Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics

Duality in refined Watanabe-Sobolev spaces and weak approximations of SPDE

A. Andersson and R. Kruse and S. Larsson

Research Report No. 2013-50
December 2013

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

DUALITY IN REFINED WATANABE-SOBOLEV SPACES AND WEAK APPROXIMATIONS OF SPDE

ADAM ANDERSSON, RAPHAEL KRUSE, AND STIG LARSSON

Abstract. In this paper we introduce a new family of refined Watanabe-Sobolev spaces that capture in a fine way integrability in time of the Malliavin derivative. We consider duality in these spaces and derive a Burkholder type inequality in a dual norm.

The theory we develop allows us to prove weak convergence with essentially optimal rate for numerical approximations in space and time of semilinear parabolic stochastic evolution equations driven by Gaussian additive noise. In particular, we combine Galerkin finite element methods with a backward Euler scheme in time. The method of proof does not rely on the use of the Kolmogorov equation or the Itō formula and is therefore in nature non-Markovian. With this method polynomial growth test functions with mild smoothness assumptions are allowed, meaning in particular that we prove convergence of arbitrary moments with essentially optimal rate. Our Gronwall argument also yields weak error estimates which are uniform in time without any additional effort.

1. Introduction

The classical Watanabe-Sobolev spaces capture the integrability in the chance parameter of the random variable and its Malliavin derivatives. In the Malliavin calculus for stochastic evolution equations the Malliavin derivative is a stochastic process. One purpose of this paper is to introduce a refined family of Watanabe-Sobolev spaces that have the ability to capture the precise integrability properties of the Malliavin derivative with respect to its time parameter. It turns out that the Malliavin derivative of the solution to a parabolic stochastic evolution equation has, depending on the regularity of the noise, good integrability properties in time and, in the case of trace class noise, it is even bounded. The main benefit with narrower spaces is that the corresponding dual spaces have good properties.

Let $(H, \|\cdot\|, \langle\cdot,\cdot\rangle)$ be a separable Hilbert space and $Q \in \mathcal{L}(H)$ be a selfadjoint positive semidefinite linear operator on H. We define the space $H_0 = Q^{\frac{1}{2}}(H)$ and let $\mathcal{L}_2^0 = \mathcal{L}_2(H, H_0)$ be the space of Hilbert-Schmidt operators $H_0 \to H$. We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbf{P})$ on which an $L^2([0,T], H_0)$-isonormal process is defined. For a differentiable random variable X the Malliavin derivative $DX = (D_tX)_{t\in[0,T]}$ with respect to the isonormal process, is an $\mathcal{L}_2^0$-valued stochastic process. We introduce for $p, q \geq 2$, the refined Watanabe-Sobolev

2010 *Mathematics Subject Classification.* 60H15, 65C30, 65M60, 65M70.

Key words and phrases. SPDE, finite element method, backward Euler, weak convergence, convergence of moments, Malliavin calculus, duality, spatio-temporal discretization.

spaces $\mathbf{M}^{1,p,q}(H)$ of random variables $X \in L^2(\Omega, H)$ defined by the norms

$$\|X\|_{\mathbf{M}^{1,p,q}(H)} = \Big(\|X\|^p_{L^p(\Omega,H)} + \|DX\|^p_{L^p(\Omega,L^q([0,T],\mathcal{L}_2^0))} \Big)^{\frac{1}{p}}.$$

The classical Watanabe-Sobolev spaces are obtained for $q = 2$. We use the refined spaces in a duality argument based on the Gelfand triple

$$\mathbf{M}^{1,p,q}(H) \subset L^2(\Omega, H) \subset \mathbf{M}^{1,p,q}(H)^*.$$

A key ingredient is the following inequality for the H-valued stochastic Itō-integral $\int_0^T \Phi \, \mathrm{d}W$ in the dual norm of $\mathbf{M}^{1,p,q}(H)$, where W is a cylindrical Q-Wiener process and $\Phi \in L^p(\Omega, L^2([0,T], \mathcal{L}_2^0))$ is a predictable stochastic process. In Theorem 3.5 we show

$$\Big\| \int_0^T \Phi(t) \, \mathrm{d}W(t) \Big\|_{\mathbf{M}^{1,p,q}(H)^*} \le \big\|\Phi\big\|_{L^{p'}(\Omega,L^{q'}([0,T],\mathcal{L}_2^0))}, \tag{1.1}$$

where p', q' are the conjugate exponents to $p, q \ge 2$. This inequality is applied in situations, where one usually relies on the Burkholder-Davis-Gundy inequality, see Lemma 2.2. In this case the $L^2(\Omega, H)$-norm of the stochastic integral is bounded in terms of the $L^p(\Omega, L^2([0,T], \mathcal{L}_2^0))$-norm of Φ. Here, the dual norm of the integral is bounded by the $L^{p'}(\Omega, L^{q'}([0,T], \mathcal{L}_2^0))$-norm of Φ and, as $p', q' \le 2$, this admits stronger singularities.

In defining the $\mathbf{M}^{1,p,q}(H)$-spaces care needs to be taken. For $q \ge 2$ the Malliavin derivative is defined on a non-standard core $\mathcal{S}^q(H)$ of smooth and cylindrical random variables, more regular than in the classical theory in which $q = 2$. By proving that the operator $D \colon \mathcal{S}^q(H) \to L^p(\Omega, L^q([0,T], \mathcal{L}_2^0))$ is well defined and closable we show that $\mathbf{M}^{1,p,q}(H)$ are Banach spaces. The proofs are rather elementary and rely to a large extent on existing results for the case $q = 2$. The spaces are new to the best of our knowledge, although there are similarities with the Hida and Kondratiev spaces, see, e.g., [20] or [4].

The motivation for introducing the spaces described above is found in our aim to develop new methods for the analysis of the weak error of the numerical approximation of semilinear stochastic partial differential equations of the form

$$\mathrm{d}X(t) + AX(t) \, \mathrm{d}t = F(X(t)) \, \mathrm{d}t + \mathrm{d}W(t), \ t \in (0,T]; \quad X(0) = X_0. \tag{1.2}$$

See Assumption 2.3 below for precise conditions on A, F, W, X_0. We treat discretizations in space and time, allowing for any spatial discretization scheme that satisfies the abstract Assumption 2.4 below. We verify this assumption for piecewise linear finite element approximations and spectral Galerkin approximations of the heat equation. Discretization in time is performed by the backward Euler method.

Weak convergence for linear stochastic evolution equations has been studied in the papers [16], [25], [26], [13], [11], [28], [32] and the works [40], [18], [19], [12], [5], [6], [7] [2] treat semilinear equations. Most of these works are based on Itō's formula and Kolmogorov's equation. It becomes apparent while reading the literature that proving weak convergence with optimal order is a challenging task. Various restrictive assumptions are imposed on the noise in all these works. In particular, multiplicative noise more general than linear has rarely been treated due to severe difficulties, and in this case only for space time white noise, see [12], [2].

Let $X, Y \in L^2(\Omega, H)$ and $\varphi\colon H \to \mathbf{R}$ have two continuous Fréchet derivatives of polynomial growth. Our technique relies on the following linearization of the weak error

$$\mathbf{E}\big[\varphi(X) - \varphi(Y)\big] = \mathbf{E}\big[\langle \tilde{\varphi}, X - Y\rangle\big], \quad \text{where} \quad \tilde{\varphi} = \int_0^1 \varphi'(\varrho X + (1-\varrho)Y)\,\mathrm{d}\varrho,$$

introduced in [8] and [28] independently. The paper [8] then proceeds by using an adjoint problem. Our method is the following: If $V \subset L^2(\Omega, H) \subset V^*$ is a Gelfand triple such that $\tilde{\varphi} \in V$, then we obtain by duality

$$\big|\mathbf{E}\big[\varphi(X) - \varphi(Y)\big]\big| \le \|\tilde{\varphi}\|_V \|X - Y\|_{V^*}.$$

With a good choice of V, the error converges in the V^*-norm with twice the rate of convergence in the $L^2(\Omega, H)$-norm and this is the desired rate of weak convergence. For linear equations we prove that $V = \mathbf{M}^{1,p,p}(H)$ is a good choice for some $p > 2$. The main part of the error $X - Y$ is then a stochastic convolution $\int_0^T E(T-t)\,\mathrm{d}W(t)$. Bounding the error operator E in the right norm yields convergence to the price of a singularity at $t = 0$. By using the inequality (1.1) on this integral with sufficiently large $p = q > 2$, we may integrate a stronger singularity and obtain a higher rate of convergence. For semilinear equations the main difference is that a term involving $F(X) - F(Y)$ appears. We then use $V = \mathbf{G}^{1,p}(H) = \mathbf{M}^{1,p,p}(H) \cap L^{2p}(\Omega, H)$. In Lemma 3.8 we show that $F\colon V^* \to V^*$ is locally Lipschitz with a constant depending on $\|X\|_{\mathbf{M}^{1,2p,p}(H)}$, $\|Y\|_{\mathbf{M}^{1,2p,p}(H)}$. The choice of a stronger V-norm is necessary in order to control the nonlinearity in this way. After bounding these norms, we may use a standard Gronwall argument to bound $\|X - Y\|_{V^*}$.

As our method does not rely on the use of Kolmogorov's equation or Itō's formula, it extends to non-Markovian equations. In the work [1] our method is used to prove weak convergence for semilinear stochastic Volterra equations driven by additive noise. Such equations suffer from the lack of a Kolmogorov equation and therefore the classical proof is not feasible. We hope that our method will enable weak error analysis for other non-Markovian equations such as for instance random evolution PDEs. For a discussion of the differences that arise in connection with a possible extension to multiplicative noise, see Subsection 4.3 below.

An additional advantage of the present work is that we only require the test function φ to have two continuous Fréchet derivatives of polynomial growth. This means, in particular, that we prove convergence of arbitrary moments with the higher rate. Except in [28] for the case of linear equations, the test functions in the weak error analysis are assumed to have bounded derivatives and convergence of moments is treated separately, for example, in [9]. In [1] it is shown that with little extra effort the method works for weak convergence of integrals with respect to general Borel-measures of the entire path of the solution, including for instance weighted sums $\sum_{m=1}^M a_m X(t_m)$, where $a_1, \dots, a_M$ are real numbers and $0 \le t_1 < \dots < t_M \le T$. As far as we can see this appears to be impossible to obtain by the classical method. In addition, our weak error estimate in Theorem 4.4 is uniform with respect to the time partition unlike earlier results in the literature.

The paper is organized as follows. In Section 2 we present preliminary material and our basic assumptions on the stochastic partial differential equation and the numerical scheme. The core of the paper is Section 3 which contains our extensions of the Malliavin calculus. In 3.1 we introduce the refined Watanabe-Sobolev spaces and prove that they are well defined. Duality of our new spaces is treated in 3.2, with the inequality (1.1) and a local Lipschitz result as the main results. In 3.3 and 3.4 regularity results in terms of our new spaces are proved for the solution to our stochastic evolution equation and its approximations, respectively. Section 4 contains the weak convergence. In 4.1 we restrict the discussion to approximations of the stochastic convolution and in 4.2 we treat semilinear equations. Finally, in Section 5 we verify our assumption on the numerical method for Galerkin finite element approximations.

2. Setting and preliminaries

2.1. **Analytic preliminaries.** Let $(U, \|\cdot\|_U, \langle\cdot,\cdot\rangle_U)$ and $(V, \|\cdot\|_V, \langle\cdot,\cdot\rangle_V)$ be separable Hilbert spaces and let $\mathcal{L}(U,V)$ be the Banach space of all bounded linear operators $U \to V$. If $U = V$, then we write $\mathcal{L}(U) = \mathcal{L}(U,U)$ and if $U = H$, we abbreviate $\mathcal{L} = \mathcal{L}(H)$. We denote by $\mathcal{L}_2(U,V) \subset \mathcal{L}(U,V)$ the subspace of all Hilbert-Schmidt operators endowed with the standard norm and inner product

$$\|T\|_{\mathcal{L}_2(U,V)} = \Big(\sum_{j\in\mathbf{N}} \|Tu_j\|_V^2\Big)^{\frac{1}{2}}, \quad \langle S,T\rangle_{\mathcal{L}_2(U,V)} = \sum_{j\in\mathbf{N}} \langle Su_j, Tu_j\rangle_V, \tag{2.1}$$

where both are independent of the particular choice of ON-basis $(u_j)_{j\in\mathbf{N}} \subset U$.

We let $\mathcal{L}^{[m]}(U,V)$ denote the space of multi-linear operators $b\colon U^m \to V$ for $m \geq 1$. We use the notation $b\cdot(u_1,\dots,u_m) = b(u_1,\dots,u_m)$ for $u_1,\dots,u_m \in U$ to emphasize that b is multi-linear and we abbreviate $b\cdot(u,\dots,u) = b\cdot(u)^m$. The norm $\|b\|_{\mathcal{L}^{[m]}(U,V)}$ is the smallest constant $C > 0$ such that

$$\|b\cdot(u_1,\dots,u_m)\|_V \leq C\|u_1\|_U\cdots\|u_m\|_U, \quad \forall u_1,\dots,u_m \in U. \tag{2.2}$$

By $\mathcal{C}^m(U,V)$ we denote the space of all mappings $\phi\colon U \to V$ with continuous Fréchet derivatives of order m, $\mathcal{C}^m_{\mathrm{b}}(U,V)$ is the subspace with $m \geq 1$ bounded derivatives $\phi',\dots,\phi^{(m)}$, and $\mathcal{C}^m_{\mathrm{p}}(U,V)$ denotes the analogous space with derivatives of polynomial growth. On $\mathcal{C}^m_{\mathrm{b}}(U,V)$ we use the natural seminorm $|\phi|_{\mathcal{C}^m_{\mathrm{b}}} = \sup_{x\in U}\|\phi^{(m)}(x)\|_{\mathcal{L}^{[m]}(U,V)}$. We define $\mathcal{C}^0_{\mathrm{b}}(U,V)$ to be all bounded continuous mappings $U \to V$, endowed with the uniform norm. The first derivative of $\phi \in \mathcal{C}^1(U,V)$ is an operator $\phi'(x) \in \mathcal{L}(U,V) = \mathcal{L}^{[1]}(U,V)$ for every $x \in U$. When $V = \mathbf{R}$ we may identify $\phi'(x) \in \mathcal{L}(U,\mathbf{R}) = U^*$ with its gradient $\phi'(x) \in U$ via $\phi'(x)\cdot u = \langle\phi'(x), u\rangle_U$ by the Riesz representation theorem. Similarly, for $\phi \in \mathcal{C}^2(U,\mathbf{R})$ we will sometimes identify $\phi''(x) \in \mathcal{L}^{[2]}(U,\mathbf{R})$ with an operator $\phi''(x) \in \mathcal{L}(U)$ via $\phi''(x)\cdot(u_1,u_2) = \langle\phi''(x)u_1, u_2\rangle_U$. By the mean value theorem we have, for $\phi \in \mathcal{C}^1(U,V)$,

$$\phi(x) = \phi(y) + \int_0^1 \phi'(y + \rho(x-y))\cdot(x-y)\,\mathrm{d}\rho, \quad x,y \in U. \tag{2.3}$$

We will use the following version of Gronwall's Lemma, for a proof see [14, Lemma 7.1].

