

An Antithetic Multilevel Monte Carlo-Milstein Scheme for Stochastic Partial Differential Equations

A. Haji-Ali and A. Stein

Research Report No. 2023-33
July 2023

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

An Antithetic Multilevel Monte Carlo-Milstein Scheme for Stochastic Partial Differential Equations

Abdul-Lateef Haji-Ali ^{*} Andreas Stein [†]

July 26, 2023

Abstract

We present a novel multilevel Monte Carlo approach for estimating quantities of interest for stochastic partial differential equations (SPDEs). Drawing inspiration from [17], we extend the antithetic Milstein scheme for finite-dimensional stochastic differential equations to Hilbert space-valued SPDEs. Our method has the advantages of both Euler and Milstein discretizations, as it is easy to implement and does not involve intractable Lévy area terms. Moreover, the antithetic correction in our method leads to the same variance decay in a MLMC algorithm as the standard Milstein method, resulting in significantly lower computational complexity than a corresponding MLMC Euler scheme. Our approach is applicable to a broader range of non-linear diffusion coefficients and does not require any commutative properties. The key component of our MLMC algorithm is a truncated Milstein-type time stepping scheme for SPDEs, which accelerates the rate of variance decay in the MLMC method when combined with an antithetic coupling on the fine scales. We combine the truncated Milstein scheme with appropriate spatial discretizations and noise approximations on all scales to obtain a fully discrete scheme and show that the antithetic coupling does not introduce an additional bias.

Keywords: Stochastic Partial Differential Equations, Multilevel Monte Carlo, Milstein Scheme, Variance Reduction, Antithetic Variates.

Subject classification: 65C05, 65C30, 65M12.

1 Introduction

Stochastic partial differential equations (SPDEs) are encountered in a range of applications spanning natural sciences, engineering, and finance. Examples include stochastic epidemic compartment models [27] and the valuation of forward contracts in interest rate or energy markets [12, 10, 3]. However, a common challenge in these applications is that SPDEs do not possess a closed-form solution and must therefore be approximated numerically. Fortunately, numerous numerical schemes for approximating various types of SPDEs have been established. A non-exhaustive list of references [21, 2, 22, 4, 14, 25, 23, 19, 1, 8] provide strong approximation results, while others [13, 25, 24, 9] offer a weak error analysis.

Once the 'pathwise' approximations are obtained, they are utilized in sampling-based approaches to estimate specific quantities of interest within the SPDE model. Monte Carlo (MC) methods are a common choice for this purpose. However, due to the low regularity of the model, standard MC approaches may become prohibitively expensive even for comparatively simple SPDEs. In addition, higher-order schemes for discretizing the stochastic space, such as stochastic Galerkin or Quasi-Monte Carlo methods, are not feasible due to the limited regularity of the model. Thus, the multilevel Monte Carlo (MLMC) method [15] seems to be the only viable option for accelerating the estimation of expectations for SPDEs. This approach has been studied in the context of SPDEs in [6, 5, 16, 26].

^{*}Maxwell Institute, Heriot-Watt University, Edinburgh. Email: a.hajiali@hw.ac.uk

[†]Seminar for Applied Mathematics, ETH Zürich. Email: andreas.stein@sam.math.ethz.ch

One common drawback of the schemes presented in [6, 5, 26] is that they rely on a simple Euler discretization in time, which leads to slow temporal convergence rates. In contrast, the authors of [16] propose a MLMC-Milstein scheme that uses a finite difference approximation in space to accelerate temporal convergence. However, their SPDE model is considerably simplified, as it is only driven by a one-dimensional Brownian motion. Consequently, it is not necessary to simulate Lévy area terms. The simulation of these terms is already a substantial issue for two-dimensional stochastic differential equations (SDEs) without certain commutativity conditions on the diffusion term. Moreover, the problem is exacerbated for infinite-dimensional driving noise, which is the natural setting for SPDEs.

1.1 Contributions

The objective of this research article is to address the previously mentioned issues by introducing an *antithetic multilevel Monte Carlo-Milstein scheme* for parabolic SPDEs. Our work is based on the antithetic MLMC scheme for SDEs presented in [17] and offers several advantages. Firstly, under natural assumptions, our scheme achieves higher-order convergence rates, similar to those of the 'standard' Milstein scheme. Secondly, the antithetic approach eliminates the need to sample Lévy area terms, making the scheme easy to implement. Our complexity analysis demonstrates that the proposed MLMC algorithm can significantly reduce computational time by several orders of magnitude. Finally, we extend the results for SDEs from [17] by allowing for unbounded, random initial conditions and not requiring a global Lipschitz condition on the Milstein correction term.

1.2 Outline

The article is structured as follows: first, in Section 2, we provide the necessary notation and background on functional analysis, infinite-dimensional Wiener processes, and parabolic SPDEs. In Section 3, we propose discretization methods for the spatial, stochastic, and temporal domains of the SPDE. The main contribution of our paper is presented in Section 4, where we introduce the antithetic Milstein scheme and prove its expected variance decay in Theorem 4.1. We then analyze the complexity of the associated antithetic MLMC Milstein scheme in Section 5 and present numerical experiments in Section 6 to complement our theoretical analysis. All proofs are provided in an appendix for clarity.

