P_h g Q g^* (I - P_h) + (I - P_h) g Q g^* P_m + (P_m + P_h) g Q g^* (P_m - I),\end{aligned}$$

and hence we get the following decomposition:

$$\begin{aligned}
2K &= \Big|\mathbf{E}\int_0^T \mathrm{Tr}\,\Big(\big[P_m g(X_h(t))Qg^*(X_h(t))P_m - P_h g(X_h(t))Qg^*(X_h(t))P_h\big] \\
&\qquad \times D^2u_m(T-t,P_mX_h(t))\Big)\,\mathrm{d}t\Big| \\
&\le \Big|\mathbf{E}\int_0^T \mathrm{Tr}\,\Big(P_h g(X_h(t))Qg^*(X_h(t))(I-P_h)D^2u_m(T-t,P_mX_h(t))\Big)\,\mathrm{d}t\Big| \\
&\quad + \Big|\mathbf{E}\int_0^T \mathrm{Tr}\,\Big((I-P_h) g(X_h(t))Qg^*(X_h(t))P_mD^2u_m(T-t,P_mX_h(t))\Big)\,\mathrm{d}t\Big| \\
&\quad + \Big|\mathbf{E}\int_0^T \mathrm{Tr}\,\Big((P_m+P_h) g(X_h(t))Qg^*(X_h(t))(P_m-I)D^2u_m(T-t,P_mX_h(t))\Big)\,\mathrm{d}t\Big| \\
&= K_1^h + K_2^h + K^m.
\end{aligned}$$

Assumption **A** is treated first; **B** is postponed. By (2.3), (2.2) and (2.7), we have

$$\begin{aligned}
&\mathrm{Tr}(P_hQ(I-P_h)D^2u_m) \\
&\quad = \mathrm{Tr}(P_hQ(I-P_h)D^2u_mA^{\frac{1-\beta}{2}}A^{\frac{\beta-1}{2}}) = \mathrm{Tr}(A^{\frac{\beta-1}{2}}P_hQ(I-P_h)D^2u_mA^{\frac{1-\beta}{2}}) \\
&\quad = \mathrm{Tr}(A^{\frac{\beta-1}{2}}P_hA^{\frac{1-\beta}{2}}A^{\frac{\beta-1}{2}}QA^{\frac{\beta-1}{2}}A^{\frac{1-\beta}{2}}(I-P_h)A^{-\frac{1+\beta}{2}+\epsilon}A^{\frac{1+\beta}{2}-\epsilon}D^2u_mA^{\frac{1-\beta}{2}}) \\
&\quad \le \|A^{\frac{\beta-1}{2}}P_hA^{\frac{1-\beta}{2}}\|_{\mathcal{L}}\|A^{\frac{\beta-1}{2}}\|^2_{\mathcal{L}_2^0}\|A^{\frac{1-\beta}{2}}(I-P_h)A^{-\frac{1+\beta}{2}+\epsilon}\|_{\mathcal{L}}\|A^{\frac{1+\beta}{2}-\epsilon}D^2u_mA^{\frac{1-\beta}{2}}\|_{\mathcal{L}},
\end{aligned}$$

where we used the fact that

$$\|A^{\frac{\beta-1}{2}}QA^{\frac{\beta-1}{2}}\|_{\mathcal{L}_1} = \mathrm{Tr}((A^{\frac{\beta-1}{2}}Q^{\frac12})(A^{\frac{\beta-1}{2}}Q^{\frac12})^*) = \|A^{\frac{\beta-1}{2}}Q^{\frac12}\|^2_{\mathcal{L}_2} = \|A^{\frac{\beta-1}{2}}\|^2_{\mathcal{L}_2^0}.$$

By (2.14) and (2.13) $\|A^{\frac{\beta-1}{2}}P_hA^{\frac{1-\beta}{2}}\|_{\mathcal{L}} \lesssim \|A_h^{\frac{\beta-1}{2}}P_hA^{\frac{1-\beta}{2}}\|_{\mathcal{L}} \lesssim \|A^{\frac{\beta-1}{2}}A^{\frac{1-\beta}{2}}\|_{\mathcal{L}} = 1$. Using (2.16), (3.2) and (4.2) gives us

$$K_1^h \lesssim h^{2\beta-2\epsilon}\|A^{\frac{\beta-1}{2}}\|^2_{\mathcal{L}_2^0}\,|G|_{\mathcal{C}_{\mathrm{b}}^2}\int_0^T (T-t)^{-1+\epsilon}\,\mathrm{d}t \lesssim h^{2\beta-2\epsilon}.$$