Lemma 2.1. *Let $T > 0$, $N \in \mathbf{N}$, $k = \frac{T}{N}$, and $t_n = nk$ for $0 \le n \le N$. If $(\varphi_j)_{j=1}^N$ are nonnegative real numbers with*

$$\varphi_n \le C_1 \, (1 + t_n^{-1+\mu}) + C_2 \, k \sum_{j=0}^{n-1} t_{n-j}^{-1+\nu} \varphi_j, \quad 1 \le n \le N,$$

for some constants $C_1, C_2 \ge 0$ and $\mu, \nu > 0$, then there exists a constant $C = C(\mu, \nu, C_2, T)$ such that

$$\varphi_n \le C \, C_1 \, (1 + t_n^{-1+\mu}), \quad 1 \le n \le N.$$

We sometimes write $a \lesssim b$ to denote $a \le Cb$ for some constant $C > 0$. Constants arising from the estimates (2.4), (2.5), (2.12) and (2.14) as well as trivial numerical constants will be suppressed with this symbol.

2.2. **Stochastic preliminaries.** Let $(H, \| \cdot \|, \langle \cdot, \cdot \rangle)$ be a separable Hilbert space and let $Q \in \mathcal{L} = \mathcal{L}(H)$ be a selfadjoint, positive semidefinite operator on H and $Q^{\frac{1}{2}}$ its unique positive square root. The space $H_0 = Q^{\frac{1}{2}}(H)$ is a Hilbert space with scalar product $\langle u, v \rangle_{H_0} = \langle Q^{-\frac{1}{2}} u, Q^{-\frac{1}{2}} v \rangle$. We denote by $\mathcal{L}_2^0 = \mathcal{L}_2(H_0, H)$ the space of Hilbert-Schmidt operators $H_0 \to H$. We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0,T]}, \mathbf{P})$ and the corresponding Bochner spaces $L^p(\Omega, V) = L^p((\Omega, \mathcal{F}, \mathbf{P}), V)$, $p \in [1, \infty]$, V a Banach space. We abbreviate $L^2(\Omega) = L^2(\Omega, \mathbf{R})$. We assume that $(W(t))_{t \in [0,T]}$ is a cylindrical Q-Wiener process, meaning that $W \in C([0,T], \mathcal{L}(H_0, L^2(\Omega)))$ is such that $t \mapsto W(t)u$ is an $\mathcal{F}_t$-predictable real-valued Brownian motion for every $u \in H_0$ and

$$\mathbf{E}\big[W(s)u \, W(t)v\big] = \min(s,t) \langle u, v \rangle_{H_0}, \quad u, v \in H_0, \; s, t \in [0,T].$$

For predictable $\Phi \in L^2([0,T] \times \Omega, \mathcal{L}_2^0)$ the H-valued stochastic Itō-integral

$$\int_0^T \Phi(t) \, \mathrm{d}W(t) \in L^2(\Omega, H),$$

is a well defined random variable. For details on the construction of cylindrical Wiener processes and the corresponding stochastic integral we refer to [10, 35, 38]. For technical reasons we assume that the σ-field $\mathcal{F}$ is generated by $(W(t))_{t \in [0,T]}$ and the filtration $(\mathcal{F}_t)_{t \in [0,T]}$ is the natural filtration associated with $(W(t))_{t \in [0,T]}$.

We cite the following special case of Burkholder's inequality [10, Lemma 7.2].

Lemma 2.2. *Let $(\Phi(t))_{t \in [0,T]}$ be a predictable and $\mathcal{L}_2^0$-valued process such that $\|\Phi\|_{L^p(\Omega, L^2([0,T], \mathcal{L}_2^0))} < \infty$ for some $p \ge 2$. Then there exists a constant C_p, such that*

$$\Big\| \int_0^T \Phi(s) \, \mathrm{d}W(s) \Big\|_{L^p(\Omega, H)} \le C_p \|\Phi\|_{L^p(\Omega, L^2([0,T], \mathcal{L}_2^0))}.$$

2.3. **The stochastic equation.** We study equation (1.2) under the following Assumption and recall that the solution X takes values in H.

Assumption 2.3. (i) *Let $(A, \mathcal{D}(A))$ be a linear operator on H such that $A^{-1} \in \mathcal{L}(H)$ exists and $-A$ is the generator of an analytic semigroup $(S(t))_{t\geq 0}$ of bounded linear operators $S(t) = \mathrm{e}^{-tA}$ on H.*

(ii) *The initial value X_0 is deterministic and satisfies $X_0 \in \dot{H}^{2\beta}$, for some $\beta \in (0,1]$, where $\dot{H}^{\alpha} \subset H$ denotes the domain of $A^{\frac{\alpha}{2}}$.*

(iii) *The covariance operator Q satisfies $\|A^{\frac{\beta-1}{2}}\|_{\mathcal{L}_2^0} = \|A^{\frac{\beta-1}{2}}Q^{\frac{1}{2}}\|_{\mathcal{L}_2} < \infty$, for the same β as in* (ii).

(iv) *The drift $F\colon H \to H$ is assumed to be twice Fréchet differentiable with bounded derivatives of order 1 and 2, i.e., $F \in \mathcal{C}_{\mathrm{b}}^2(H,H)$.*

Under Assumption 2.3 (i) the fractional powers $A^{\frac{r}{2}}$ for $r \in \mathbf{R}$ are well defined, see [34, Section 2.6]. We define the norms $\|v\|_r = \|A^{\frac{r}{2}}v\|$ and let $\dot{H}^r = \mathcal{D}(A^{\frac{r}{2}})$ for $r \geq 0$. For $r < 0$ we define $\dot{H}^r$ as the closure of H under the norm $\|v\|_r$. The spaces $\dot{H}^r \subset H \subset \dot{H}^{-r}$ form a Gelfand triple for $r > 0$.

The analytic semigroup $(S(t))_{t\geq 0}$ generated by $-A$ satisfies, see [34, Section 2.6],

$$\|A^{\varrho}S(t)\|_{\mathcal{L}} \leq C_{\varrho}t^{-\varrho}, \quad t > 0,\ \varrho \geq 0, \tag{2.4}$$

$$\|(S(t) - I)A^{-\varrho}\|_{\mathcal{L}} \leq C_{\varrho}t^{\varrho}, \quad t \geq 0,\ 0 < \varrho \leq 1, \tag{2.5}$$

where $C_{\varrho} = \frac{2}{\varrho}$. For later reference we state the semigroup property:

$$S(s+t) = S(s)S(t),\ s,t \geq 0; \quad S(0) = I. \tag{2.6}$$

Under Assumption 2.3, the stochastic equation (1.2) has a mild solution $X \in \mathcal{C}([0,T], L^p(\Omega, H))$, for every $p \geq 2$, in the sense that it satisfies the integral equation

$$X(t) = S(t)X_0 + \int_0^t S(t-s)F(X(s))\,\mathrm{d}s + \int_0^t S(t-s)\,\mathrm{d}W(s), \quad t \in [0,T], \tag{2.7}$$

and

$$\sup_{t\in[0,T]} \|X(t)\|_{L^p(\Omega,H)} \leq C(1 + \|X_0\|). \tag{2.8}$$

For every $\gamma \in [0,\beta)$ the solution satisfies $X(t) \in \dot{H}^{\gamma}$, $\mathbf{P}$-a.s., for all $t \in [0,T]$. For more details we refer to [10, 22, 23, 29] and the references therein.

2.4. **Approximation of the solution.** We approximate equation (1.2) in finite-dimensional approximation spaces $V_h \subseteq H$, $h \in (0,1]$. The parameter $h \in (0,1]$ is a refinement parameter. We denote by $P_h\colon H \to V_h$ the orthogonal projector onto V_h and by $(A_h)_{h\in(0,1]}$ a family of operators $A_h\colon V_h \to V_h$ approximating A. The assumptions on $(V_h)_{h\in(0,1]}$, and $(A_h)_{h\in(0,1]}$ are given in Assumption 2.4 below.

For the time discretization let $k \in (0,1)$ be the constant step size. We define the discrete time points by $t_n = nk$, $n = 0, \ldots, N$, where $N = N(k) \in \mathbf{N}$ is determined by $t_N \leq T < t_N + k$. We define the operator $S_{h,k} = (I + kA_h)^{-1}P_h$ and notice that $S_{h,k}Q^{\frac{1}{2}} \in \mathcal{L}_2(H)$ since $S_{h,k}$ is a finite rank operator. Hence it is a valid integrand for the stochastic integral. Our completely discrete scheme is to find the recursive sequence $(X^n_{h,k})_{n=0}^N \subset V_h$ given by the semi-implicit Euler-Maruyama method:

$$
\begin{aligned}
&X^{n+1}_{h,k} = S_{h,k}X^n_{h,k} + kS_{h,k}F(X^n_{h,k}) + \int_{t_n}^{t_{n+1}} S_{h,k}\,\mathrm{d}W(s), \quad n = 0,\dots,N-1;\\
&X^0_{h,k} = P_h X_0.
\end{aligned} \tag{2.9}
$$

By iterating (2.9) we obtain the discrete analog of (2.7)

$$
X^n_{h,k} = S^n_{h,k}P_hX_0 + k\sum_{j=0}^{n-1} S^{n-j}_{h,k}F(X^j_{h,k}) + \sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} S^{n-j}_{h,k}\,\mathrm{d}W(t), \quad n = 0,\dots,N. \tag{2.10}
$$

Further, we define the error operators $E^n_{h,k}$, $h,k \in (0,1]$, by

$$
E^n_{h,k} := S(nk) - S^n_{h,k}. \tag{2.11}
$$

We now state our assumption on the numerical discretization.

Assumption 2.4. *The linear operators $A_h\colon V_h \to V_h$ and the orthogonal projectors $P_h\colon H \to V_h$, $h \in (0,1]$, satisfy for all $0 \le \varrho \le \frac{1}{2}$*

$$
\|A_h^{\varrho} S^n_{h,k}\|_{\mathcal{L}} \le Ct_n^{-\varrho}, \quad n = 1,\dots,N, \tag{2.12}
$$

$$
\|A_h^{-\varrho} P_h A^{\varrho}\|_{\mathcal{L}} \le C, \tag{2.13}
$$

uniformly in $h,k \in (0,1]$, and for all $0 \le \theta \le 2$, $0 \le \varrho \le \min(1, 2-\theta)$,

$$
\|E^n_{h,k}A^{\frac{\varrho}{2}}\|_{\mathcal{L}} \le C\big(h^{\theta} + k^{\frac{\theta}{2}}\big)t_n^{-\frac{\theta+\varrho}{2}}, \quad n = 1,\dots,N. \tag{2.14}
$$

We remark that the error estimate (2.14) is non-standard, due to the low regularity regime we consider. In fact, it corresponds to an error estimate for the deterministic linear equation with rough initial data, i.e., $S(t)X_0 = S(t)A^{\frac{\varrho}{2}}x$ with $x \in H$, so that $X_0 = A^{\frac{\varrho}{2}}x \in \dot{H}^{-\varrho}$. We verify (2.14) in Section 5 for the finite element method and the heat equation by means of interpolation techniques, using already established results from [27, 28]. By [27, Example 3.4], spectral Galerkin approximations also fit under our Assumption 2.4.

3. Malliavin calculus

The papers [17] and [31] are the earliest works to treat Malliavin calculus for stochastic evolution equations in the Hilbert space framework. Later it was used in several papers related to optimal control of stochastic partial differential equations, in particular, in connection with backward stochastic differential equations [15] and backward Volterra integral equations in Hilbert spaces [3]. Malliavin differentiability of solutions to stochastic evolution equations is proved in [15]. There are also works using the Malliavin calculus for specific equations outside the setting of the present paper and it is more extensively developed for equations studied in the framework of [39], see the book [36]. We mention also the papers [2], [5], [6], [8], [12], [18], [19], [24], [40], where the Malliavin calculus is applied to the problem of proving weak convergence. Below we take a new direction and introduce in Subsection 4.1 a family of refined Watanabe-Sobolev spaces. We show in Subsection 4.2 that these spaces are particularly useful in connection with duality.

3.1. **Refined Watanabe-Sobolev spaces.** Let $I\colon L^2([0,T],H_0) \to L^2(\Omega)$ be the mapping given by

$$I(\phi) = \int_0^T \phi(t)\,\mathrm{d}W(t), \quad \phi \in L^2([0,T],H_0),$$

where we identify $L^2([0,T],H_0) \cong L^2([0,T],\mathcal{L}_2(H_0,\mathbf{R}))$. This identification is important since an $\mathbf{R}$-valued stochastic integral has an $L^2([0,T],\mathcal{L}_2(H_0,\mathbf{R}))$-valued integrand. Fix an ON-basis $(\phi_j)_{j\in\mathbf{N}} \subset L^2([0,T],H_0)$, let $\mathcal{P}_n$ be the set of random variables given by n:th order polynomials of the random variables $(I(\phi_j))_{j\in\mathbf{N}}$. The set $\cup_{n\in\mathbf{N}}\mathcal{P}_n$ is independent of the choice of basis, see [21], and

$$\bigcup_{n\in\mathbf{N}} \mathcal{P}_n \subset L^p(\Omega) \text{ is dense for } 1 \le p < \infty. \tag{3.1}$$

Let $2 \le q \le \infty$ and let the mapping $i\colon L^q([0,T],H_0) \to L^2([0,T],H_0)$ denote the canonical embedding. Let $\mathcal{S}^q$ be the set of random variables F of the form

$$\begin{aligned} &F = f(I(i(\phi_1)),\dots,I(i(\phi_n))),\\ &f \in \mathcal{C}^1_{\mathrm{p}}(\mathbf{R}^n,\mathbf{R}),\ (\phi_j)_{j=1}^n \subset L^q([0,T],H_0),\ n\in\mathbf{N}. \end{aligned} \tag{3.2}$$

The class $\mathcal{S}^2$ is standard in Malliavin calculus and is usually denoted by $\mathcal{S}$. Our definition coincides with that in [28] but in the standard work [33] and many other works $\mathcal{C}^\infty_{\mathrm{p}}(\mathbf{R}^n,\mathbf{R})$ is used instead of $\mathcal{C}^1_{\mathrm{p}}(\mathbf{R}^n,\mathbf{R})$. The classes $\mathcal{S}^q$ for $q > 2$ are new to our knowledge.

Lemma 3.1. *For every $1 \le p < \infty$ and $2 \le q \le \infty$, $\mathcal{S}^q \subset L^p(\Omega)$ is dense.*

Proof. Without causing confusion we also let i denote the canonical embedding from $L^q([0,T],\mathbf{R})$ to $L^2([0,T],\mathbf{R})$. We notice the isomorphism $L^2([0,T],H_0) \cong L^2([0,T],\mathbf{R}) \otimes H_0$.

Since there even exists a bounded ON-basis of the space $L^2([0,T],\mathbf{R})$ we clearly find a sequence $(f_n)_{n\in\mathbf{N}} \subset L^q([0,T],\mathbf{R})$ such that $(i(f_n))_{n\in\mathbf{N}}$ is an ON-basis for $L^2([0,T],\mathbf{R})$. If $(h_n)_{n\in\mathbf{N}}$ is an ON-basis for H_0, then $(i(f_m)\otimes h_n)_{m,n\in\mathbf{N}}$ is an ON-basis for $L^2([0,T],\mathbf{R}) \otimes H_0$. In particular, we have that $i(f_m \otimes h_n) = i(f_m) \otimes h_n$.