2 Preliminaries

2.1 Basic Notation

Let $(\mathcal{Y}, \|\cdot\|_{\mathcal{Y}})$ and $(\mathcal{Z}, \|\cdot\|_{\mathcal{Z}})$ be two Banach spaces. The Borel σ -algebra of \mathcal{Y} is generated by the open sets in \mathcal{Y} and denoted by $\mathcal{B}(\mathcal{Y})$. We further denote by $\mathcal{L}(\mathcal{Y}, \mathcal{Z})$ and $\mathcal{L}(\mathcal{Y})$ the set of linear bounded operators $O : \mathcal{Y} \rightarrow \mathcal{Z}$ and $O : \mathcal{Y} \rightarrow \mathcal{Y}$, respectively. For any (bounded or unbounded) operator $O : \mathcal{Y} \rightarrow \mathcal{Z}$, we denote its adjoint by $O^* : \mathcal{Z} \rightarrow \mathcal{Y}$. Let $\mathcal{Y}_0 \subseteq \mathcal{Y}$ be an open subset and let $F : \mathcal{Y} \rightarrow \mathcal{Z}$ be a twice Fréchet differentiable mapping on \mathcal{Y}_0 . The first two Fréchet derivatives of F are given by $F' : \mathcal{Y}_0 \rightarrow \mathcal{L}(\mathcal{Y}, \mathcal{Z})$ and $F'' : \mathcal{Y}_0 \rightarrow \mathcal{L}(\mathcal{Y}, \mathcal{L}(\mathcal{Y}, \mathcal{Z})) \simeq \mathcal{L}(\mathcal{Y} \times \mathcal{Y}, \mathcal{Z})$. For the remainder of this article, $C > 0$ denotes a generic positive constant which may change from one line to another. The dependency of C on certain parameters is made explicit if necessary.

2.2 Hilbert-Schmidt Operators and RKHS

Throughout this article, we consider two separable Hilbert spaces $(U, (\cdot, \cdot)_U)$ and $(H, (\cdot, \cdot)_H)$. The space of *Hilbert-Schmidt operators* on U is given by

$$\mathcal{L}_{HS}(U, H) := \{O \in \mathcal{L}(U, H) \mid \|O\|_{\mathcal{L}_{HS}(U, H)}^2 := \sum_{k \in \mathbb{N}} \|Ou_k\|_H^2 < +\infty\},$$

where $(u_k, k \in \mathbb{N})$ is some orthonormal basis of U . Recall that $(\mathcal{L}_{HS}(U, H), \|\cdot\|_{\mathcal{L}_{HS}(U, H)})$ is separable, while this is in general not true for $\mathcal{L}(U, H)$. Further, $\mathcal{L}_{HS}(U, H)$ is a Hilbert space equipped with the tensor product

$$(O_1, O_2)_{\mathcal{L}_{HS}(U, H)} := \sum_{k \in \mathbb{N}} (O_1 u_k, O_2 u_k)_H, \quad O_1, O_2 \in \mathcal{L}_{HS}(U, H).$$

The tensor product of U and H is denoted by $(U \otimes H, (\cdot, \cdot)_{U \otimes H})$. For $\phi \in U$ and $\psi \in H$ we associate to $\phi \otimes \psi \in U \otimes H$ the rank one operator $O_{\phi, \psi} \in \mathcal{L}_{HS}(U, H)$, such that $O_{\phi, \psi} u = (\phi, u)_U \psi$ for all $u \in U$. Thus, we use the identification $U \otimes H \simeq \mathcal{L}_{HS}(U, H)$, as $U \otimes H$ and $\mathcal{L}_{HS}(U, H)$ are isometrically isomorphic.

We denote by $\mathcal{L}_1(U)$ the space of all trace class operators on U , and by $\mathcal{L}_1^+(U)$ the subset of all non-negative, self-adjoint operators on U with finite trace. The trace of $Q \in \mathcal{L}_1^+(U)$ is denoted by $\text{Tr}(Q) < \infty$. For any $Q \in \mathcal{L}_1^+(U)$, the Hilbert-Schmidt theorem yields that the ordered eigenvalues $\eta_1 \geq \eta_2 \geq \dots \geq 0$ are non-negative with zero as only accumulation point, and the corresponding eigenfunctions $(e_k, k \in \mathbb{N}) \subset U$ form an orthonormal basis of U . The *square-root* of $Q \in \mathcal{L}_1^+(U)$ is defined via

$$Q^{1/2} \phi := \sum_{k \in \mathbb{N}} \sqrt{\eta_k} (\phi, e_k)_U e_k, \quad \phi \in U.$$

Since $Q^{1/2}$ is not necessarily injective, the *pseudo-inverse* of $Q^{1/2}$ is given by

$$Q^{-1/2} \varphi := \phi, \quad \text{if } Q^{1/2} \phi = \varphi \quad \text{and} \quad \|\phi\|_U = \inf \left\{ \|\varphi\|_U : \varphi \in U \text{ is such that } Q^{1/2} \varphi = \phi \right\}.$$

We define *reproducing kernel Hilbert space* (RKHS) associated to Q as the set $\mathcal{U} := Q^{1/2}(U)$ equipped with the scalar-product

$$(\varphi_1, \varphi_2)_{\mathcal{U}} := (Q^{-1/2} \varphi_1, Q^{-1/2} \varphi_2)_U, \quad \varphi_1, \varphi_2 \in \mathcal{U}.$$

Note that $(\sqrt{\eta_k} e_k, k \in \mathbb{N})$ forms an orthonormal system in \mathcal{U} , hence

$$\|O\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 = \sum_{k \in \mathbb{N}} \eta_k \|O e_k\|_H^2, \quad O \in \mathcal{L}_{HS}(\mathcal{U}, H).$$