For K_2^h we compute similarly

$$\begin{aligned}
\mathrm{Tr}((I-P_h)QP_mD^2u) &= \mathrm{Tr}(A^{-\frac{1+\beta}{2}+\epsilon}(I-P_h)A^{\frac{1-\beta}{2}}A^{\frac{\beta-1}{2}}QA^{\frac{\beta-1}{2}}A^{\frac{1-\beta}{2}}D^2u_mA^{\frac{1+\beta}{2}-\epsilon}) \\
&\le \|A^{-\frac{1+\beta}{2}+\epsilon}(I-P_h)A^{\frac{1-\beta}{2}}\|_{\mathcal{L}}\|A^{\frac{\beta-1}{2}}\|^2_{\mathcal{L}_2^0}\|A^{\frac{1-\beta}{2}}D^2u_mA^{\frac{1+\beta}{2}-\epsilon}\|_{\mathcal{L}},
\end{aligned}$$

where

$$\|A^{-\frac{1+\beta}{2}+\epsilon}(I-P_h)A^{\frac{1-\beta}{2}}\|_{\mathcal{L}} \le \|(A^{-\frac{1+\beta}{2}+\epsilon}(I-P_h)A^{\frac{1-\beta}{2}})^*\|_{\mathcal{L}} = \|A^{\frac{1-\beta}{2}}(I-P_h)A^{-\frac{1+\beta}{2}+\epsilon}\|_{\mathcal{L}},$$

so that (2.16) applies. Hence,

$$K_2^h \lesssim h^{2\beta-2\epsilon}\|A^{\frac{\beta-1}{2}}\|^2_{\mathcal{L}_2^0}\,|G|_{C_{\mathrm{b}}^2}\int_0^T (T-t)^{-1+\epsilon}\,\mathrm{d}t \lesssim h^{2\beta-2\epsilon}.$$

Term K^m is treated analogously as K_1^h. We have $K^m \lesssim \lambda_m^{-\beta+\epsilon}$.

Finally, by the Lipschitz continuity of G, the Dominated Convergence Theorem and the strong convergence of $P_m \to I$ we get

$$e_4^m(T) \le |G|_{\mathcal{C}_{\mathrm{b}}^1}\mathbf{E}\|(P_m-I)X_h(t)\| \to 0, \quad \text{as } m\to\infty.$$

We conclude that $e(T) = O(h^{2\gamma})$ for any $\gamma < \beta$, which completes the proof in case **A**.

5.2. **The case of white noise.** Now consider the case of Assumption **B**. All estimates above, except those for I_3^h, K_1^h, K_2^h and K^m, hold under **B** by setting $Q = I$ and $\beta = \frac{1}{2}$ and recalling that $U_0 = H$ and $\mathcal{L}_2^0 = \mathcal{L}_2$. We now complete the proof with the remaining estimates.

Using Hölder's inequality (2.8) yields

$$\begin{aligned} I_3^h &= \Big| \mathbf{E} \int_0^T \int_0^t \Big\langle E_h(t-s) P_h g(X_h(s)), A_h P_h (R_h - I) \\ &\qquad \times P_m D^2 u_m(T-t, P_m X_h(t)) P_m D_s X_h(t) \Big\rangle_{\mathcal{L}_2} \mathrm{d}s \, \mathrm{d}t \Big| \\ &\le \mathbf{E} \int_0^T \int_0^t \|A_h^{1-3\epsilon} E_h(t-s) P_h\|_{\mathcal{L}} \|g(X_h(s))\|_{\mathcal{L}} \|A_h^{3\epsilon} P_h (R_h - I) A^{-\frac{1}{2}+\epsilon}\|_{\mathcal{L}} \\ &\qquad \times \sup_{x \in H^m} \|A^{\frac{1}{2}-\epsilon} D^2 u_m(T-t, x) A^{\frac{1}{2}-\epsilon}\|_{\mathcal{L}} \|A^{-\frac{1}{2}+\epsilon} A_h^{-2\epsilon} P_h\|_{\mathcal{L}_1} \|A_h^{2\epsilon} D_s X_h(t)\|_{\mathcal{L}} \, \mathrm{d}s \, \mathrm{d}t. \end{aligned}$$