Since the result (3.1) is independent of the choice of the basis, we conclude our assertion by using the sequence $(I(i(f_m \otimes h_n)))_{m,n\in\mathbf{N}}$. □

For $1 \le p < \infty$ and $2 \le q \le \infty$ we define the action of the Malliavin derivative $D\colon \mathcal{S}^q \to L^p(\Omega, L^q([0,T],H_0))$ on a random variable F of the form (3.2) by

$$D_t F = \sum_{j=1}^n \partial_j f(I(i(\phi_1)),\dots,I(i(\phi_n))) \otimes \phi_j(t), \quad t \in [0,T].$$

This is well defined because $\phi_1,\dots,\phi_n \in L^q([0,T],H_0)$, the random variables $I(\phi_1),\dots,I(\phi_n)$ are Gaussian with all existing moments and since f has polynomial growth. By a direct modification of [28, Proposition 4.2] it does not depend on the specific representation of F.

We remark that for $q = 2$ the linear operator $D\colon \mathcal{S}^2 \to L^p(\Omega, L^2([0,T], H_0))$ is the standard Malliavin derivative. Technically speaking, we have restricted the domain of the Malliavin derivative to $\mathcal{S}^q \subset \mathcal{S}^2$ for $2 < q \leq \infty$. By this we have ensured that $D|_{\mathcal{S}^q}$ maps into the smaller space $L^p(\Omega, L^q([0,T], H_0)) \subset L^p(\Omega, L^2([0,T], H_0))$.

We define the Malliavin derivative for H-valued random variables as in [28, Chap. 4]. For this we denote by $\mathcal{S}^q(H)$ the collection of all H-valued smooth random variables of the form

$$X = \sum_{j=1}^n h_j \otimes F_j, \quad h_1, \ldots, h_n \in H, \quad F_1, \ldots, F_n \in \mathcal{S}^q,\ n \in \mathbf{N}.$$

Since H is separable and by Lemma 3.1 it follows that $\mathcal{S}^q(H)$ is dense in $L^p(\Omega, H)$ for all $1 \leq p < \infty$. The Malliavin derivative $D\colon \mathcal{S}^q(H) \to L^p(\Omega, L^q([0,T], \mathcal{L}_2^0))$ acts in the following way:

$$D_t X = D_t \sum_{j=1}^n h_j \otimes F_j = \sum_{j=1}^n h_j \otimes D_t F_j, \quad t \in [0,T].$$

Here we did the identifications

$$H \otimes L^p(\Omega, L^q([0,T], H_0)) \cong L^p(\Omega, H \otimes L^q([0,T], H_0)) \cong L^p(\Omega, L^q([0,T], \mathcal{L}_2^0)).$$

We write $D_t^u X = D_t X u \in L^2(\Omega, H)$ for the derivative in the direction $u \in H_0$.

In the final step of its construction we extend the domain of the Malliavin derivative to its closure with respect to the graph norm. For this we recall that an unbounded operator $A\colon U \to V$ is *closable* if and only if for every $(u_n)_{n \in \mathbf{N}} \subset U$ such that $\lim_{n\to\infty} u_n = 0$ and $\lim_{n\to\infty} Au_n = v$, we have $v = 0$.

Lemma 3.2. *The Malliavin derivative $D\colon \mathcal{S}^q \to L^p(\Omega, L^q([0,T], \mathcal{L}_2^0))$ is closable for every $1 < p < \infty$ and $2 \leq q \leq \infty$.*

Proof. We will use the fact that $D\colon \mathcal{S}^2(H) \to L^p(\Omega, L^2([0,T], \mathcal{L}_2^0))$ is closable for $p > 1$, [28, Proposition 4.4]. Let $(X_n)_{n\in\mathbf{N}} \subset \mathcal{S}^q(H) \subset \mathcal{S}^2(H)$ be a sequence satisfying $\lim_{n\to\infty} X_n = 0$ in $L^p(\Omega, H)$ such that $\lim_{n\to\infty} DX_n = Z$ in $L^p(\Omega, L^q([0,T], \mathcal{L}_2^0))$, and hence also in $L^p(\Omega, L^2([0,T], \mathcal{L}_2^0))$. By the closability we have $Z = 0$ in $L^p(\Omega, L^2([0,T], \mathcal{L}_2^0))$ and, consequently, also in $L^p(\Omega, L^q([0,T], \mathcal{L}_2^0))$. □

For $1 < p < \infty$ and $2 \leq q \leq \infty$ we can therefore consider the closure $\mathbf{M}^{1,p,q}(H)$ of $\mathcal{S}^q(H)$ with respect to the norm

$$\|X\|_{\mathbf{M}^{1,p,q}(H)} = \Big(\|X\|^p_{L^p(\Omega,H)} + \|DX\|^p_{L^p(\Omega, L^q([0,T], \mathcal{L}_2^0))} \Big)^{\frac{1}{p}}.$$

Clearly, the spaces $\mathbf{M}^{1,p,2}(H)$, $p > 1$, coincide with the classical Watanabe-Sobolev spaces of the Malliavin calculus, which are usually denoted by $\mathbf{D}^{1,p}(H)$. The standard Malliavin derivative is uniquely extended to an operator from $\mathbf{M}^{1,p,2}(H)$ to $L^p(\Omega, L^2([0,T], \mathcal{L}_2^0))$. In addition it holds $\mathbf{M}^{1,p,q_1}(H) \subset \mathbf{M}^{1,p,q_2}(H)$ for all $\infty \geq q_1 \geq q_2 \geq 2$ and from Lemma 3.2 it follows that the restriction of the standard Malliavin derivative $D|_{\mathbf{M}^{1,p,q}(H)}$ is a well-defined operator from $\mathbf{M}^{1,p,q}(H)$ to $L^p(\Omega, L^q([0,T], \mathcal{L}_2^0))$. If $p = q$ we abbreviate $\mathbf{M}^{1,p}(H) = \mathbf{M}^{1,p,p}(H)$.

The space $\mathbf{M}^{1,2}(H)$ is a Hilbert space and it has a well developed theory of Malliavin calculus. The adjoint of the Malliavin derivative $D\colon \mathbf{M}^{1,2}(H) \subset L^2(\Omega, H) \to L^2([0,T] \times \Omega, \mathcal{L}_2^0)$ is called the *divergence* operator or the *Skorohod* integral and denoted by $\delta\colon L^2([0,T] \times \Omega, \mathcal{L}_2^0) \to L^2(\Omega, H)$ with domain $\mathcal{D}(\delta)$. The duality reads

$$\big\langle X, \delta\Phi \big\rangle_{L^2(\Omega,H)} = \big\langle DX, \Phi \big\rangle_{L^2([0,T]\times\Omega,\mathcal{L}_2^0)}, \quad X \in \mathbf{M}^{1,2}(H),\ \Phi \in \mathcal{D}(\delta). \tag{3.3}$$

We refer to this as the Malliavin integration by parts formula. It is well known that for predictable $\Phi \in \mathcal{D}(\delta)$ the action of δ coincides with that of the H-valued Itō integral, i.e., $\delta\Phi = \int_0^T \Phi(t) \,\mathrm{d}W(t)$, [28, Proposition 4.12].

In the remainder of this subsection we state a modification of the chain rule from [28, Lemma 4.7] and a product rule for the Malliavin derivative.

Lemma 3.3. *Let U, V be two separable Hilbert spaces and let $\sigma \in \mathcal{C}^1(U,V)$, be such that there exist constants C and $r \geq 0$ with*

$$\|\sigma(u)\|_V \leq C\big(1 + \|u\|_U^{1+r}\big), \quad \|\sigma'(u)\|_{\mathcal{L}(U,V)} \leq C\big(1 + \|u\|_U^r\big),$$

for all $u \in U$. Then for every $1 < p < \infty$, $2 \leq q \leq \infty$ and $X \in \mathbf{M}^{1,(1+r)p,q}(U)$ it follows that $\sigma(X) \in \mathbf{M}^{1,p,q}(V)$ with $\|\sigma(X)\|_{\mathbf{M}^{1,p,q}(V)} \lesssim (1 + \|X\|^{1+r}_{\mathbf{M}^{1,(1+r)p,q}(U)})$ and

$$D_t(\sigma(X)) = \sigma'(X) \cdot D_t X, \quad t \in [0,T]. \tag{3.4}$$

Proof. Let $p > 1$ be arbitrary. For $q = 2$ the result follows directly from [28, Lemma 4.7]. Therefore, it suffices to show that $\|\sigma(X)\|_{\mathbf{M}^{1,p,q}(V)} < \infty$ if $X \in \mathbf{M}^{1,(1+r)p,q}(U)$ for $q > 2$. Indeed, from the polynomial growth condition it follows that

$$\big\|\sigma(X)\big\|_{L^p(\Omega,V)} \leq C\big(1 + \|X\|^{1+r}_{L^{(1+r)p}(\Omega,U)}\big) \leq C\big(1 + \|X\|^{1+r}_{\mathbf{M}^{1,(1+r)p,q}(U)}\big).$$

Moreover, it holds

$$\begin{aligned}
&\big\|D\sigma(X)\big\|_{L^p(\Omega,L^q([0,T],\mathcal{L}_2(H_0,V)))}\\
&\quad = \Big(\mathbf{E}\big[\big\|\sigma'(X) \cdot DX\big\|^p_{L^q([0,T],\mathcal{L}_2(H_0,V))}\big]\Big)^{\frac{1}{p}}\\
&\quad \lesssim \Big(\mathbf{E}\big[\big(1 + \big\|X\big\|_U^r\big)^p \|DX\|^p_{L^q([0,T],\mathcal{L}_2(H_0,U))}\big]\Big)^{\frac{1}{p}}\\
&\quad \leq \big(1 + \|X\|^r_{L^{(1+r)p}(\Omega,U)}\big)\big\|DX\big\|_{L^{(1+r)p}(\Omega,L^q([0,T],\mathcal{L}_2(H_0,U)))}\\
&\quad \lesssim \big(1 + \|X\|^{1+r}_{\mathbf{M}^{1,(1+r)p,q}(U)}\big),
\end{aligned}$$

where we applied the polynomial growth condition on σ' and the Hölder inequality with exponents $(r+1)/r$ and $(r+1)$. This completes the proof. □

Lemma 3.4. *For $\sigma \in \mathcal{C}_{\mathrm{b}}^2(H)$ it holds $\sigma'(X) \cdot Y \in \mathbf{M}^{1,p,q}(H)$ for all $X, Y \in \mathbf{M}^{1,2p,q}(H)$, $1 \leq p < \infty$, $2 \leq q \leq \infty$. In addition, we have*

$$D_t(\sigma'(X) \cdot Y) = \sigma''(X) \cdot (D_t X, Y) + \sigma'(X) \cdot D_t Y, \quad t \in [0,T]. \tag{3.5}$$

Proof. The proof is done by an application of the chain rule. For this define the mapping $\gamma\colon H \times H \to H$ given by $\gamma(x,y) = \sigma'(x) \cdot y$. Certainly, it holds $\gamma \in \mathcal{C}^1(H \times H, H)$ and we have $\|\gamma(x,y)\| = \|\sigma'(x) \cdot y\| \leq |\sigma|_{\mathcal{C}_{\mathrm{b}}^1} \|y\|$ for all $(x,y) \in H \times H$. Further, it holds

$$\gamma'(x,y) \cdot (z_1, z_2) = \sigma''(x) \cdot (z_1, y) + \sigma'(x) \cdot z_2,$$

for all $(x,y),(z_1,z_2) \in H \times H$. Therefore,

$$\begin{aligned}\big\|\gamma'(x,y)\cdot(z_1,z_2)\big\| &\le |\sigma|_{\mathcal{C}^2_{\mathrm{b}}}\|z_1\|\|y\| + |\sigma|_{\mathcal{C}^1_{\mathrm{b}}}\|z_2\| \\ &\le \max\{|\sigma|_{\mathcal{C}^1_{\mathrm{b}}}, |\sigma|_{\mathcal{C}^2_{\mathrm{b}}}\}\big(1+\|y\|\big)\big(\|z_1\|+\|z_2\|\big).\end{aligned}$$

Hence, γ satisfies the assumption of Lemma 3.3 with $r=1$ and the result follows. ☐

3.2. **Duality.** For any $2 \le p < \infty$, $2 \le q \le \infty$ the inclusion $\mathbf{M}^{1,p,q}(H) \subset L^2(\Omega, H)$ is dense and continuous and hence the spaces

$$\mathbf{M}^{1,p,q}(H) \subset L^2(\Omega,H) \subset \mathbf{M}^{1,p,q}(H)^*, \tag{3.6}$$

define a Gelfand triple, where we identify $L^2(\Omega,H) \cong L^2(\Omega,H)^*$ by the Riesz Representation Theorem. We denote the dual pairing of $\mathbf{M}^{1,p,q}(H)^*$ and $\mathbf{M}^{1,p,q}(H)$ by $[Z,Y]$ for $Z \in \mathbf{M}^{1,p,q}(H)^*$, $Y \in \mathbf{M}^{1,p,q}(H)$. The inclusion $L^2(\Omega,H) \subset \mathbf{M}^{1,p,q}(H)^*$ is realized through the definition $[Z,Y] = \langle Z,Y\rangle_{L^2(\Omega,H)}$ for all $Z \in L^2(\Omega,H)$, $Y \in \mathbf{M}^{1,p,q}(H)$, with the norm

$$\|Z\|_{\mathbf{M}^{1,p,q}(H)^*} = \sup_{Y \in \mathbf{M}^{1,p,q}(H)} \frac{\langle Y,Z\rangle_{L^2(\Omega,H)}}{\|Y\|_{\mathbf{M}^{1,p,q}(H)}}, \quad Z \in L^2(\Omega,H). \tag{3.7}$$

We will now prove an inequality that is a sort of Burkholder inequality in the $\mathbf{M}^{1,p,q}(H)^*$-norm. Lemma 2.2 gives an estimate that is L^2 in time. In this weaker norm we get an estimate that is $L^{q'}$ in time, where q' is the conjugate exponent to q given by $\frac{1}{q} + \frac{1}{q'} = 1$ if $q < \infty$ and $q' = 1$ otherwise. Since $q \in [2,\infty]$, and hence $q' \in [1,2]$, we can integrate worse singularities.