2.3 Martingales on Hilbert Spaces

We consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t, t \geq 0))$ with normal filtration and a finite time interval $\mathbb{T} = [0, T]$. The Lebesgue-Bochner space of all p -integrable, H -valued random variables is given as

$$L^p(\Omega; H) := \left\{ Y : \Omega \rightarrow H \text{ is measurable with } \|Y\|_{L^p(\Omega; H)} := \mathbb{E}(\|Y\|_H^p)^{1/p} < \infty \right\}, \quad p \in [1, \infty).$$

Solutions to stochastic partial differential equations (SPDEs) are defined as predictable H -valued processes. The *predictable σ -algebra* $\mathcal{P}_{\mathbb{T}}$ is the smallest σ -field on $\Omega \times \mathbb{T}$ containing all sets of the form $\mathcal{A} \times (s, t]$, where $\mathcal{A} \in \mathcal{F}_s$ and $s, t \in \mathbb{T}$ with $s < t$. An H -valued stochastic process $Y : \Omega \times \mathbb{T} \rightarrow H$ is called *predictable* if it is a $\mathcal{P}_{\mathbb{T}}/\mathcal{B}(H)$ -measurable mapping. The set of all square-integrable, H -valued predictable processes is denoted by

$$\mathcal{X}_{\mathbb{T}} := \left\{ X : \Omega \times \mathbb{T} \rightarrow H \mid X \text{ is predictable and } \sup_{t \in \mathbb{T}} \mathbb{E}(\|X(t)\|_H^2) < \infty \right\}. \quad (1)$$

All appearing equalities and estimates involving stochastic terms are in the path-wise sense and are assumed to hold almost surely, thus we omit the stochastic argument $\omega \in \Omega$ for notational convenience.

Definition 2.1. Let $(e_k, k \in \mathbb{N})$ be an arbitrary orthonormal basis of U and denote $\mathcal{M}^2(U)$ the set of all square-integrable, U -valued martingales.

1. For $Y \in \mathcal{M}^2(U)$, denote by $\langle Y, Y \rangle : \Omega \times \mathbb{T} \rightarrow \mathbb{R}$ the unique predictable (quadratic variation) process, such that $\mathbb{T} \ni t \mapsto \|Y(t)\|_U^2 - \langle Y, Y \rangle_t$ is a real-valued martingale. The covariation of two martingales $Y, Z \in \mathcal{M}^2(U)$ is given by the polarization identity

$$\langle Y, Z \rangle := \frac{1}{2} (\langle Y + Z, Y + Z \rangle - \langle Y, Y \rangle - \langle Z, Z \rangle).$$

2. The *operator-valued angle bracket process* $\langle\langle Y, Y \rangle\rangle : \Omega \times \mathbb{T} \rightarrow \mathcal{L}_1^+(U)$ of $Y \in \mathcal{M}^2(U)$ is defined as

$$\langle\langle Y, Y \rangle\rangle : \Omega \times \mathbb{T} \rightarrow \mathcal{L}_{HS}(U), \quad t \mapsto \sum_{k, l \in \mathbb{N}} \langle (Y(\cdot), e_k)_U, (Y(\cdot), e_l)_U \rangle_t e_k \otimes e_l.$$

It holds that $\langle\langle Y, Y \rangle\rangle$ is the unique process such that $\mathbb{T} \ni t \mapsto Y(t) \otimes Y(t) - \langle\langle Y, Y \rangle\rangle_t$ is an $\mathcal{L}_1(U)$ -valued martingale. Further, there exists a unique process $\mathcal{Q}_Y : \Omega \times \mathbb{T} \rightarrow \mathcal{L}_1^+(U)$, called the *martingale covariance* of Y , such that

$$\langle\langle Y, Y \rangle\rangle_t = \int_0^t \mathcal{Q}_Y(s) d\langle Y, Y \rangle_s, \quad t \in \mathbb{T}, \quad (2)$$

see e.g. [30, Theorem 8.2/Definition 8.3]. We consider H -valued stochastic integrals $\int_0^t G(s) dY(s)$ with predictable, operator-valued integrands $G : \Omega \times \mathbb{T} \mapsto \mathcal{L}(U, H)$ such that $G \circ \mathcal{Q}_Y^{1/2} : \Omega \times \mathbb{T} \mapsto \mathcal{L}_{HS}(U, H)$.

2.4 Wiener Process on a Hilbert Space

Definition 2.2. [31, Definition 2.1.9] Let $Q \in \mathcal{L}_1^+(U)$. A U -valued stochastic process $W = (W(t), t \in \mathbb{T})$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *Q-Wiener process* if

- $W(0) = 0$,
- W has \mathbb{P} -almost surely continuous trajectories,
- W has independent increments, and
- for all $0 \leq s \leq t \leq T$ there holds that $W(t) - W(s) \sim \mathcal{N}(0, (t-s)Q)$.

For any Q -Wiener process there holds the identity

$$\mathbb{E}(\langle W(t) - \mathbb{E}(W(t)), \phi \rangle_U \langle W(t) - \mathbb{E}(W(t)), \psi \rangle_U) = t \langle Q\phi, \psi \rangle_U, \quad \phi, \psi \in U, t \in \mathbb{T}.$$

It follows that $\langle W, W \rangle_t = t \operatorname{Tr}(Q)$ and $\langle\langle W, W \rangle\rangle_t = tQ$ (note that $\mathcal{Q}_Y = Q \operatorname{Tr}(Q)^{-1}$ in (2) is constant with respect to t in this case). Further, recall that W admits the *Karhunen-Loève expansion*

$$W(t) = \sum_{k \in \mathbb{N}} \langle W(t), e_k \rangle_U e_k \stackrel{d}{=} \sum_{k \in \mathbb{N}} \sqrt{\eta_k} w_k(t) e_k, \quad t \in \mathbb{T}, \quad (3)$$

where the relation $\stackrel{d}{=}$ signifies equality in distribution and $(w_k, k \in \mathbb{N})$ is a sequence of real-valued and independent standard Brownian motions.