First, using (2.13) and (2.9), we have

$$\|A^{-\frac{1}{2}+\epsilon} A_h^{-2\epsilon} P_h\|_{\mathcal{L}_1} = \|A_h^{-2\epsilon} P_h A^{-\frac{1}{2}+\epsilon}\|_{\mathcal{L}_1} \lesssim \|A_h^{-2\epsilon} P_h A^{2\epsilon}\|_{\mathcal{L}} \|A^{-\frac{1}{2}-\epsilon}\|_{\mathcal{L}_1} \lesssim \|A^{-\frac{1}{2}-\epsilon}\|_{\mathcal{L}_1}.$$

Now we apply (2.15), (2.18), (4.2) with $\rho = \lambda = \frac{1}{2} - \epsilon < \alpha = \frac{1}{2}$, to get

$$I_3^h \lesssim h^{1-8\epsilon} |G|_{\mathcal{C}_{\mathrm{b}}^2} \int_0^T \int_0^t \big(\mathbf{E}\|g(X_h(s))\|_{\mathcal{L}}^2\big)^{\frac{1}{2}} \big(\mathbf{E}\|A_h^{2\epsilon} D_s X_h(t)\|_{\mathcal{L}}^2\big)^{\frac{1}{2}} (T-t)^{-1+2\epsilon} (t-s)^{-1+3\epsilon} \, \mathrm{d}s \, \mathrm{d}t.$$

Finally using Lemma 3.2 together with (3.2) finishes the estimate of I_3^h. Indeed,

$$I_3^h \lesssim h^{1-8\epsilon} |G|_{C_{\mathrm{b}}^2} \int_0^T (T-t)^{-1+2\epsilon} (t-s)^{-1+\epsilon} \, \mathrm{d}s \, \mathrm{d}t \lesssim h^{1-8\epsilon}.$$

For K_1 we use Hölder's inequality (2.6), (2.16), and Lemma 4.2 to get

$$\begin{aligned} 2K_1 &\le \int_0^T \mathbf{E} \|A^{-\frac{1-\epsilon}{2}} P_h g(X_h(t)) g^*(X_h(t)) A^{-\epsilon}\|_{\mathcal{L}_1} \|A^{\epsilon}(I - P_h) A^{-\frac{1-\epsilon}{2}}\|_{\mathcal{L}} \\ &\qquad \times \sup_{x \in H} \|A^{\frac{1-\epsilon}{2}} D^2 u_m(T-t, x) A^{\frac{1-\epsilon}{2}}\|_{\mathcal{L}} \, \mathrm{d}t \\ &\lesssim h^{1-3\epsilon} \sup_{t \in [0,T]} \mathbf{E} \|A^{-\frac{1-\epsilon}{2}} P_h g(X_h(t)) g^*(X_h(t))\|_{\mathcal{L}_{2/(2-3\epsilon)}} \|A^{-\epsilon}\|_{\mathcal{L}_{2/3\epsilon}} |G|_{\mathcal{C}_{\mathrm{b}}^2} \int_0^T (T-t)^{-1+\epsilon} \, \mathrm{d}t \\ &\lesssim h^{1-3\epsilon} \sup_{t \in [0,T]} \mathbf{E} \|g(X_h(t))\|_{\mathcal{L}}^2 \|A^{-\frac{1-\epsilon}{2}}\|_{\mathcal{L}_{2/(2-3\epsilon)}} \|A^{-\epsilon}\|_{\mathcal{L}_{2/3\epsilon}} |G|_{\mathcal{C}_{\mathrm{b}}^2}. \end{aligned}$$

We compute and use (2.9) to conclude

$$\begin{aligned} \|A^{-\epsilon}\|_{\mathcal{L}_{2/3\epsilon}}^{3\epsilon/2} &= \sum_{i \in \mathbf{N}} (\lambda_i^{-\epsilon})^{\frac{2}{3\epsilon}} = \sum_{i \in \mathbf{N}} \lambda_i^{-\frac{2}{3}} = \mathrm{Tr}(A^{-\frac{2}{3}}) < \infty \\ \|A^{-\frac{1-\epsilon}{2}}\|_{\mathcal{L}_{2/(2-3\epsilon)}}^{(2-3\epsilon)/2} &= \sum_{i \in \mathbf{N}} (\lambda_i^{-\frac{1-\epsilon}{2}})^{\frac{2}{2-3\epsilon}} = \sum_{i \in \mathbf{N}} \lambda_i^{-\frac{1-\epsilon}{2-3\epsilon}} = \mathrm{Tr}\left(A^{-\frac{1}{2}\left(\frac{2-2\epsilon}{2-3\epsilon}\right)}\right) < \infty. \end{aligned}$$

The terms K_2^h and K^m admits the same treatment, so that

$$K_2^h \lesssim h^{1-3\epsilon}, \quad \text{and} \quad K^m \lesssim \lambda_m^{-\frac{1}{2}+\frac{3\epsilon}{2}}.$$

We have $e(T) = \mathcal{O}(h^{2\gamma})$ for any $\gamma < \frac{1}{2}$.