Theorem 3.5. *Let $p \in [2,\infty)$, $q \in [2,\infty]$ and p', q' denote the conjugate exponents given by $\frac{1}{p}+\frac{1}{p'}=1$ and $\frac{1}{q}+\frac{1}{q'}=1$ with $q'=1$ for $q=\infty$. If $\Phi \in L^2([0,T]\times\Omega, \mathcal{L}_2^0)$ is predictable, then*

$$\Big\| \int_0^T \Phi(t)\,\mathrm{d}W(t)\Big\|_{\mathbf{M}^{1,p,q}(H)^*} \le \|\Phi\|_{L^{p'}(\Omega, L^{q'}([0,T],\mathcal{L}_2^0))}.$$

Proof. We use the fact that the stochastic integral of Φ equals $\delta\Phi$. By (3.7), (3.3), and Hölder's inequality, we get

$$\begin{aligned}\big\|\delta\Phi\big\|_{\mathbf{M}^{1,p,q}(H)^*} &= \sup_{Y \in \mathbf{M}^{1,p,q}(H)} \frac{\big\langle Y,\delta\Phi\big\rangle_{L^2(\Omega,H)}}{\|Y\|_{\mathbf{M}^{1,p,q}(H)}} = \sup_{Y \in \mathbf{M}^{1,p,q}(H)} \frac{\big\langle DY,\Phi\big\rangle_{L^2([0,T]\times\Omega,\mathcal{L}_2^0)}}{\|Y\|_{\mathbf{M}^{1,p,q}(H)}} \\ &\le \sup_{Y \in \mathbf{M}^{1,p,q}(H)} \frac{\|DY\|_{L^p(\Omega,L^q([0,T],\mathcal{L}_2^0))}\big\|\Phi\big\|_{L^{p'}(\Omega,L^{q'}([0,T],\mathcal{L}_2^0))}}{\|Y\|_{\mathbf{M}^{1,p,q}(H)}} \\ &\le \|\Phi\|_{L^{p'}(\Omega,L^{q'}([0,T],\mathcal{L}_2^0))},\end{aligned}$$

which finishes the proof. ☐

Remark 3.6. We prove the reverse inequality for deterministic $\Phi \in L^2(\Omega \times [0,T], \mathcal{L}_2^0)$. Since $\mathcal{H}_1^q(H) = \{\delta\Psi : \Psi \in L^q([0,T],\mathcal{L}_2^0)\} \subset \mathbf{M}^{1,p,q}(H)$ we get an inequality by taking the supremum over $\mathcal{H}_1^q(H)$ in (3.7) instead of $\mathbf{M}^{1,p,q}(H)$. We also use the identity $D\delta\Psi = \Psi + \delta D\Psi = \Psi$ for deterministic $\Psi \in L^q([0,T],\mathcal{L}_2^0)$. By Burkholder's inequality Lemma 2.2 and Hölder's inequality we get

$$\begin{aligned}
\big\|\delta\Phi\big\|_{\mathbf{M}^{1,p,q}(H)^*} &= \sup_{Y\in\mathbf{M}^{1,p,q}(H)} \frac{\big\langle Y,\delta\Phi\big\rangle_{L^2([0,T]\times\Omega,\mathcal{L}_2^0)}}{\|Y\|_{\mathbf{M}^{1,p,q}(H)}} \\
&\geq \sup_{Y\in\mathcal{H}_1^q(H)} \frac{\big\langle DY,\Phi\big\rangle_{L^2([0,T]\times\Omega,\mathcal{L}_2^0)}}{\|Y\|_{\mathbf{M}^{1,p,q}(H)}} \\
&= \sup_{\Psi\in L^q([0,T],\mathcal{L}_2^0)} \frac{\big\langle D\delta\Psi,\Phi\big\rangle_{L^2([0,T]\times\Omega,\mathcal{L}_2^0)}}{\Big(\|\delta\Psi\|^p_{L^p(\Omega,H)} + \|D\delta\Psi\|^p_{L^p(\Omega,L^q([0,T],\mathcal{L}_2^0))}\Big)^{\frac{1}{p}}} \\
&\geq \sup_{\Psi\in L^q([0,T],\mathcal{L}_2^0)} \frac{\big\langle \Psi,\Phi\big\rangle_{L^2([0,T],\mathcal{L}_2^0)}}{\Big(C_p^p\|\Psi\|^p_{L^2([0,T],\mathcal{L}_2^0)} + \|\Psi\|^p_{L^q([0,T],\mathcal{L}_2^0)}\Big)^{\frac{1}{p}}} \\
&\geq \frac{1}{\big(C_p^p T^{\frac{q}{q-2}}+1\big)^{\frac{1}{p}}} \sup_{\Psi\in L^q([0,T],\mathcal{L}_2^0)} \frac{\big\langle \Psi,\Phi\big\rangle_{L^2([0,T],\mathcal{L}_2^0)}}{\|\Psi\|_{L^q([0,T],\mathcal{L}_2^0)}} \\
&= \frac{1}{\big(C_p^p T^{\frac{q}{q-2}}+1\big)^{\frac{1}{p}}} \|\Phi\|_{L^{q'}([0,T],\mathcal{L}_2^0)}.
\end{aligned}$$

The proof relies strongly on the fact that $D\Psi = 0$. For random Φ one needs random $\Psi \in L^p(\Omega, L^q([0,T],\mathcal{L}_2^0))$ and, since $\delta D\Psi \neq 0$ in this case, the proof above does not work.

Remark 3.7. One consequence of Theorem 3.5 is that the stochastic integral can be extended in $\mathbf{M}^{1,p,q}(H)^*$ to integrands in $L^{p'}(\Omega, L^{q'}([0,T],\mathcal{L}_2^0))$. The elements of $\mathbf{M}^{1,p,q}(H)^*$ are distributions defined by their action on random variables in $\mathbf{M}^{1,p,q}(H)$. One can show that the solution of the linear stochastic heat equation driven by space-time white noise in two space dimensions is a stochastic process $X \in \mathcal{C}([0,T],\mathbf{M}^{1,p,q}(H)^*)$ for every $p \geq 2$ and $q > 2$. In three space dimensions the same is valid for every $p \geq 2$ and $q > 4$. In higher space dimensions than three the solution is not $\mathbf{M}^{1,p,q}(H)^*$-valued since this would force $q' < 1$. Hölder continuity in time in the $\mathbf{M}^{1,p,q}(H)^*$-norms can be shown for the solution in two and three space dimensions for the p, q for which the solution is defined. See Lemma 3.8 below for the regular case.

Solutions defined in a distributional sense with respect to Ω is not conceptually new. This is the heart of the white noise approach to SPDE, see, e.g., [4],[20]. In [4] semilinear equations are considered in the random field framework and a fixed point argument is performed. In a future work we will explore if there are suitable spaces in which such an analysis can be performed for equations within the semigroup framework and then, in particular, for the case of multiplicative noise.

Theorem 3.5 is a key result in the present work. But to be able to perform error estimates for semilinear equations we also need an intermediate space between $\mathbf{M}^{1,p,p}(H)$ and $\mathbf{M}^{1,2p,p}(H)$. For $2 \leq p < \infty$ we define $\mathbf{G}^{1,p}(H)$, to be the subspace of $\mathbf{M}^{1,p,p}(H)$ with the additional integrability property $\mathbf{G}^{1,p}(H) \subset L^{2p}(\Omega, H)$, i.e.,

$$\mathbf{G}^{1,p}(H) = \mathbf{M}^{1,p,p}(H) \cap L^{2p}(\Omega, H).$$

It is a Banach space equipped with the norm

$$\|Y\|_{\mathbf{G}^{1,p}(H)} = \max\Big(\|Y\|_{\mathbf{M}^{1,p,p}(H)}, \|Y\|_{L^{2p}(\Omega,H)}\Big).$$

It holds $\mathbf{M}^{1,2p,p}(H) \subset \mathbf{G}^{1,p}(H) \subset \mathbf{M}^{1,p,p}(H)$ and we obtain a new Gelfand triple

$$\mathbf{G}^{1,p}(H) \subset L^2(\Omega, H) \subset \mathbf{G}^{1,p}(H)^*. \tag{3.8}$$

Our next key result is stated in Lemma 3.8 below. It establishes a local Lipschitz bound in the $\mathbf{G}^{1,p}(H)^*$-norm. This helps us to perform a Gronwall argument in this norm in Section 4.2.

Lemma 3.8. *Let $p \geq 2$. For $F \in \mathcal{C}^2_{\mathrm{b}}(H)$ the following local Lipschitz bound holds*

$$\begin{aligned}&\big\|F(X_1) - F(X_2)\big\|_{\mathbf{G}^{1,p}(H)^*}\\ &\quad\leq \max\big(|F|_{\mathcal{C}^1_{\mathrm{b}}}, |F|_{\mathcal{C}^2_{\mathrm{b}}}\big)\Big(1 + \sum_{i=1}^{2}\|X_i\|_{\mathbf{M}^{1,2p,p}(H)}\Big)\big\|X_1 - X_2\big\|_{\mathbf{G}^{1,p}(H)^*},\end{aligned}$$

for all $X_1, X_2 \in \mathbf{M}^{1,2p,p}(H)$.

Proof. In view of (2.3) it suffices to show

$$\begin{aligned}&\|F'(X)\cdot Y\|_{\mathbf{G}^{1,p}(H)^*}\\ &\quad\leq \max\big(|F|_{\mathcal{C}^1_{\mathrm{b}}}, |F|_{\mathcal{C}^2_{\mathrm{b}}}\big)\Big(1 + \|X\|_{\mathbf{M}^{1,2p,p}(H)}\Big)\|Y\|_{\mathbf{G}^{1,p}(H)^*},\end{aligned} \tag{3.9}$$

for all $X, Y \in \mathbf{M}^{1,2p,p}(H)$. We have

$$\|F'(X)\cdot Y\big\|_{\mathbf{G}^{1,p}(H)^*} \leq \big\|F'(X)^*\big\|_{\mathcal{L}(\mathbf{G}^{1,p}(H))}\big\|Y\big\|_{\mathbf{G}^{1,p}(H)^*},$$

since $\|F'(X)\|_{\mathcal{L}(\mathbf{G}^{1,p}(H)^*)} = \|F'(X)^*\|_{\mathcal{L}(\mathbf{G}^{1,p}(H))}$. To bound the latter norm we calculate

$$\begin{aligned}\|F'(X)^*\cdot Z\|_{L^{2p}(\Omega,H)} &\leq \big(\mathbf{E}\big[\|F'(X)^*\|^{2p}_{\mathcal{L}}\|Z\|^{2p}\big]\big)^{\frac{1}{2p}}\\ &\leq |F|_{\mathcal{C}^1_{\mathrm{b}}}\|Z\|_{L^{2p}(\Omega,H)} \leq |F|_{\mathcal{C}^1_{\mathrm{b}}}\|Z\|_{\mathbf{M}^{1,2p,p}(H)}.\end{aligned}$$

Further, by the product rule (3.5) and the Cauchy-Schwarz inequality we get

$$\begin{aligned}&\big\|DF'(X)^*\cdot Z\big\|_{L^p(\Omega,L^p([0,T],\mathcal{L}^0_2))}\\ &\quad= \big\|F''(X)^*\cdot\big(DX, Z\big) + F'(X)^*\cdot DZ\big\|_{L^p(\Omega,L^p([0,T],\mathcal{L}^0_2))}\\ &\quad\leq |F|_{\mathcal{C}^2_{\mathrm{b}}}\Big(\mathbf{E}\Big[\int_0^T\|D_sX\|^p_{\mathcal{L}^0_2}\,\mathrm{d}s\,\|Z\|^p\Big]\Big)^{\frac{1}{p}} + |F|_{\mathcal{C}^1_{\mathrm{b}}}\|DZ\|_{L^p(\Omega,L^p([0,T],\mathcal{L}^0_2))}\\ &\quad\leq |F|_{\mathcal{C}^2_{\mathrm{b}}}\|DX\|_{L^{2p}(\Omega,L^p([0,T],\mathcal{L}^0_2))}\|Z\|_{L^{2p}(\Omega,H)} + |F|_{\mathcal{C}^1_{\mathrm{b}}}\|DZ\|_{L^p(\Omega,L^p([0,T],\mathcal{L}^0_2))}\\ &\quad\leq \max\big(|F|_{\mathcal{C}^1_{\mathrm{b}}}, |F|_{\mathcal{C}^2_{\mathrm{b}}}\big)\Big(1 + \|X\|_{\mathbf{M}^{1,2p,p}(H)}\Big)\|Z\|_{\mathbf{G}^{1,p}(H)}.\end{aligned}$$

Taken together, these bounds prove (3.9), which completes the proof. □

3.3. **Regularity of the solution.** Here we prove a regularity result for the Malliavin derivative as well as Hölder continuity in the $\mathbf{M}^{1,p}(H)^*$-norm of the solution X to (2.7) under Assumption 2.3. For a suitably chosen p the latter turns out to be twice as high as in the $L^2(\Omega,H)$-norm. The Malliavin derivative $D_rX(t)$ of $X(t)$ at time $r\in[0,T]$ satisfies the equation [15, Proposition 3.5 (ii)]

$$\begin{aligned} &D_rX(t) = S(t-r) + \int_r^t S(t-s)F'(X(s))D_rX(s)\,\mathrm{d}s, \quad t\in[r,T],\\ &D_rX(t)=0, \quad t\in[0,r]. \end{aligned} \tag{3.10}$$

Proposition 3.9. *Let Assumption 2.3 hold. If $\beta\in(0,1)$, then*

$$\sup_{t\in[0,T]} \big\|X(t)\big\|_{\mathbf{M}^{1,p,q}(H)} < \infty,$$

for every $2\le q<\frac{2}{1-\beta}$ and $2\le p<\infty$. If $\beta=1$, then the same holds for every $2\le q\le\infty$.

Proof. We remark that the case $p=q=2$ was already proved in [15]. The moment estimate (2.8) tells that $\sup_{t\in[0,T]}\|X(t)\|_{L^p(\Omega,H)}<\infty$ for every $2\le p<\infty$ and $t\in[0,T]$. Using (3.10) and $D_rX(t)=0$ for $r\in[t,T]$ as well as Assumption 2.3 (iii), we get

$$\begin{aligned} &\big\|DX(t)\big\|_{L^p(\Omega,L^q([0,T],\mathcal{L}_2^0))} = \Big(\mathbf{E}\Big[\Big(\int_0^t \|D_rX(t)\|^q_{\mathcal{L}_2^0}\,\mathrm{d}r\Big)^{\frac{p}{q}}\Big]\Big)^{\frac1p}\\ &\quad\le \big\|S(t-\cdot)\big\|_{L^q([0,t],\mathcal{L}_2^0)} + \Big\|\int_0^t S(t-s)F'(X(s))DX(s)\,\mathrm{d}s\Big\|_{L^p(\Omega,L^q([0,T],\mathcal{L}_2^0))}\\ &\quad\le \big\|S\big\|_{L^q([0,T],\mathcal{L}_2^0)} + \int_0^t \big\|S(t-s)F'(X(s))DX(s)\big\|_{L^p(\Omega,L^q([0,T],\mathcal{L}_2^0))}\,\mathrm{d}s\\ &\quad\le \big\|SA^{\frac{1-\beta}{2}}\big\|_{L^q([0,T],\mathcal{L})}\big\|A^{\frac{\beta-1}{2}}\big\|_{\mathcal{L}_2^0} + |F|_{\mathcal{C}_\mathrm{b}^1}\int_0^t \big\|DX(s)\big\|_{L^p(\Omega,L^q([0,T],\mathcal{L}_2^0))}\,\mathrm{d}s. \end{aligned} \tag{3.11}$$

With our choice of q we have, by (2.4) with $\varrho=\frac{1-\beta}{2}$,

$$\big\|SA^{\frac{1-\beta}{2}}\big\|_{L^q([0,T],\mathcal{L})} = \Big(\int_0^T \|S(t)A^{\frac{1-\beta}{2}}\|^q_{\mathcal{L}}\,\mathrm{d}t\Big)^{\frac1q} \lesssim \Big(\int_0^T t^{-q\frac{1-\beta}{2}}\,\mathrm{d}t\Big)^{\frac1q}<\infty.$$

We conclude by a standard Gronwall lemma. □

Proposition 3.10. *Let Assumption 2.3 hold with $\beta\in(0,1]$ and denote by X the solution to (2.7). For $\gamma\in[0,\beta)$ set $q=\frac{2}{1-\gamma}$. Then there exists for every $p\in[2,\infty)$ a constant $C>0$ such that*

$$\big\|X(t_2)-X(t_1)\big\|_{\mathbf{M}^{1,p,q}(H)^*} \le C\big|t_2-t_1\big|^{\gamma}, \quad t_1,t_2\in(0,T].$$

Proof. Without loss of generality we assume $t_2 > t_1 > 0$. From (2.7) we then get

$$\begin{aligned} X(t_2) - X(t_1) = &\big(S(t_2-t_1) - I\big)S(t_1)X_0 \\ &+ \big(S(t_2-t_1)-I\big)\int_0^{t_1} S(t_1-s)F(X(s))\,\mathrm{d}s \\ &+ \big(S(t_2-t_1)-I\big)\int_0^{t_1} S(t_1-s)\,\mathrm{d}W(s) \\ &+ \int_{t_1}^{t_2} S(t_2-s)F(X(s))\,\mathrm{d}s + \int_{t_1}^{t_2} S(t_2-s)\,\mathrm{d}W(s). \end{aligned}$$

In the following we study the $\mathbf{M}^{1,p,q}(H)^*$-norms of these five summands. For the first, second, and fourth terms we use the fact that $\|Z\|_{\mathbf{M}^{1,p,q}(H)^*} \le \|Z\|_{L^2(\Omega,H)}$.