2.5 Stochastic Partial Differential Equations

We consider the stochastic partial differential equation (SPDE)

$$dX(t) = (AX(t) + F(X(t)))dt + G(X(t))dW(t), \quad X(0) = X_0, \quad (4)$$

where $A : D(A) \subset H \rightarrow H$ is a densely defined and unbounded linear (differential) operator. The initial value X_0 is a H -valued random variable, W is a Q -Wiener process, and the coefficients F and G in

Eq. (4) are (possibly) non-linear measurable mappings $F : H \rightarrow H$ and $G : H \rightarrow \mathcal{L}_{HS}(\mathcal{U}, H)$, respectively. Throughout this article we will assume that $(-A)$ is self-adjoint, positive definite and boundedly invertible. Consequently, the eigenvalues $(\lambda_n, n \in \mathbb{N})$ of $(-A)$ are positive, non-decreasing and only accumulate at infinity, with the corresponding eigenfunctions $(f_n, n \in \mathbb{N})$ spanning an orthonormal basis of H .

By the Hille-Yosida Theorem, A is the generator of an analytic semigroup $S = (S(t), t \geq 0) \subset \mathcal{L}(H)$ (see e.g. [25, Appendix B.2]). The fractional powers of $(-A)$, given by

$$(-A)^\alpha v := \sum_{n \in \mathbb{N}} \lambda_n^\alpha (v, f_n)_H f_n \quad v \in H,$$

are well-defined for any $\alpha \in \mathbb{R}$. Moreover, $(-A)^\alpha : D((-A)^\alpha) \rightarrow H$ is a closed operator, with $D((-A)^\alpha)$ being dense in H (see e.g. [29, Chapter 2, Theorem 6.8]). We define the Hilbert space $\dot{H}^\alpha := D((-A)^{\alpha/2})$ equipped with the inner product $(\cdot, \cdot)_\alpha := ((-A)^{\frac{\alpha}{2}} \cdot, (-A)^{\frac{\alpha}{2}} \cdot)_H$, which will in turn be used to quantify smoothness of solutions to (4).

Example 2.3. Let $H = L^2(\mathcal{D})$ for on a bounded, convex domain $\mathcal{D} \subset \mathbb{R}^d$, and let $A = \Delta$ be the Laplace operator with zero Dirichlet boundary conditions on \mathcal{D} . It then holds that $\dot{H}^2 = D((-A)) = H^2(\mathcal{D}) \cap H_0^1(\mathcal{D})$. More generally, it holds for $\alpha \in [1, 2]$ that $\dot{H}^\alpha = D((-A)^{\alpha/2}) = H^\alpha(\mathcal{D}) \cap H_0^1(\mathcal{D})$, see [11, Proposition 4.1].

We formulate suitable, but natural assumptions on the initial value and the coefficients of the SPDE (4) in the following. We also repeat the above conditions on A for the reader's convenience.

Assumption 2.4.

- (i) The operator $A : D(A) \subset H \rightarrow H$ is self-adjoint, densely defined in H and the infinitesimal generator of an analytic semigroup $S = (S(t), t \geq 0) \subset \mathcal{L}(H)$. Moreover, $(-A) : D(A) \rightarrow H$ is boundedly invertible, i.e. $0 \in \rho(A)$, where $\rho(A)$ is the resolvent set of A .
- (ii) $X_0 \in L^8(\Omega; H)$ is a \mathcal{F}_0 -measurable random variable.
- (iii) The mappings $F : H \rightarrow H$ and $G : H \rightarrow \mathcal{L}_{HS}(\mathcal{U}, H)$ are twice Fréchet differentiable on H with bounded derivatives, i.e. there is a $C > 0$ such that for all $v \in H$ there holds

$$\begin{aligned} \|F'(v)\|_{\mathcal{L}(H)} + \|F''(v)\|_{\mathcal{L}(H \times H, H)} &\leq C \\ \|G'(v)\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))} + \|G''(v)\|_{\mathcal{L}(H \times H, \mathcal{L}_{HS}(\mathcal{U}, H))} &\leq C. \end{aligned}$$

- (iv) There are constants $C > 0$ and $\alpha \geq 1$ such that for all $v \in \dot{H}^\alpha$ there hold the linear growth bounds

$$\begin{aligned} \|F(v)\|_{\dot{H}^\alpha} + \|G(v)\|_{\mathcal{L}_{HS}(\mathcal{U}, \dot{H}^\alpha)} &\leq C(1 + \|v\|_{\dot{H}^\alpha}), \\ \|G'(v)\|_{\mathcal{L}(\dot{H}^\alpha, \mathcal{L}_{HS}(\mathcal{U}, \dot{H}^\alpha))} &\leq C. \end{aligned}$$

Remark 2.5. We require $X_0 \in L^8(\Omega; H)$, rather than $X_0 \in L^2(\Omega; H)$, in Item (ii) for some technical steps in the proofs (cf. Lemma C.2 in the Appendix), as we apply Hölder's inequality to obtain suitable mean-square error bounds.