References

[1] C. E. Brehier, *Approximation of the invariant measure with an Euler scheme for stochastic PDE's driven by space-time white noise*, preprint (2012).

[2] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, third ed., Texts in Applied Mathematics, vol. 15, Springer, New York, 2008.

[3] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Encyclopedia of Mathematics and its Applications, vol. 44, Cambridge University Press, Cambridge, 1992.

[4] A. de Bouard and A. Debussche, *Weak and strong order of convergence of a semidiscrete scheme for the stochastic nonlinear Schrödinger equation*, Appl. Math. Optim. **54** (2006), 369–399.

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

[6] A. Debussche and J. Printems, *Weak order for the discretization of the stochastic heat equation*, Math. Comp. **78** (2009), 845–863.

[7] N. Dunford and J. T. Schwartz, *Linear Operators. Part II*, Wiley Classics Library, John Wiley & Sons Inc., New York, 1988, Spectral theory. Selfadjoint operators in Hilbert space, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1963 original, A Wiley-Interscience Publication.

[8] A. Grorud and É. Pardoux, *Intégrales hilbertiennes anticipantes par rapport à un processus de Wiener cylindrique et calcul stochastique associé*, Appl. Math. Optim. **25** (1992), 31–49.

[9] E. Hausenblas, *Weak approximation for semilinear stochastic evolution equations*, Stochastic analysis and related topics VIII, Progr. Probab., vol. 53, Birkhäuser, Basel, 2003, pp. 111–128.

[10] ———, *Weak approximation of the stochastic wave equation*, J. Comput. Appl. Math. **235** (2010), 33–58.

[11] S. Janson, *Gaussian Hilbert spaces*, Cambridge Tracts in Mathematics, vol. 129, Cambridge University Press, Cambridge, 1997.

[12] A. Jentzen and M. Röckner, *Regularity analysis for stochastic partial differential equations with nonlinear multiplicative trace class noise*, J. Differential Equations **252** (2012), 114–136.

[13] 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 (2012), `doi:10.1007/s10543-012-0405-1`.

[14] Mihály Kovács, Stig Larsson, and Fredrik Lindgren, *Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise*, BIT **52** (2012), 85–108.

[15] R. Kruse, *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*, PhD thesis, University of Bielefeld, 2012.

[16] ———, *Optimal error estimates of Galerkin finite element methods for stochastic partial differential equations with multiplicative noise*, preprint (2011).

[17] R. Kruse and S. Larsson, *Optimal regularity for semilinear stochastic partial differential equations with multiplicative noise*, Electron. J. Probab. **65** (2012), 1–19.

[18] J. A. León and D. Nualart, *Stochastic evolution equations with random generators*, Ann. Probab. **26** (1998), 149–186.

[19] F. Lindner and R. L. Schilling, *Weak order for the discretization of the stochastic heat equation driven by impulsive noise*, Potential Anal. (2012), `doi:10.1007/s11118-012-9276-y`.

[20] Alessandra Lunardi, *Interpolation Theory*, second ed., Appunti. Scuola Normale Superiore di Pisa (Nuova Serie). [Lecture Notes. Scuola Normale Superiore di Pisa (New Series)], Edizioni della Normale, Pisa, 2009.

[21] D. Nualart, *The Malliavin calculus and related topics*, second ed., Probability and its Applications (New York), Springer-Verlag, Berlin, 2006.

[22] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, second ed., Springer Series in Computational Mathematics, vol. 25, Springer-Verlag, Berlin, 2006.

[23] J. M. A. M. van Neerven, *Stochastic Evolution Equations*, ISEM lecture notes (2008).

[24] X. Wang and S. Gan, *Weak convergence analysis of the linear implicit euler method for semilinear stochastic partial differential equations with additive noise*, J. Math. Anal. Appl. (2012).