For the first summand, we use (2.5) with $\varrho = \gamma$ and (2.4) with $\varrho = 0$ as well as Assumption 2.3 (ii). This yields

$$\begin{aligned} \big\|\big(S(t_2-t_1)-I\big)S(t_1)X_0\big\|_{\mathbf{M}^{1,p,q}(H)^*} &\le \big\|\big(S(t_2-t_1)-I\big)A^{-\gamma}S(t_1)A^{\gamma}X_0\big\|_{L^2(\Omega,H)} \\ &\lesssim \big|t_2-t_1\big|^{\gamma}\|A^{\gamma}X_0\| \lesssim \big|t_2-t_1\big|^{\gamma}\|X_0\|_{\dot{H}^{2\beta}}. \end{aligned}$$

The estimate of the second summand is done by applying Assumption 2.3 (iv) and the same arguments as for the first term. To be more precise, we get

$$\begin{aligned} &\Big\|\big(S(t_2-t_1)-I\big)\int_0^{t_1} S(t_1-s)F(X(s))\,\mathrm{d}s\Big\|_{\mathbf{M}^{1,p,q}(H)^*} \\ &\le \big\|\big(S(t_2-t_1)-I\big)A^{-\gamma}\big\|_{\mathcal{L}}\int_0^{t_1}\big\|A^{\gamma}S(t_1-s)\big\|_{\mathcal{L}}\big\|F(X(s))\big\|_{L^2(\Omega,H)}\,\mathrm{d}s \\ &\lesssim \big|t_2-t_1\big|^{\gamma}\int_0^{t_1}(t_1-s)^{-\gamma}\,\mathrm{d}s\Big(1+\sup_{s\in[0,T]}\|X(s)\|_{L^2(\Omega,H)}\Big) \lesssim \big|t_2-t_1\big|^{\gamma}, \end{aligned}$$

where the last term is bounded by (2.8).

We now turn to the third term. We recall that $q = 2/(1-\gamma)$ and $q' = 2/(1+\gamma)$. Since $\gamma < \beta$ we have

$$q'\frac{2\gamma+1-\beta}{2} = \frac{2\gamma+1-\beta}{1+\gamma} = 1 - \frac{\beta-\gamma}{1+\gamma} < 1. \tag{3.12}$$

We apply Theorem 3.5 to the third summand. Then by (2.4), (2.5) with $\varrho = \gamma$, Assumption 2.3 (iii), and (3.12), we obtain

$$\begin{aligned} &\Big\|\big(S(t_2-t_1)-I\big)\int_0^{t_1} S(t_1-s)\,\mathrm{d}W(s)\Big\|_{\mathbf{M}^{1,p,q}(H)^*} \\ &\le \big\|\big(S(t_2-t_1)-I\big)S(t_1-\cdot)\big\|_{L^{p'}(\Omega,L^{q'}([0,t_1],\mathcal{L}_2^0))} \\ &\le \big\|\big(S(t_2-t_1)-I\big)A^{-\gamma}\big\|_{\mathcal{L}}\Big(\int_0^{t_1}\big\|A^{\gamma}S(t_1-s)\big\|_{\mathcal{L}_2^0}^{q'}\,\mathrm{d}s\Big)^{\frac{1}{q'}} \\ &\lesssim \big|t_2-t_1\big|^{\gamma}\Big(\int_0^{t_1}(t_1-s)^{-q'\frac{2\gamma+1-\beta}{2}}\big\|A^{\frac{\beta-1}{2}}\big\|_{\mathcal{L}_2^0}^{q'}\,\mathrm{d}s\Big)^{\frac{1}{q'}} \lesssim \big|t_2-t_1\big|^{\gamma}. \end{aligned}$$

Next we turn to the fourth term. By applying the same arguments as for the second summand, we derive the bound

$$\begin{aligned}\Big\| \int_{t_1}^{t_2} S(t_2-s)F(X(s))\,\mathrm{d}s \Big\|_{\mathbf{M}^{1,p,q}(H)^*} &\le \int_{t_1}^{t_2} \big\| S(t_2-s)F(X(s)) \big\|_{L^2(\Omega,H)}\,\mathrm{d}s \\ &\lesssim |t_2-t_1| \Big(1 + \sup_{s\in[0,T]} \big\| X(s) \big\|_{L^2(\Omega,H)} \Big),\end{aligned}$$

since the integrand is bounded with respect to the $L^2(\Omega,H)$-norm.

Finally, a further application of Theorem 3.5 and (2.4) with $\varrho = \frac{1-\beta}{2}$ yields for the fifth summand

$$\begin{aligned}\Big\| \int_{t_1}^{t_2} S(t_2-s)\,\mathrm{d}W(s) \Big\|_{\mathbf{M}^{1,p,q}(H)^*} &\le \Big(\int_{t_1}^{t_2} \big\| S(t_2-s)A^{\frac{1-\beta}{2}} \big\|_{\mathcal{L}}^{q'} \big\| A^{\frac{\beta-1}{2}} \big\|_{\mathcal{L}_2^0}^{q'} \,\mathrm{d}s \Big)^{\frac{1}{q'}} \\ &\lesssim |t_2-t_1|^{\frac{2-q'(1-\beta)-2q'\gamma}{2q'}} |t_2-t_1|^{\gamma}.\end{aligned}$$

Now, by inserting $q' = 2/(1+\gamma)$ and since $\beta > \gamma$ we get

$$\frac{2-q'(1-\beta)-2q'\gamma}{2q'} = \frac{1}{2}(1+\gamma-(1-\beta)-2\gamma) = \frac{1}{2}(\beta-\gamma) > 0.$$

Thus it follows $|t_2-t_1|^{\frac{2-q'(1-\beta)-2q'\gamma}{2q'}} \le T^{\frac{1}{2}(\beta-\gamma)}$. This completes the proof. □

3.4. **Regularity of the numerical solution.** Here we first show a bound on the p:th-moment of the discrete solutions $X_{h,k}$ to (2.9), uniformly in $h,k \in (0,1]$, and then we prove a discrete analogy of Proposition 3.9.

Proposition 3.11. *Let Assumptions 2.3 and 2.4 hold with $\beta \in (0,1]$. Then, for every $p \ge 2$ there exists a constant C such that*

$$\max_{n\in\{0,\dots,N\}} \sup_{h,k\in(0,1]} \big\| X_{h,k}^n \big\|_{L^p(\Omega,H)} \le C.$$

Proof. Fix arbitrary $h,k \in (0,1]$. For $n \in \{1,\dots,N\}$ we recall the representation (2.10) of $X_{h,k}^n$. Hence, it follows

$$\begin{aligned}\big\| X_{h,k}^n \big\|_{L^p(\Omega,H)} &\le \big\| S_{h,k}^n P_h X_0 \big\| + k \sum_{j=0}^{n-1} \big\| S_{h,k}^{n-j} F(X_{h,k}^j) \big\|_{L^p(\Omega,H)} \\ &\quad + \Big\| \int_0^T \Big(\sum_{j=0}^{n-1} \chi_{(t_j,t_{j+1}]}(t) S_{h,k}^{n-j} \Big) \,\mathrm{d}W(t) \Big\|_{L^p(\Omega,H)}.\end{aligned}$$

By (2.12) with $\varrho = 0$ we have

$$\sup_{n\in\{1,\dots,N\}} \big\| S_{h,k}^n \big\|_{\mathcal{L}} \lesssim 1, \tag{3.13}$$

uniformly in $h,k \in (0,1]$. Therefore, by also applying Lemma 2.2,

$$\begin{aligned}\big\| X_{h,k}^n \big\|_{L^p(\Omega,H)} &\le \| X_0 \| + \sum_{j=0}^{n-1} k \big\| F(X_{h,k}^j) \big\|_{L^p(\Omega,H)} \\ &\quad + C \Big\| \sum_{j=0}^{n-1} \chi_{(t_j,t_{j+1}]} S_{h,k}^{n-j} \Big\|_{L^2([0,T],\mathcal{L}_2^0)}.\end{aligned}$$

From (2.12) and (2.13) with $\varrho = \frac{1-\beta}{2}$ we obtain

$$\Big\| \sum_{j=0}^{n-1} \chi_{(t_j,t_{j+1}]} S_{h,k}^{n-j} P_h \Big\|^2_{L^2([0,T],\mathcal{L}_2^0)} \leq k \sum_{j=0}^{n-1} \big\| S_{h,k}^{n-j} A_h^{\frac{1-\beta}{2}} \big\|^2_{\mathcal{L}} \big\| A_h^{\frac{\beta-1}{2}} P_h \big\|^2_{\mathcal{L}_2^0}$$

$$\leq Ck \sum_{j=0}^{n-1} t_{n-j}^{-1+\beta} \big\| A_h^{\frac{\beta-1}{2}} P_h A^{\frac{1-\beta}{2}} \big\|^2_{\mathcal{L}} \big\| A^{\frac{\beta-1}{2}} \big\|^2_{\mathcal{L}_2^0} \lesssim 1,$$

since $\beta \in (0,1]$. In particular this estimate is independent of $h, k \in (0,1]$.

Further, since the drift $F\colon H \to H$ satisfies a linear growth bound under Assumption 2.3 (iv) it follows

$$\big\| X_{h,k}^n \big\|_{L^p(\Omega,H)} \lesssim 1 + \|X_0\| + k \sum_{j=0}^{n-1} \big\| X_{h,k}^j \big\|_{L^p(\Omega,H)}$$

and the result follows from an application of Gronwall's Lemma 2.1. □

Proposition 3.12. *Let Assumptions 2.3 and 2.4 hold with $\beta \in (0,1]$. If $\beta \in (0,1)$ then*

$$\max_{n\in\{1,\dots,N\}} \sup_{h,k\in(0,1]} \big\| X_{h,k}^n \big\|_{\mathbf{M}^{1,p,q}} < \infty,$$

for every $2 \leq q < \frac{2}{1-\beta}$ and $2 \leq p < \infty$. If $\beta = 1$ then the same holds for every $2 \leq q \leq \infty$.

Proof. The Malliavin differentiability of $X_{h,k}^1, \dots, X_{h,k}^N$ is proved inductively using (2.10). First, for a given $n = 0, \dots, N-1$ let us note that $D_r X_{h,k}^n = 0$ for all $r \geq t_n$ since $X_{h,k}^n$ is $\mathcal{F}_r$-measurable. In addition, we have that $D_r X_{h,k}^1 = \chi_{[0,k]}(r) S_{h,k}$ from (2.9). For the inductive step let $n = 2, \dots, N-1$ and fix $r \in [t_j, t_{j+1})$ for some $j = 0, \dots, n-1$. By using (3.4) we apply the Malliavin derivative termwise to equation (2.10) and obtain

$$D_r X_{h,k}^n = S_{h,k}^{n-j} + k \sum_{i=j+1}^{n-1} S_{h,k}^{n-i} F'(X_{h,k}^i) D_r X_{h,k}^i. \tag{3.14}$$

Let $l(r) = j$ for $r \in [t_j, t_{j+1})$, $j = 0, \dots, N-1$, and $l = l(\cdot)$. By (3.13) it holds

$$\big\| DX_{h,k}^n \big\|_{L^p(\Omega,L^q([0,T],\mathcal{L}_2^0))} \leq \big\| \chi_{(0,t_n]} S_{h,k}^{n-l} \big\|_{L^q([0,T],\mathcal{L}_2^0)}$$

$$+ k \sup_{j\in\{1,\dots,N\}} \big\| S_{h,k}^j \big\|_{\mathcal{L}} |F|_{\mathcal{C}_{\mathrm{b}}^1} \sum_{i=1}^{n-1} \big\| D_r X_{h,k}^i \big\|_{L^p(\Omega,L^q([0,T],\mathcal{L}_2^0))}$$

$$\lesssim \big\| \chi_{(0,t_n]} S_{h,k}^{n-l} A^{\frac{1-\beta}{2}} \big\|_{L^q([0,T],\mathcal{L})} \big\| A^{\frac{\beta-1}{2}} \big\|_{\mathcal{L}_2^0} + k \sum_{i=1}^{n-1} \big\| D_r X_{h,k}^i \big\|_{L^p(\Omega,L^q([0,T],\mathcal{L}_2^0))}.$$

With our choice of q we have

$$\big\| \chi_{(0,t_n]} S_{h,k}^{n-l} A^{\frac{1-\beta}{2}} \big\|_{L^q([0,T],\mathcal{L})} \leq \Big(k \sum_{i=0}^{N-1} \big\| S_{h,k}^{N-i} A^{\frac{1-\beta}{2}} \big\|^q_{\mathcal{L}} \Big)^{\frac{1}{q}} \lesssim \Big(k \sum_{i=0}^{N-1} t_{N-i}^{-q\frac{1-\beta}{2}} \Big)^{\frac{1}{q}} \leq C,$$

uniformly in $h, k \in (0,1]$. By the discrete Gronwall Lemma 2.1 we conclude. □

4. Weak convergence by duality

Let X be the solution to equation (2.7) and $X_{h,k}$ be the discretization given by the recursive scheme (2.9) and take at this stage $\varphi \in \mathcal{C}^1(H, \mathbf{R})$. We start our approach to weak convergence by an application of (2.3) to get

$$\max_{n\in\{1,\dots,N\}} \big|\mathbf{E}\big[\varphi(X(t_n)) - \varphi(X^n_{h,k})\big]\big| = \max_{n\in\{1,\dots,N\}} \big|\big\langle \Phi^n_{h,k}, X(t_n) - X^n_{h,k}\big\rangle_{L^2(\Omega,H)}\big|, \tag{4.1}$$

where

$$\Phi^n_{h,k} = \int_0^1 \varphi'(\Theta^n_{h,k}(\varrho))\,\mathrm{d}\varrho \quad \text{and} \quad \Theta^n_{h,k}(\varrho) = \varrho X(t_n) + (1-\varrho)X^n_{h,k}, \tag{4.2}$$

for $n \in \{1,\dots,N\}$. This linearization was first proposed in [8] for nonlinear stochastic ordinary differential equations. Their approach after the linearization is based on duality in the sense of an adjoint equation. This approach applies also in our setting by using backward stochastic evolution equations, although the available theory of Malliavin calculus for such equations is slightly insufficient for the purpose. Independently, the linearization was applied in [28] for linear stochastic partial differential equations.