Mild solutions to SPDEs are characterized by path-wise identities that hold almost surely as follows:

Definition 2.6. [30, Chapter 9] Let $\mathcal{X}_\mathbb{T}$ be as in (1). A process $X \in \mathcal{X}_\mathbb{T}$ is called a *mild solution* to Eq. (4) if for all $t \in \mathbb{T}$ there holds \mathbb{P} -a.s.

$$X(t) = S(t)X_0 + \int_0^t S(t-s)F(X(s))ds + \int_0^t S(t-s)G(X(s))dW(s). \quad (5)$$

In Eq. (5), $S : \mathbb{T} \rightarrow \mathcal{L}(H)$ is the semigroup generated by $(-A)$, thus $S(t) = e^{-tA}$ and Eq. (5) may be interpreted as a *variation-of-constants* formula. Well-posedness of (4) in the mild sense, and regularity of solutions has been investigated under suitable assumptions on F, G and X_0 , see e.g. [30, Theorems 9.15 and 9.29] or [25, Chapters 2.4-2.6]. We condense the main results in the following statement.

Theorem 2.7. *Under Assumption 2.4, there exists a unique mild solution $X \in \mathcal{X}_{\mathbb{T}}$ to (4), such that for all $p \in (0, 8]$ and $\kappa \in [0, \alpha]$ it holds that*

$$\sup_{t \in \mathbb{T}} \mathbb{E}(\|X(t)\|_{\dot{H}^\alpha}^p) < \infty \quad \text{and} \quad \sup_{t, s \in \mathbb{T}} \frac{\mathbb{E}(\|X(t) - X(s)\|_{\dot{H}^\kappa}^p)^{1/p}}{|t - s|^{\min(1/2, (\alpha - \kappa)/2)}} < \infty.$$

3 Pathwise Approximations

3.1 Spatial Discretization

To derive a spatial approximation based define $V := \dot{H}^1 = D((-A)^{1/2})$ and consider the bilinear form

$$B : V \times V \rightarrow \mathbb{R}, \quad (v, w) \mapsto ((-A)^{1/2}v, (-A)^{1/2}w). \quad (6)$$

In Example 2.3, where A is the Laplacian with zero Dirichlet boundary conditions on a convex domain $\mathcal{D} \subset \mathbb{R}^d$, we have $V = H_0^1(\mathcal{D})$, and $B(v, w) = (\nabla v, \nabla w)_H$.

We replace V by a finite dimensional subspace V_N with $N := \dim(V_N) \in \mathbb{N}$. This encompasses several spatial approximations, for instance spectral Galerkin methods, where N is number of terms in expansion, and finite element methods, where the mesh refinement parameter $h > 0$ is related to N via $N = \mathcal{O}(h^{-d})$. We introduce the *discrete* operator $A_N : V_N \rightarrow V_N$ by

$$(-A_N v_N, w_N) = B(v_N, w_N), \quad v_N, w_N \in V_N. \quad (7)$$

Then, $(-A_N)$ generates an analytic semigroup $(S_N(t), t \geq 0)$ of linear operators $S_N(t) : V_N \rightarrow V_N$ via $S_N(t) := \exp(-tA_N)$. Let $P_N : H \rightarrow V_N$ be the H -orthogonal projection onto V_N . The semi-discrete (mild) problem is then to find $X_N : \Omega \times \mathbb{T} \rightarrow V_N$ such that for all $t \in \mathbb{T}$ there holds \mathbb{P} -a.s.

$$X_N(t) = S_N(t)P_N X_0 + \int_0^t S_N(t-s)P_N F(X_N(s))ds + \int_0^t S_N(t-s)P_N G(X_N(s))dW(s). \quad (8)$$

3.2 Noise Approximation

Recall the Karhunen-Loève expansion of W from Equation (3), where the scalar products $(W(\cdot), e_k)_H$ are real-valued, independent and scaled Brownian motions with variance $\eta_k \geq 0$ (the k -th eigenvalue of Q). In general, infinitely many of the eigenvalues η_k are strictly greater than zero, hence we truncate the series in Eq. (3) after $K \in \mathbb{N}$ terms to obtain

$$W_K(t) := \sum_{k=1}^K (W(t), e_k)_U e_k, \quad t \in \mathbb{T}.$$

It can be shown, see for example [7], that W_K converges to W in mean-square uniformly on \mathbb{T} with truncation error given by

$$\mathbb{E}(\|W_K(t) - W(t)\|_U^2) = t \sum_{k>K} \eta_k, \quad t \in \mathbb{T}.$$

Combining the semi-discrete mild formulation from (8) with the noise truncation then yields the problem to find $X_{N,K} : \Omega \times \mathbb{T} \rightarrow V_N$ such that for all $t \in \mathbb{T}$ there holds \mathbb{P} -a.s.

$$X_{N,K}(t) = S_N(t)P_N X_0 + \int_0^t S_N(t-s)P_N F(X_{N,K}(s))ds + \int_0^t S_N(t-s)P_N G(X_{N,K}(s))dW_K(s). \quad (9)$$

3.3 Time Stepping

The temporal discretization is based on rational approximations of S_N . Recall that $(-A_N) : V_N \rightarrow V_N$ is a linear, positive definite, self-adjoint operator and that $N = \dim(V_N) \in \mathbb{N}$. There exists an H -orthonormal eigenbasis $(\tilde{f}_1, \dots, \tilde{f}_N) \subseteq V_N$ of eigenfunctions of $(-A_N)$, with corresponding non-decreasing eigenvalues $(\tilde{\lambda}_1, \dots, \tilde{\lambda}_N)$ such that $\tilde{\lambda}_1 > 0$. We denote the spectrum of $(-A_N)$ by $\sigma(-A_N)$ and consider a rational function r defined on $\sigma(-A_N)$.