Extending the idea of [28], we proceed as follows: choose a Gelfand triple $V \subset L^2(\Omega, H) \subset V^*$ such that $\Phi^n_{h,k} \in V$. By duality we have

$$\big|\mathbf{E}\big[\varphi(X(t_n)) - \varphi(X^n_{h,k})\big]\big| \le \Big(\sup_{h,k\in(0,1]} \big\|\Phi^n_{h,k}\big\|_V\Big) \big\|X(t_n) - X^n_{h,k}\big\|_{V^*}. \tag{4.3}$$

The proof of our weak convergence result in Theorem 4.4 then amounts to showing that, for every $\gamma \in [0, \beta)$, we can find a suitable space V such that

$$\begin{aligned} \max_{n\in\{1,\dots,N\}} \sup_{h,k\in(0,1]} \big\|\Phi^n_{h,k}\big\|_V &\le C, \\ \max_{n\in\{1,\dots,N\}} \big\|X(t_n) - X^n_{h,k}\big\|_{V^*} &\le C\big(h^{2\gamma} + k^{\gamma}\big), \quad h,k \in (0,1]. \end{aligned} \tag{4.4}$$

For a comparison, the strong error converges with half this rate, i.e., for every $\gamma \in [0, \beta)$ there exist a $C > 0$ such that

$$\max_{n\in\{1,\dots,N\}} \|X(t_n) - X^n_{h,k}\|_{L^2(\Omega,H)} \le C(h^{\gamma} + k^{\frac{\gamma}{2}}), \quad h,k \in (0,1].$$

This is not to be found in the literature, although the proof is straightforward using our low regularity error estimate Assumption 2.4.

We will explain our method by gradually choosing more sophisticated spaces V and start with the simpler problem of the weak approximation of the stochastic convolution. This problem is treated in [13], [16] [25], [26], [28] and to some extent in [42]. We show that in this case $V = L^2(\Omega, \dot{H}^{\gamma})$, and $V = \mathbf{M}^{1,p,p}(H)$ with $p = \frac{2}{1-\gamma}$, suffice with different degrees of success. The proofs are simpler than in the mentioned papers, except for [28] to which the present paper is an extension. We continue with a subsection containing our main result on the case of semilinear equations with additive noise. This is where we need the space $V = \mathbf{G}^{1,p}(H)$, whose dual norm allows for a Gronwall argument in view of Lemma 3.8. Finally, we discuss multiplicative noise in Subsection 4.3 and illustrate why our approach is not yet sufficient for this generality.

In general we assume that test functions are $\mathcal{C}^2$ with polynomial growth:

Assumption 4.1. *The function $\varphi\colon H \to \mathbf{R}$ is twice continuously Fréchet differentiable and there exists an integer $m \geq 2$ such that (recall the norm in (2.2))*

$$\|\varphi^{(j)}(x)\|_{\mathcal{L}^{[j]}(H,\mathbf{R})} \leq C\big(1 + \|x\|^{m-j}\big), \quad x \in H, \ j = 1, 2. \tag{4.5}$$

4.1. **The stochastic convolution.** We consider the stochastic convolution W^A and its approximations $W_{h,k}^{A_h}$,

$$W^A(t_n) = \int_0^{t_n} S(t_n - s)\,\mathrm{d}W(s) \quad \text{and} \quad W_{h,k}^{A_h,n} = \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} S_{h,k}^{n-j}\,\mathrm{d}W(s)$$

for $n \in \{1, \dots, N\}$. For arbitrary $\gamma \in (0, \beta)$ consider first the Gelfand triple

$$L^2\big(\Omega, \dot{H}^\gamma\big) \subset L^2(\Omega, H) \subset L^2\big(\Omega, \dot{H}^{-\gamma}\big).$$

In order to have $\Phi_{h,k}^n \in L^2(\Omega, \dot{H}^\gamma)$ we need to impose an extra assumption on φ, namely that, for some $m \geq 1$ and every $\gamma \in (0, \beta)$, it holds

$$\big\|\varphi'(x)\big\|_{\dot{H}^\gamma} \leq C\Big(1 + \|x\|_{\dot{H}^\gamma}^{m-1}\Big), \quad x \in \dot{H}^\gamma. \tag{4.6}$$

Then we first get by the Sobolev regularity of W^A and $W_{h,k}^{A_h}$

$$\big\|\Phi_{h,k}^n\big\|_{L^2(\Omega,\dot{H}^\gamma)} \lesssim \big\|W^A(t_n)\big\|_{L^{2(m-1)}(\Omega,\dot{H}^\gamma)}^{m-1} + \big\|W_{h,k}^{A_h,n}\big\|_{L^{2(m-1)}(\Omega,\dot{H}^\gamma)}^{m-1} \lesssim 1,$$

uniformly in $h, k \in (0, 1]$. To prove convergence in $L^2(\Omega, \dot{H}^{-\gamma})$ we write the difference of the stochastic convolution and its numerical discretization in the form

$$W^A(t_n) - W_{h,k}^{A_h,n} = \int_0^{t_n} \tilde{E}_{h,k}(t_n - t)\,\mathrm{d}W(t), \tag{4.7}$$

where $\tilde{E}_{h,k}\colon [0, T] \to \mathcal{L}_2^0$ is given by

$$\tilde{E}_{h,k}(t) := S(t) - S_{h,k}^{j+1}, \qquad \text{for } t \in [t_j, t_{j+1}), \ j = 0, \dots, N-1. \tag{4.8}$$

Provided that this error operator satisfies

$$\big\|A^{-\frac{\gamma}{2}}\tilde{E}_{h,k}(t_n - s)A^{\frac{1-\beta}{2}}\big\|_{\mathcal{L}}^2 \lesssim \big(h^{2\gamma} + k^\gamma\big)(t_n - s)^{-1+\beta-\gamma}, \tag{4.9}$$

for all $n \in \{1, \dots, N\}$ and $s \in [0, t_n)$, we obtain by the Itō isometry and Assumption 2.3 (iii)

$$\begin{aligned}
\big\|W^A(t_n) - W_{h,k}^{A_h,n}\big\|_{L^2(\Omega,\dot{H}^{-\gamma})} &= \Big(\int_0^{t_n} \big\|A^{-\frac{\gamma}{2}}\tilde{E}_{h,k}(t_n - s)\big\|_{\mathcal{L}_2^0}^2\,\mathrm{d}s\Big)^{\frac{1}{2}} \\
&\leq \Big(\int_0^{t_n} \big\|A^{-\frac{\gamma}{2}}\tilde{E}_{h,k}(t_n - s)A^{\frac{1-\beta}{2}}\big\|_{\mathcal{L}}^2 \big\|A^{\frac{\beta-1}{2}}\big\|_{\mathcal{L}_2^0}^2\,\mathrm{d}s\Big)^{\frac{1}{2}} \\
&\lesssim \big(h^{2\gamma} + k^\gamma\big)\Big(\int_0^{t_n} (t_n - s)^{-1+\beta-\gamma}\,\mathrm{d}s\Big)^{\frac{1}{2}} \lesssim h^{2\gamma} + k^\gamma.
\end{aligned}$$

The error estimate (4.9) is verified for Galerkin finite element approximations in Section 5 for $\gamma = 0$, but the case $\gamma > 0$ is not to be found in the literature, so for this particular choice of Gelfand triple we do not work in full rigor. An integrated version of (4.9) is found in [42], details in [41], and we find no reason to doubt the validity of (4.9). In view of (4.3) and assuming (4.9) we have proved weak convergence of the same rate.

Actually, [42, Theorem 1.2] shows convergence of order $O(h^{2\beta}+k^{\beta})$ in $L^2(\Omega,\dot{H}^{-1})$ (except for a logarithmic factor). However, the fact that $L^2(\Omega,\dot{H}^{-1})$-convergence implies weak convergence for more than linear test functions was not realized in the early work [42]. Subsequent works except [28] rely on the use of Kolmogorov's equation. In the paper [16] this was done for test functions (4.6), while [13] only assumed $\varphi \in \mathcal{C}^2_{\mathrm{b}}(H,\mathbf{R})$. We also remark that the only technical ingredient used in the present proof is the Itō isometry. Therefore this proof carries over without additional difficulties to the case when the cylindrical Q-Wiener process W is replaced by a square integrable martingale M, by just introducing the suitable notation. This gives a partial extension of the results in [32] in which impulsive noise was considered. In that paper the additional assumption (4.6) was not used but instead the test functions were assumed to be in $\mathcal{C}^2_{\mathrm{b}}(H,\mathbf{R})$.

Fix $\gamma \in (0,\beta)$ and let $p = \frac{2}{1-\gamma}$. Consider next the choice of Gelfand triple

$$\mathbf{M}^{1,p,p}(H) \subset L^2(\Omega,H) \subset \mathbf{M}^{1,p,p}(H)^*.$$

With these spaces we need no assumptions on the test functions other than Assumption 4.1. We state the two parts of our weak convergence proof as two separate Lemmas. Notice that the first lemma is not restricted to the stochastic convolution.

Lemma 4.2. *Let Assumptions 2.3, 2.4, and 4.1 hold with $\beta \in (0,1]$. Then for $p = \frac{2}{1-\gamma}$ with $\gamma \in (0,\beta)$ it holds*

$$\max_{n\in\{1,\dots,N\}} \sup_{h,k\in(0,1]} \big\| \Phi^n_{h,k} \big\|_{\mathbf{M}^{1,p,p}(H)} \le C,$$

where $\Phi^n_{h,k}$ is defined in (4.2).

Proof. First note that φ' satisfies the condition of the chain rule in Lemma 3.3 with $r = m-2$ by Assumption 4.1. Thus, it holds

$$\Phi^n_{h,k} = \int_0^1 \varphi'(\Theta^n_{h,k}(\varrho))\,\mathrm{d}\varrho \in \mathbf{M}^{1,p,p}(H),$$

since $\Theta^n_{h,k}(\varrho) = \varrho X(t_n) + (1-\varrho)X^n_{h,k} \in \mathbf{M}^{1,(m-1)p,p}(H)$ by Propositions 3.9 and 3.12. Further, from Lemma 3.3 we also get

$$\begin{aligned}\big\| \Phi^n_{h,k} \big\|_{\mathbf{M}^{1,p,p}(H)} &\lesssim \big(1 + \sup_{\varrho\in[0,1]} \big\| \Theta^n_{h,k} \big\|^{m-1}_{\mathbf{M}^{1,(m-1)p,p}(H)}\big) \\ &\lesssim \big(1 + \big\| X(t_n) \big\|^{m-1}_{\mathbf{M}^{1,(m-1)p,p}(H)} + \big\| X^n_{h,k} \big\|^{m-1}_{\mathbf{M}^{1,(m-1)p,p}(H)}\big).\end{aligned}$$

By Propositions 3.9 and 3.12 this can be bounded independently of $h,k \in (0,1]$. □

Lemma 4.3. *Let Assumptions 2.3 and 2.4 hold with $\beta \in (0,1]$. For an arbitrary $\gamma \in (0,\beta)$ set $p = \frac{2}{1-\gamma}$. It holds*

$$\max_{n\in\{1,\dots,N\}} \big\| W^A(t_n) - W^{A_h,n}_{h,k} \big\|_{\mathbf{M}^{1,p,p}(H)^*} \le C\big(h^{2\gamma} + k^{\gamma}\big), \quad h,k \in (0,1].$$

Proof. By (4.7), Theorem 3.5 and Assumption 2.3 (iii) we get with $p' = \frac{2}{1+\gamma}$

(4.10)
$$\begin{aligned}\big\|W^A(t_n) - W_{h,k}^{A_h,n}\big\|_{\mathbf{M}^{1,p,p}(H)^*} &\le \Big(\int_0^{t_n} \big\|\tilde{E}_{h,k}(t_n-t)\big\|_{\mathcal{L}_2^0}^{p'}\,\mathrm{d}t\Big)^{\frac{1}{p'}} \\ &\le \Big(\int_0^{t_n} \big\|\tilde{E}_{h,k}(t_n-t)A^{\frac{1-\beta}{2}}\big\|_{\mathcal{L}}^{p'}\big\|A^{\frac{\beta-1}{2}}\big\|_{\mathcal{L}_2^0}^{p'}\,\mathrm{d}t\Big)^{\frac{1}{p'}}.\end{aligned}$$

Recalling the error operator (2.11) we obtain for $t \in (t_j, t_{j+1}]$, $j = 0, \dots, n-1$,

$$\begin{aligned}&\big\|\tilde{E}_{h,k}(t_n-t)A^{\frac{1-\beta}{2}}\big\|_{\mathcal{L}} \\ &\le \big\|\big(S(t_n-t) - S(t_n-t_j)\big)A^{\frac{1-\beta}{2}}\big\|_{\mathcal{L}} + \big\|E_{h,k}^{n-j}A^{\frac{1-\beta}{2}}\big\|_{\mathcal{L}} \\ &\le \big\|\big(I - S(t-t_j)\big)A^{-\gamma}\big\|_{\mathcal{L}}\big\|S(t_n-t)A^{\frac{2\gamma+1-\beta}{2}}\big\|_{\mathcal{L}} + \big\|E_{h,k}^{n-j}A^{\frac{1-\beta}{2}}\big\|_{\mathcal{L}} \\ &\lesssim (t_n-t)^{-\frac{2\gamma+1-\beta}{2}}\big(h^{2\gamma} + k^{\gamma}\big),\end{aligned} \tag{4.11}$$

where we applied (2.4), (2.5) and (2.14) in the last step. By recalling (3.12) the estimate in (4.10) is completed by

$$\begin{aligned}\big\|W^A(t_n) - W_{h,k}^{A_h,n}\big\|_{\mathbf{M}^{1,p}(H)^*} &\lesssim \big(h^{2\gamma} + k^{\gamma}\big)\Big(\int_0^{t_n} (t_n-t)^{-p'\frac{2\gamma+1-\beta}{2}}\,\mathrm{d}t\Big)^{\frac{1}{p'}} \\ &\lesssim h^{2\gamma} + k^{\gamma},\end{aligned}$$

where the integral is bounded independently of t_n. □

4.2. **Semilinear equations with additive noise.** Above we demonstrated that $V = \mathbf{M}^{1,p,p}(H)$ is suitable for the weak error analysis for the stochastic convolution. In order to treat semilinear equations we need an even smaller space. Here we work with the Gelfand triple

$$\mathbf{G}^{1,p}(H) \subset L^2(\Omega, H) \subset \mathbf{G}^{1,p}(H)^*.$$

The line of proof is the same as above only that the convergence in the dual norm is more involved and relies on the local Lipschitz continuity stated in Lemma 3.8, the inequality Lemma 3.5 and a classical Gronwall argument.

Theorem 4.4. *Let Assumptions 2.3 and 2.4 hold with $\beta \in (0,1]$. Let X and $X_{h,k}^N$ be the solutions to equations (2.7) and (2.9), respectively. For every function $\varphi\colon H \to H$ that satisfies Assumption 4.1 and every $\gamma \in [0,\beta)$ we have for $h,k \in (0,1]$ the weak convergence*

$$\max_{n\in\{1,\dots,N\}} \big|\mathbf{E}\big[\varphi(X(t_n)) - \varphi(X_{h,k}^n)\big]\big| \le C\big(h^{2\gamma} + k^{\gamma}\big).$$

Proof. This is a direct consequence of (4.4) and Lemma 4.5 and 4.6 below. □

Lemma 4.5. *Let the assumptions of Theorem 4.4 hold. For an arbitrary $\gamma \in (0,\beta)$ set $p = \frac{2}{1-\gamma}$. It holds*

$$\max_{n\in\{1,\dots,N\}} \sup_{h,k\in(0,1]} \big\|\Phi_{h,k}^n\big\|_{\mathbf{G}^{1,p}(H)} \le C.$$

Proof. By Lemma 4.2 we have $\|\Phi^n_{h,k}\|_{\mathbf{M}^{1,p,p}(H)} \leq C$ uniformly in n and h, k. In addition, by (2.8), Proposition 3.11 and Assumption 4.1 it holds $\|\Phi^n_{h,k}\|_{L^{2p}(\Omega,H)} \leq C$ uniformly in n and h, k. □

Lemma 4.6. *Let the assumptions of Theorem 4.4 hold. For an arbitrary $\gamma \in (0, \beta)$ set $p = \frac{2}{1-\gamma}$. Then there exists a constant $C > 0$ independent of $h, k \in (0, 1]$ such that*

$$\max_{n\in\{1,\dots,N\}} \big\|X(t_n) - X^n_{h,k}\big\|_{\mathbf{G}^{1,p}(H)^*} \leq C\big(h^{2\gamma} + k^\gamma\big), \quad h, k \in (0, 1].$$

Proof. Let $n \in \{1, \dots, N\}$ be arbitrary. Using mild solution (2.7) and its discrete counterpart (2.10), we can write

$$\begin{aligned} X(t_n) - X^n_{h,k} &= \big(S(t_n) - S^n_{h,k}\big)X_0 \\ &\quad + \sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} \big(S(t_n - t) - S^{n-j}_{h,k}\big)F(X(t))\,\mathrm{d}t \\ &\quad + \sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} S^{n-j}_{h,k}\big(F(X(t)) - F(X^j_{h,k})\big)\,\mathrm{d}t + W^A(t_n) - W^{A_h,n}_{h,k}. \end{aligned} \tag{4.12}$$

By recalling the error operators $\tilde{E}_{h,k}(t)$ from (4.8) we obtain

$$\begin{aligned} \big\|X(t_n) - X^n_{h,k}\big\|_{\mathbf{G}^{1,p}(H)^*} &\leq \big\|\big(S(t_n) - S^n_{h,k}\big)X_0\big\| \\ &\quad + \Big\|\int_0^{t_n} \tilde{E}_{h,k}(t_n - t)F(X(t))\,\mathrm{d}t\Big\|_{\mathbf{G}^{1,p}(H)^*} \\ &\quad + \Big\|\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} S^{n-j}_{h,k}\big(F(X(t)) - F(X^j_{h,k})\big)\,\mathrm{d}t\Big\|_{\mathbf{G}^{1,p}(H)^*} \\ &\quad + \big\|W^A(t_n) - W^{A_h,n}_{h,k}\big\|_{\mathbf{G}^{1,p}(H)^*}. \end{aligned} \tag{4.13}$$

By (2.14) with $\varrho = 0$ and $\theta = 2\gamma$ and Assumption 2.3 (ii) we get

$$\big\|\big(S(t_n) - S^n_{h,k}\big)X_0\big\| \leq \big\|(S(t_n) - S^n_{h,k})A^{-\frac{\gamma}{2}}\big\|_{\mathcal{L}}\big\|A^{\frac{\gamma}{2}}X_0\big\| \lesssim \big(h^{2\gamma} + k^\gamma\big)\big\|A^{\frac{\gamma}{2}}X_0\big\|.$$

For the second term in (4.13) we first use that $\|Z\|_{\mathbf{G}^{1,p}(H)^*} \leq \|Z\|_{L^2(\Omega,H)}$ for all $Z \in L^2(\Omega, H)$. Then by (4.11) with $\beta = 1$, the linear growth of F, and (2.8) we have

$$\begin{aligned} \Big\|\int_0^{t_n} \tilde{E}_{h,k}(t_n - t)F(X(t))\,\mathrm{d}t\Big\|_{\mathbf{G}^{1,p}(H)^*} &\leq \int_0^{t_n} \big\|\tilde{E}_{h,k}(t_n - t)\big\|_{\mathcal{L}}\big\|F(X(t))\big\|_{L^2(\Omega,H)}\,\mathrm{d}t \\ \lesssim \big(h^{2\gamma} + k^\gamma\big)\int_0^{t_n} (t_n - t)^{-\gamma}\,\mathrm{d}t\Big(1 + \sup_{t\in[0,T]}&\big\|X(t)\big\|_{L^2(\Omega,H)}\Big) \lesssim h^{2\gamma} + k^\gamma. \end{aligned}$$

For the third summand we first notice that Proposition 3.9 and Proposition 3.12 justify the use of Lemma 3.8 for $Y_1 = X(t)$ and $Y_2 = X^j_{h,k}$ with $t \in (t_j, t_{j+1}]$. We get

$$\begin{aligned}&\big\|F(X(t)) - F(X^j_{h,k})\big\|_{\mathbf{G}^{1,p}(H)^*}\\&\quad\le \max_{i\in\{1,2\}} |F|_{\mathcal{C}^i_{\mathrm{b}}}\Big(1 + \|DX(t)\|_{L^p([0,T],L^{2p}(\Omega,\mathcal{L}^0_2))} + \|DX^j_{h,k}\|_{L^p([0,T],L^{2p}(\Omega,\mathcal{L}^0_2))}\Big)\\&\qquad\times \|X(t) - X^j_{h,k}\|_{\mathbf{G}^{1,p}(H)^*} \lesssim \|X(t) - X^j_{h,k}\|_{\mathbf{G}^{1,p}(H)^*}.\end{aligned}$$

By (3.13) we get

$$\begin{aligned}&\Big\|\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} S^{n-j}_{h,k}\big(F(X(t)) - F(X^j_{h,k})\big)\,\mathrm{d}t\Big\|_{\mathbf{G}^{1,p}(H)^*}\\&\quad\le \sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} \max_{i\in\{1,\dots,N\}} \big\|S^i_{h,k}\big\|_{\mathcal{L}}\big\|F(X(t)) - F(X^j_{h,k})\big\|_{\mathbf{G}^{1,p}(H)^*}\,\mathrm{d}t\\&\quad\lesssim \sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} \big\|X(t) - X^j_{h,k}\big\|_{\mathbf{G}^{1,p}(H)^*}\,\mathrm{d}t.\end{aligned}\tag{4.14}$$

We split to get

$$\big\|X(t) - X^j_{h,k}\big\|_{\mathbf{G}^{1,p}(H)^*} \le \big\|X(t) - X(t_j)\big\|_{\mathbf{G}^{1,p}(H)^*} + \big\|X(t_j) - X^j_{h,k}\big\|_{\mathbf{G}^{1,p}(H)^*}$$

By the Hölder continuity Proposition 3.10 it holds $\|X(t) - X(t_j)\|_{\mathbf{G}^{1,p}(H)^*} \lesssim k^\gamma$ and therefore

$$\begin{aligned}&\Big\|\sum_{j=0}^{n-1}\int_{t_j}^{t_{j+1}} S^{n-j}_{h,k}\big(F(X(t)) - F(X^j_{h,k})\big)\,\mathrm{d}t\Big\|_{\mathbf{G}^{1,p}(H)^*}\\&\qquad\lesssim \sum_{j=0}^{n-1} k^{1+\gamma} + k\sum_{j=0}^{n-1}\big\|X(t_j) - X^j_{h,k}\big\|_{\mathbf{G}^{1,p}(H)^*}.\end{aligned}$$

The fourth summand is estimated in Lemma 4.3. Altogether we conclude that

$$\big\|X(t_n) - X^n_{h,k}\big\|_{\mathbf{G}^{1,p}(H)^*} \lesssim \big(h^{2\gamma} + k^\gamma\big) + k\sum_{j=0}^{n-1}\big\|X(t_j) - X^j_{h,k}\big\|_{\mathbf{G}^{1,p}(H)^*}.$$

By the discrete Gronwall Lemma 2.1 the assertion follows. □

Remark 4.7. In all the works we know of the rate of weak convergence is twice that of strong convergence. This can be understood from the point of view we have presented here. Let $V \subset L^2(\Omega, H) \subset V^*$ denote any of the above mentioned Gelfand triples. Although we estimate $X - X_{h,k}$ in the V^*-norm we have $X - X_{h,k} \in V$. The strong error is measured in the $L^2(\Omega, H)$-norm, exactly in the middle of V and V^* in the regularity scale, giving an intuition for the relationship between the strong and weak convergence.

Remark 4.8. A second observation is that we have used two different regularity scales, regularity in the Malliavin sense and regularity in space. Our analysis shows that exploiting the fine property of the first Malliavin derivative is similar to exploiting the spatial regularity of order β. This is somehow analogous to the parabolic scaling, where one time derivative corresponds to two space derivatives.

4.3. **Multiplicative noise.** The choice $V = \mathbf{G}^{1,p}(H)$ of Subsection 4.2 only works for equations with additive noise. We demonstrate this here by considering the following equation with multiplicative noise

$$\mathrm{d}X(t) + AX(t)\,\mathrm{d}t = F(X(t))\,\mathrm{d}t + G(X(t))\,\mathrm{d}W(t),\ t \in (0,T]; \quad X(0) = X_0.$$

Here $G \in \mathcal{C}_{\mathrm{b}}^2(H, \mathcal{L}(H_0, \dot{H}^{\beta-1}))$. In order to perform the Gronwall argument in the $\mathbf{G}^{1,p}(H)^*$-norm for this equation, one would need a bound

$$\Big\| \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} S_{h,k}^{n-j} \big(G(X(t)) - G(X_{h,k}^j) \big)\,\mathrm{d}W(t) \Big\|_{\mathbf{G}^{1,p}(H)^*} \lesssim \sum_{j=0}^{n-1} \int_{t_j}^{t_{j+1}} \big\| X(t) - X_{h,k}^j \big\|_{\mathbf{G}^{1,p}(H)^*}\,\mathrm{d}t,$$

cf. (4.14). In order to simplify the presentation, we consider proving

$$\Big\| \int_0^T \big(G(Y_1(t)) - G(Y_2(t)) \big)\,\mathrm{d}W(t) \Big\|_{\mathbf{G}^{1,p}(H)^*} \lesssim \int_0^T \big\| Y_1(t) - Y_2(t) \big\|_{\mathbf{G}^{1,p}(H)^*}\,\mathrm{d}t \tag{4.15}$$

for Y_1, Y_2 sufficiently regular. Define

$$K(t) = \int_0^1 G'(Y_2(t) + \varrho(Y_1(t)(t) - Y_2(t)))\,\mathrm{d}\varrho, \quad t \in [0,T].$$

Recall that $\|Y\|_{\mathbf{G}^{1,p}(H)^*} = \sup_{Z\in B} \langle Z, Y \rangle_{L^2(\Omega,H)}$, where $B \subset \mathbf{G}^{1,p}(H)$ is the unit ball. We integrate by parts and move the supremum inside the integral to get

$$\begin{aligned} \Big\| \int_0^T \big(G(Y_1(t)) - G(Y_2(t)) \big)\,\mathrm{d}W(t) \Big\|_{\mathbf{G}^{1,p}(H)^*} &= \sup_{Z\in B} \Big\langle Z, \int_0^T K(t) \cdot \big(Y_1(t) - Y_2(t) \big)\,\mathrm{d}W(t) \Big\rangle_{L^2(\Omega,H)} \\ &= \sup_{Z\in B} \int_0^T \big\langle D_t Z, K(t) \cdot \big(Y_1(t) - Y_2(t) \big) \big\rangle_{L^2(\Omega,\mathcal{L}_2^0)}\,\mathrm{d}t \\ &\le \int_0^T \sup_{Z\in B} \big\langle K(t)^* D_t Z, Y_1(t) - Y_2(t) \big\rangle_{L^2(\Omega,H)}\,\mathrm{d}t. \end{aligned}$$

Certainly $K(t)^* D_t Z \notin \mathbf{G}^{1,p}(H)$, since Z is not more than once Malliavin differentiable. Therefore, there is no constant C such that, for all $Y \in L^2(\Omega, H)$,

$$\sup_{Z\in B} \big\langle K(t)^* D_t Z, Y \big\rangle_{L^2(\Omega,H)} \le C \|Y\|_{\mathbf{G}^{1,p}(H)^*} = C \sup_{Z\in B} \big\langle Z, Y \big\rangle_{L^2(\Omega,H)}. \tag{4.16}$$

This shows that $V = \mathbf{G}^{1,p}(H)$ only suffices for equations with additive noise.

In view of (4.16), for any feasible choice of V, we need $\|Z\|_V \lesssim \|K(t)^* D_t Z\|_V$ uniformly in t. Therefore it is clear that V must be a space of infinitely smooth random variables. Such spaces are Fréchet spaces but not normed spaces, which makes the formulation (4.15) with $\mathbf{G}^{1,p}(H)^*$ replaced by V^* invalid. But more importantly, in view of (4.4) we need $\|\Phi_{h,k}^n\|_V \le C$ uniformly in n, h, k. We thus require infinite smoothness of the solution X. To our knowledge, there are no results on Malliavin differentiability of arbitrary order for stochastic evolution equations in the Hilbert space framework. It is clear that such a smooth solution requires $F \in \mathcal{C}_{\mathrm{b}}^{\infty}(H,H)$ and $G \in \mathcal{C}_{\mathrm{b}}^{\infty}(H,\mathcal{L}_2^0)$.

5. Approximation by the finite element method

In this section we describe a more explicit example for the linear operator A and its corresponding numerical discretization by the finite element method.

For this we consider the Hilbert space $H = L^2(D)$, where $D \subset \mathbf{R}^d$, $d = 1,2,3$, is a bounded, convex and polygonal domain. The linear operator $(A, \mathcal{D}(A))$ is defined to be $Au = \nabla \cdot (a \nabla u) + cu$ with Dirichlet boundary conditions, where $a, c \colon D \to \mathbb{R}$ are sufficiently smooth with $c(x) \geq 0$ and $a(x) \geq a_0 > 0$. Under these conditions A is an elliptic, selfadjoint, second order differential operator with compact inverse, see for instance [30, Chap. 6.1]. In particular, A satisfies Assumption 2.3 (i).