Now fix $M \in \mathbb{N}$ and let $0 = t_0 < t_1 < \dots < t_M = T$ be an equidistant grid of $[0, T]$ with time step $\Delta t := T/M$. Further, let $r(\Delta t A_N)$ be a rational approximation of $S_N(\Delta t) = \exp(-\Delta t A_N)$, given by

$$r(\Delta t A_N)v = \sum_{n=1}^N r(\Delta t \tilde{\lambda}_n)(v, \tilde{f}_n)_H \tilde{f}_n, \quad v \in H. \quad (10)$$

The drift part in (9) is then approximated in each time step by the forward difference

$$\int_{t_m}^{t_{m+1}} S_N(t_{m+1} - s) P_N F(X_{N,K}(s)) ds \approx r(\Delta t A_N) P_N F(X_{N,K}(t_m)) \Delta t.$$

To introduce the approximation of the stochastic integral, recall that $G' : H \rightarrow \mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))$ denotes the Fréchet derivative of G . For any $k \in \mathbb{N}$ such that $\eta_k > 0$, we define $w_k := \eta_k^{-1/2}(W, e_k)_U$, hence $(w_k, k \in \mathbb{N})$ is the sequence of independent Brownian motions in the Karhunen-Loève expansion of W . Further, for $m = 0, \dots, M-1$ and any stochastic process $\mathfrak{W} : \Omega \times \mathbb{T} \rightarrow \mathcal{H}$ with $\mathcal{H} \in \{\mathbb{R}, U, \mathcal{L}_1(U)\}$, we denote by $\Delta_m \mathfrak{W} := \mathfrak{W}(t_{m+1}) - \mathfrak{W}(t_m)$ the increment with timestep $[t_m, t_{m+1}]$ (we will use in particular $\mathfrak{W} \in \{W, W_K, w_k\}$). We employ a truncated Milstein scheme to approximate the stochastic integral in (9) by a first order Taylor expansion of G via

$$\begin{aligned} & \int_{t_m}^{t_{m+1}} S_N(t_{m+1} - s) P_N G(X_{N,K}(s)) dW_K(s) \\ & \approx \int_{t_m}^{t_{m+1}} S_N(t_{m+1} - s) P_N G(X_{N,K}(t_m)) dW_K(s) \\ & \quad + \int_{t_m}^{t_{m+1}} S_N(t_{m+1} - s) P_N \left[G'(X_{N,K}(t_m)) \left(\int_{t_m}^s S_N(s-r) P_N G(X_{N,K}(r)) dW_K(r) \right) \right] dW_K(s) \\ & \approx r(\Delta t A_N) P_N G(X_{N,K}(t_m)) \Delta_m W_K \\ & \quad + r(\Delta t A_N) P_N \int_{t_m}^{t_{m+1}} G'(X_{N,K}(t_m)) \left(P_N G(X_{N,K}(t_m)) \int_{t_m}^s dW_K(r) \right) dW_K(s) \\ & \approx r(\Delta t A_N) P_N G(X_{N,K}(t_m)) \Delta_m W_K \\ & \quad + \frac{r(\Delta t A_N) P_N}{2} \sum_{k,l=1}^K G'(X_{N,K}(t_m)) (P_N G(X_{N,K}(t_m)) e_l) e_k (\sqrt{\eta_k \eta_l} \Delta_m w_k \Delta_m w_l - \delta_{k,l} \eta_k \Delta t), \end{aligned}$$

where $\delta_{k,l}$ is the Kronecker delta. This approximation corresponds to the truncated Milstein scheme in [17] for finite-dimensional SDEs. Now define for any $s \in [t_m, T]$ the $\mathcal{L}_1(U)$ -valued process

$$\mathcal{W}_{m,K}(s) := (W_K(s) - W_K(t_m)) \otimes (W_K(s) - W_K(t_m)) - (s - t_m) \sum_{k=1}^K \eta_k e_k \otimes e_k, \quad (11)$$

and note that $\mathcal{W}_{m,K}$ is a continuous, square-integrable, $\mathcal{L}_1(U)$ -valued martingale on $[t_m, T]$. Further, let $Q \otimes Q \in \mathcal{L}(\mathcal{L}_1(U))$ be given by $Q \otimes Q(\phi \otimes \varphi) = Q\phi \otimes Q\varphi$ for all $\phi \otimes \varphi \in \mathcal{L}_1^+(U)$. As $W_K(s) - W_K(t_m)$ is Gaussian, there is a $C > 0$ such that for all $s, t \in [t_m, T]$ with $t \geq s$ there holds

$$\langle \langle \mathcal{W}_{m,K}, \mathcal{W}_{m,K} \rangle \rangle_t - \langle \langle \mathcal{W}_{m,K}, \mathcal{W}_{m,K} \rangle \rangle_s \leq C(t-s)^2 Q \otimes Q. \quad (12)$$

We use the operator-valued processes $\mathcal{W}_{m,K}$ to write the truncated correction term in a compact form.