As before we measure spatial regularity in terms of the spaces $\dot{H}^\theta$, $\theta \in \mathbf{R}$, which now correspond to classical Sobolev spaces, for example $\dot{H}^1 = H_0^1(D)$ and $\dot{H}^2 = H_0^1(D) \cap H^2(D)$. For more details we refer to [28, App. B.2] and the references therein.

Let $(T_h)_{h\in(0,1]}$ be a regular family of triangulations of D with maximal mesh size $h \in (0,1]$. We define a family of subspaces $(V_h)_{h\in(0,1]}$ of $\dot{H}^1$, consisting of continuous piecewise linear functions corresponding to $(T_h)_{h\in(0,1]}$. By equipping the space $\dot{H}^1$ with the inner product $\langle \cdot,\cdot \rangle_1 := \langle A^{\frac12}\cdot, A^{\frac12}\cdot\rangle$, we define $A_h \colon V_h \to V_h$, $h \in (0,1]$, to be the linear operators given by the relationship

$$\langle A_h v_h, u_h \rangle = \langle v_h, u_h \rangle_1, \quad \forall v_h, u_h \in V_h.$$

Now, from [28, (3.15)] we get $\|A_h^{-1} P_h x\| \leq \|x\|_{-1}$ for all $x \in \dot{H}^{-1}$. Hence it holds

$$\|A_h^{-\frac12} P_h A^{\frac12}\|_{\mathcal{L}} \leq 1.$$

An interpolation between this and $\|P_h\|_{\mathcal{L}} \leq 1$ yields (2.13) for $\varrho \in [0,1]$.

As in Subsection 2.3 we denote by $(S(t))_{t\geq 0}$ the semigroup generated by $-A$ and $S_{h,k} := (I + kA_h)^{-1} P_h$. The standard literature on finite element methods, for instance [37], provides error estimates for the approximation of the semigroup with smooth and nonsmooth initial data. More precisely, it holds for the error operator (4.8)

$$\|\tilde{E}_{h,k}(t)x\| \leq C\big(h^2 + k\big) t^{-\frac{2-q}{2}} \|x\|_{\dot{H}^q}, \quad x \in \dot{H}^q,\ q = 0,2.$$

For the purpose of the present work we need to extend this to less regular initial data. This is done by the next lemma, which is a consequence of [28, Lemma 3.12].

Lemma 5.1. *Under the above assumptions and for $0 \leq \theta \leq 2$ and $0 \leq \varrho \leq \min(1, 2-\theta)$, the following estimate holds true*

$$\|\tilde{E}_{h,k}(t)x\| \leq C\big(h^\theta + k^{\frac{\theta}{2}}\big) t^{-\frac{\theta+\varrho}{2}} \|x\|_{-\varrho}, \quad x \in \dot{H}^{-\varrho},\ t > 0,\ h,k \in (0,1].$$

Proof. By [28, Lemma 3.12 (i)] the estimate

$$\|\tilde{E}_{h,k}(t)x\| \leq C\big(h^\theta + k^{\frac{\theta}{2}}\big) t^{-\frac{\theta}{2}} \|x\|, \quad t > 0,\ 0 \leq \theta \leq 2, \tag{5.1}$$

holds for all $h,k \in (0,1]$. By [28, Lemma 3.12 (iii)] the error operator $\tilde{E}_{h,k}$ also satisfies for $1 \le \theta \le 2$

$$\|\tilde{E}_{h,k}(t)x\| \le C\big(h^\theta + k^{\frac{\theta}{2}}\big)t^{-1}\|x\|_{-(2-\theta)}, \quad t > 0. \tag{5.2}$$

Interpolation of (5.1) and (5.2) with fixed $1 \le \theta \le 2$ gives that, for $\lambda \in [0,1]$,

$$\begin{aligned}\|\tilde{E}_{h,k}(t)x\| &\le C\big(h^\theta + k^{\frac{\theta}{2}}\big)t^{-(1-\lambda)\frac{\theta}{2}}t^{-\lambda}\|x\|_{-\lambda(2-\theta)}\\ &= C\big(h^\theta + k^{\frac{\theta}{2}}\big)t^{-\frac{\theta}{2}-\frac{\lambda(2-\theta)}{2}}\|x\|_{-\lambda(2-\theta)}, \quad t > 0.\end{aligned}$$

If we let $\varrho = \lambda(2-\theta)$, then we get the following estimate: for $1 \le \theta \le 2$ and $0 \le \varrho \le 2-\theta$

$$\|\tilde{E}_{h,k}(t)x\| \le C\big(h^\theta + k^{\frac{\theta}{2}}\big)t^{-\frac{\theta+\varrho}{2}}\|x\|_{-\varrho}, \quad t \ge 0. \tag{5.3}$$

By [28, Lemma 3.12 (ii)] it holds

$$\|\tilde{E}_{h,k}(t)x\| \le Ct^{-\frac{\varrho}{2}}\|x\|_{-\varrho}, \quad t > 0,\ 0 \le \varrho \le 1, \tag{5.4}$$

and using (5.3) with $\theta = 1$ and (5.4), both with the same $0 \le \varrho \le 1$, yields

$$\begin{aligned}\|\tilde{E}_{h,k}(t)x\| &= \|\tilde{E}_{h,k}(t)x\|^\lambda\|\tilde{E}_{h,k}(t)x\|^{1-\lambda} \le C\big(h + k^{\frac{1}{2}}\big)^\lambda t^{-\frac{\lambda+\varrho}{2}}\|x\|_{-\varrho}\\ &\le C\big(h^\lambda + k^{\frac{\lambda}{2}}\big)t^{-\frac{\lambda+\varrho}{2}}\|A^{-\frac{\varrho}{2}}x\|, \quad t > 0,\ 0 \le \lambda \le 1.\end{aligned} \tag{5.5}$$

Combining (5.3) and (5.5) concludes the proof. □

Writing the statement of the lemma in operator form reads

$$\|\tilde{E}_{h,k}(t)A^{\frac{\varrho}{2}}\|_{\mathcal{L}} \le C\big(h^\theta + k^{\frac{\theta}{2}}\big)t^{-\frac{\theta+\varrho}{2}}, \quad t > 0,\ 0 \le \theta \le 2,\ 0 \le \varrho \le \min(1, 2-\theta).$$

In particular, this shows (2.14) for the finite element method. To verify Assumption 2.4 it remains to show (2.12). By [28, (3.42)]

$$\|S^n_{h,k}x\| \le Ct^{-\frac{1}{2}}\|x\|_{-1}.$$

Interpolating between this and $\|S^n_{h,k}x\| \le C\|x\|$ yields (2.12).

Acknowledgement. The authors wish to thank Dr. Mihály Kovács for fruitful discussions, under the preparation of the work [1], which led to true improvements of the present paper.

The first two authors also acknowledge the kind support by Prof. Wolf-Jürgen Beyn, Prof. Barbara Gentz and the DFG-funded CRC 701 'Spectral Structures and Topological Methods in Mathematics' by making possible an inspiring research stay at Bielefeld University, where part of this work has been written.

References

[1] A. Andersson, M. Kovacs, and S. Larsson, *Weak and strong error analysis for semilinear stochastic Volterra equations*, 2014. (in preparation).

[2] A. Andersson and S. Larsson, *Weak convergence for a spatial approximation of the nonlinear stochastic heat equation*, ArXiv Preprint, arXiv:1212.5564 (2012).

[3] V. V. Anh, W. Grecksch, and J. Yong, *Regularity of backward stochastic Volterra integral equations in Hilbert spaces*, Stoch. Anal. Appl. **29** (2011), no. 1, 146–168.

[4] F. E. Benth, T. Deck, and J. Potthoff, *A white noise approach to a class of non-linear stochastic heat equations*, J. Funct. Anal. **146** (1997), no. 2, 382–415.

[5] C. E. Brehier, *Approximation of the invariant measure with an Euler scheme for stochastic PDE's driven by space-time white noise*, ArXiv Preprint, arXiv:1202.2707 (2012).

[6] ———, *Strong and weak order in averaging for SPDEs*, ArXiv Preprint, arXiv:1202.2708 (2012).

[7] C. E. Brehier and M. Kopec, *Approximation of the invariant law of spdes: error analysis using a poisson equation for a full-discretization scheme*, ArXiv Preprint, arXiv:1311.7030 (2013).

[8] E. Clément, A. Kohatsu-Higa, and D. Lamberton, *A duality approach for the weak approximation of stochastic differential equations*, Ann. Appl. Probab. **16** (2006), no. 3, 1124–1154.

[9] D. Cohen and M. Sigg, *Convergence analysis of trigonometric methods for stiff second-order stochastic differential equations*, Numer. Math. **121** (2012), no. 1, 1–29.

[10] G. Da Prato and J. Zabczyk, *Stochastic equations in infinite dimensions*, Encyclopedia of Mathematics and its Applications, vol. 44, Cambridge University Press, Cambridge, 1992.

[11] 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.

[12] A. Debussche, *Weak approximation of stochastic partial differential equations: the nonlinear case*, Math. Comp. **80** (2011), no. 273, 89–117.

[13] A. Debussche and J. Printems, *Weak order for the discretization of the stochastic heat equation*, Math. Comp. **78** (2009), no. 266, 845–863.

[14] C. M. Elliott and S. Larsson, *Error estimates with smooth and nonsmooth data for a finite element method for the Cahn-Hilliard equation*, Math. Comp. **58** (1992), no. 198, 603–630, S33–S36.

[15] M. Fuhrman and G. Tessitore, *Nonlinear Kolmogorov equations in infinite dimensional spaces: the backward stochastic differential equations approach and applications to optimal control*, Ann. Probab. **30** (2002), 1397–1465.

[16] M. Geissert, M. Kovács, and S. Larsson, *Rate of weak convergence of the finite element method for the stochastic heat equation with additive noise*, BIT **49** (2009), 343–356.

[17] 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), no. 1, 31–49.

[18] E. Hausenblas, *Approximation for semilinear stochastic evolution equations*, Potential Anal. **18** (2003), no. 2, 141–186.

[19] ———, *Weak approximation of the stochastic wave equation*, J. Comput. Appl. Math. **235** (2010), no. 1, 33–58.

[20] H. Holden, B. Øksendal, J. Ubøe, and T. Zhang, *Stochastic partial differential equations*, Second, Universitext, Springer, New York, 2010. A modeling, white noise functional approach.

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

[22] A. Jentzen and P. E. Kloeden, *A unified existence and uniqueness theorem for stochastic evolution equations*, Bull. Aust. Math. Soc. **81** (2010), no. 1, 33–46.

[23] A. Jentzen and M. Röckner, *Regularity analysis for stochastic partial differential equations with nonlinear multiplicative trace class noise*, J. Differential Equations **252** (2012), no. 1, 114–136.

[24] A. Kohatsu-Higa, *Weak approximations. A Malliavin calculus approach*, Math. Comp. **70** (2001), no. 233, 135–172.

[25] M. Kovács, S. Larsson, and F. Lindgren, *Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise*, BIT Numer. Math. **52** (2012), no. 1, 85–108.

[26] 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 Numer. Math. **53** (2013), no. 2, 497–525.

[27] R. Kruse, *Optimal error estimates of Galerkin finite element methods for stochastic partial differential equations with multiplicative noise*, IMA J. Numer. Anal. (2013).

[28] ———, *Strong and weak approximation of stochastic evolution equations*, Lecture Notes in Mathematics, vol. 2093, Springer, 2014.

[29] R. Kruse and S. Larsson, *Optimal regularity for semilinear stochastic partial differential equations with multiplicative noise*, Electron. J. Probab. **17** (2012), no. 65, 1–19.

[30] S. Larsson and V. Thomée, *Partial differential equations with numerical methods*, Texts in Applied Mathematics, vol. 45, Springer-Verlag, Berlin, 2003.

[31] J. A. León and D. Nualart, *Stochastic evolution equations with random generators*, Ann. Probab. **26** (1998), no. 1, 149–186.

[32] F. Lindner and R. L. Schilling, *Weak order for the discretization of the stochastic heat equation driven by impulsive noise*, Potential Anal. **38** (2012), no. 2, 345–179.

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

[34] A. Pazy, *Semigroups of linear operators and applications to partial differential equations*, Applied Mathematical Sciences, vol. 44, Springer, New York, 1983.

[35] C. Prévôt and M. Röckner, *A Concise Course on Stochastic Partial Differential Equations*, Lecture Notes in Mathematics, vol. 1905, Springer, Berlin, 2007.

[36] M. Sanz-Solé, *Malliavin Calculus: with Applications to Stochastic Partial Differential Equations*, 1. ed., Fundamental sciences : Mathematics, EPFL Press, Lausanne, Switzerland, 2005.

[37] V. Thomée, *Galerkin finite element methods for parabolic problems*, Second, Springer Series in Computational Mathematics, vol. 25, Springer-Verlag, Berlin, 2006.

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

[39] J. B. Walsh, *An introduction to stochastic partial differential equations*, École d'été de probabilités de Saint-Flour, XIV—1984, 1986, pp. 265–439.

[40] 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. **398** (2013), no. 1, 151–169.

[41] Y. Yan, *Error Analysis and Smoothing Properties of Discretized Deterministic and Stochastic Parabolic Problems*, Ph.D. Thesis, 2003. `http://www.math.chalmers.se/Math/Research/Preprints/Doctoral/2003/3.pdf`.

[42] ———, *Galerkin finite element methods for stochastic parabolic partial differential equations*, SIAM J. Numer. Anal. **43** (2005), no. 4, 1363–1384.

Adam Andersson, Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-412 96 Gothenburg, Sweden

E-mail address: `adam.andersson@chalmers.se`

Raphael Kruse, ETH Zürich, Seminar für Angewandte Mathematik, Rämistrasse 101, CH-8092 Zürich, Switzerland

E-mail address: `raphael.kruse@sam.math.ethz.ch`

Stig Larsson, Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-412 96 Gothenburg, Sweden

E-mail address: `stig@chalmers.se`

Recent Research Reports

Nr.	Authors/Title
2013-40	S. Mishra and F. Fuchs and A. McMurry and N.H. Risebro Explicit and implicit finite volume schemes for radiation mhd and the effects of radiation on wave propagation in stratified atmospheres.
2013-41	J. Ernest and P. LeFloch and S. Mishra Schemes with Well controlled Dissipation (WCD) I: Non-classical shock waves
2013-42	S. Mishra and N. Risebro and F. Weber Convergence rates of finite difference schemes for the wave equation with rough coeffiicients
2013-43	U. Koley and S. Mishra and N. Risebro and F. Weber Robust finite difference schemes for a nonlinear variational wave equation modeling liquid crystals
2013-44	G. Coclite and S. Mishra and N. Risebro and F. Weber Analysis and Numerical approximation of Brinkman regularization of two-phase flows in porous media
2013-45	M. Hutzenthaler and A. Jentzen Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients
2013-46	G. Da Prato and A. Jentzen and M. Röckner A mild Itô formula for SPDEs
2013-47	P. Grohs and S. Keiper and G. Kutyniok and M. Schaefer Parabolic Molecules: Curvelets, Shearlets, and Beyond
2013-48	R. Hiptmair and C. Jerez-Hanckes and C. Urzua Optimal Operator Preconditioning for Boundary Elements on Open Curves
2013-49	P. Grohs and S. Hosseini ε-Subgradient Algorithms for Locally Lipschitz Functions on Riemannian Manifolds