Proposition 3.1. *Let Assumption 2.4 hold and let $\mathcal{W}_{m,K}$ be defined as in (11) for $m = 0, \dots, M-1$ and $M \in \mathbb{N}$, and let further $\Delta_m w_k := (W(t_{m+1}) - W(t_m), e_k)_U$ for $k \in \mathbb{N}$. There exists a mapping $\mathcal{G} : H \mapsto \mathcal{L}_{HS}(\mathcal{L}_{HS}(U), H)$, such that for any $X \in H$ and $M \in \mathbb{N}$ there holds*

$$\int_{t_m}^{t_{m+1}} \mathcal{G}(X) d\mathcal{W}_{m,K}(s) = \frac{1}{2} \sum_{k,l=1}^K G'(X) (P_N G(X) e_l) e_k (\sqrt{\eta_k \eta_l} \Delta_m w_k \Delta_m w_l - \delta_{k,l} \eta_k \Delta t). \quad (13)$$

Moreover, \mathcal{G} is Fréchet differentiable on H and satisfies the linear growth bounds

$$\|\mathcal{G}(X)\|_{\mathcal{L}_{HS}(\mathcal{L}_{HS}(U), H)} + \|\mathcal{G}'(X)\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{L}_{HS}(U), H))} \leq C(1 + \|X\|_H), \quad X \in H, \quad (14)$$

and

$$\|\mathcal{G}(X)\|_{\mathcal{L}_{HS}(\mathcal{L}_{HS}(U), \dot{H}^\alpha)} \leq C(1 + \|X\|_{\dot{H}^\alpha}), \quad X \in \dot{H}^\alpha. \quad (15)$$

Proof. See Appendix B. □

Based on Proposition 3.1, we obtain the fully discrete problem as to find $Y_0^{N,K}, Y_1^{N,K}, \dots, Y_M^{N,K} : \Omega \rightarrow V_N$ such that $Y_0^{N,K} = P_N X_0$ and for all $m = 0, \dots, M-1$ there holds

$$Y_{m+1}^{N,K} = r(\Delta t A_N) P_N (Y_m^{N,K} + F(Y_m^{N,K}) \Delta t + G(Y_m^{N,K}) \Delta_m \mathcal{W}_K + \mathcal{G}(Y_m^{N,K}) \Delta_m \mathcal{W}_{m,K}), \quad (16)$$

where we have used (13) to define the last term in (16) as

$$\mathcal{G}(Y_m^{N,K}) \Delta_m \mathcal{W}_{m,K} := \int_{t_m}^{t_{m+1}} \mathcal{G}(Y_m^{N,K}) d\mathcal{W}_{m,K}(s), \quad m = 0, \dots, M-1.$$

The first three terms on the right hand side of (16) correspond to an Euler approximation of X , the fourth term is the truncated Milstein correction. We emphasize that *the scheme in (16) does not require the simulation of any iterated stochastic integrals*, and is therefore straightforward to implement. We formulate the following assumption on strong and weak convergence of the fully discrete scheme.

Assumption 3.2.

- (i) The rational approximation r of S_N is of order $q \in \mathbb{N}$ and stable. That is, $r(z) = e^{-z} + \mathcal{O}(z^{q+1})$ as $z \rightarrow 0$, $|r(z)| < 1$ for $z > 0$ and $\lim_{z \rightarrow \infty} r(z) = 0$.
- (ii) Subspace approximation property: Fix $\alpha > 0$ and let $(V_N, N \in \mathbb{N})$ be a sequence of subspaces $V_N \subset V$ such that $\dim(V_N) = N$. There are constants $C, \tilde{\alpha} > 0$, depending on α and d , such that for any $N \in \mathbb{N}$ and any $v \in \dot{H}^\alpha$ there holds

$$\|v - P_N v\|_H \leq C N^{-\tilde{\alpha}} \|v\|_{\dot{H}^\alpha}, \quad \text{and} \quad \left\| A_N^{\min(\alpha, 2)/2} P_N v \right\|_H \leq C \|v\|_{\dot{H}^{\min(\alpha, 2)}}.$$

- (iii) Strong convergence: There are constants $C, \tilde{\alpha}, \beta > 0$ such that for $p \in (0, 8]$ and all discretization parameters $M, N, K \in \mathbb{N}$ there holds the strong error estimate

$$\max_{m=0, \dots, M} \|X(m\Delta t) - Y_m^{N,K}\|_{L^p(\Omega; H)} \leq C \left(M^{-1/2} + N^{-\tilde{\alpha}} + K^{-\beta} \right).$$

Remark 3.3. Item (iii) essentially states that the truncated Milstein term does neither increase, nor spoil the strong rates of convergence, as compared to the standard Euler scheme. We justify this in Theorem 3.5 below. However, the Milstein correction accelerates variance decay for an antithetic coupling in the MLMC estimator. If the eigenvalues of Q exhibit the decay $\eta_j \leq C j^{-\beta_0}$ for some $\beta_0 > 1$, we may choose $\beta = 1/2(\beta_0 - 1 - \delta) > 0$ for an arbitrary small $\delta \in (0, \beta_0 - 1)$ in Item (iii).

Example 3.4.

1. Assume the setting in Example 2.3, i.e., we consider the stochastic heat equation on a convex domain \mathcal{D} . Let V_N be a space of linear finite elements with respect to a regular triangulation of \mathcal{D} with mesh width $h = \mathcal{O}(N^{-d})$ for $N \in \mathbb{N}$. Assumption 3.2 then holds with $\tilde{\alpha} = \min(\alpha, 2)/d$, where α is the spatial Sobolev regularity of X as in Assumption 2.4, see e.g. [25, Chapters 3 and 5].
2. For the Dirichlet-Laplacian in Example 2.3, we have by Weyl's law that $\lambda_n = \mathcal{O}(n^{2/d})$ as $n \rightarrow \infty$. For a spectral Galerkin approach with $V_N = \text{span}\{f_1, \dots, f_N\}$, Assumption 3.2 thus holds with $\tilde{\alpha} = \alpha/d$ and we obtain the stronger relation

$$\left\| A_N^{\alpha/2} P_N v \right\|_H \leq \|v\|_{\dot{H}^\alpha},$$

for $\alpha \geq 2$. However, this will not affect efficiency of our antithetic scheme in Section 4, therefore we formulated Item (ii) in a unified way to encompass spectral Galerkin and finite element approaches.

We conclude this section by recording an error estimate on the truncated Milstein approximation.

Theorem 3.5. *Let Assumptions 2.4 and 3.2 (i) hold, and denote by $Y^{N,K} : \{0, \dots, M\} \times \Omega \rightarrow H$ the truncated Milstein approximation in (16).*

1. For any $p \in [2, 8]$ there is a constant $C > 0$, independent of M, N and K , such that

$$\max_{m=0, \dots, M} \mathbb{E} \left(\|Y_m^{N,K}\|_H^p \right) \leq C(1 + \mathbb{E}(\|X_0\|_H^p)) < \infty. \quad (17)$$

2. Let the standard Euler discretization of (2.4) be given by $\hat{Y}_0 = P_N X_0$ and

$$\hat{Y}_{m+1}^{N,K} = r(\Delta t A_N) \hat{Y}_m^{N,K} + r(\Delta t A_N) P_N F(\hat{Y}_m^{N,K}) \Delta t + r(\Delta t A_N) P_N G(\hat{Y}_m^{N,K}) \Delta_m W_K,$$

for $m = 0, \dots, M-1$. Then, for any $p \in (0, 4]$ there exists a constant $C > 0$, independent of M, N and K , such that

$$\max_{m=0, \dots, M} \mathbb{E} \left(\|Y_m^{N,K} - \hat{Y}_m^{N,K}\|_H^p \right)^{1/p} \leq CM^{-1/2}.$$

Proof. See Appendix B. □

Remark 3.6. The second part of Theorem 3.5 justifies Assumption 3.2 (iii), since the truncated Milstein correction does not affect the temporal convergence of order $\mathcal{O}(M^{-1/2})$ for the Euler scheme, neither does it introduce an additional bias with respect to N or K .

4 Antithetic Multilevel Monte Carlo-Milstein Scheme

To construct an antithetic estimator, we consider coupled "coarse" and "fine" approximations of X given by the truncated Milstein scheme in (16), with refinement parameters M, N and K adjusted accordingly. First, let $Y^c := Y^{N,K}$ be the coarse step discretization with a fixed time step $\Delta t = T/M$ and fixed $N, K \in \mathbb{N}$ as in (16). That is, Y_m^c is given by $Y_0^c := P_N Y_0$ and

$$Y_{m+1}^c = r(\Delta t A_N) P_N (Y_m^c + F(Y_m^c) \Delta t + G(Y_m^c) \Delta_m W_K + \mathcal{G}(Y_m^c) \Delta_m \mathcal{W}_{m,K}), \quad m = 0, \dots, M-1. \quad (18)$$

For the corresponding fine approximation, let $\mathfrak{d}t := \Delta t/2$ and set a cutoff index $K_f \in \mathbb{N}$ such that $K_f \geq K$ for the noise approximation. We define the fine increments

$$\begin{aligned} \mathfrak{d}_m W_{K_f} &:= W_{K_f}(t_m + \mathfrak{d}t) - W_{K_f}(t_m), & \mathfrak{d}_{m+1/2} W_{K_f} &:= W_{K_f}(t_{m+1}) - W_{K_f}(t_m + \mathfrak{d}t) \\ \mathfrak{d}_m \mathcal{W}_{m,K_f} &:= \mathcal{W}_{m,K_f}(t_m + \mathfrak{d}t), & \mathfrak{d}_{m+1/2} \mathcal{W}_{m,K_f} &:= \mathfrak{d}_{m+1/2} W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f} - \mathfrak{d}t \sum_{k=1}^{K_f} \eta_k e_k \otimes e_k. \end{aligned}$$

Note that $\Delta_m \mathcal{W}_{m,K_f} \neq \mathfrak{d}_{m+1/2} \mathcal{W}_{m,K_f} + \mathfrak{d}_m \mathcal{W}_{m,K_f}$, but there holds

$$\Delta_m \mathcal{W}_{m,K_f} = \mathfrak{d}_{m+1/2} \mathcal{W}_{m,K_f} + \mathfrak{d}_m \mathcal{W}_{m,K_f} + \mathfrak{d}_{m+1/2} W_{K_f} \otimes \mathfrak{d}_m W_{K_f} + \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f}. \quad (19)$$

We further consider a finite dimensional subspace $V_{N_f} \subset V$ with $\dim(V_{N_f}) = N_f$ such that $N_f \geq N = \dim(V_N)$. The corresponding discrete operator is denoted by $A_{N_f} : V_{N_f} \rightarrow V_{N_f}$, its associated semigroup by S_{N_f} , and $P_{N_f} : H \rightarrow V_{N_f}$ is the H -orthogonal projection onto V_{N_f} . Finally, we set a cutoff index $K_f \in \mathbb{N}$ such that $K_f \geq K$ for the noise approximation on the fine scale.

The *fine step discretization* $Y^f : \Omega \times \{0, 1/2, 1, \dots, M-1/2, M\} \rightarrow V_{N_f}$ is then given by $Y_0^f := P_{N_f} X_0$,

$$Y_{m+1/2}^f = r(\mathfrak{d}t A_{N_f}) P_{N_f} \left(Y_m^f + F(Y_m^f) \mathfrak{d}t + G(Y_m^f) \mathfrak{d}_m W_{K_f} + \mathcal{G}(Y_m^f) \mathfrak{d}_m \mathcal{W}_{m,K_f} \right), \quad (20)$$

and

$$Y_{m+1}^f = r(\mathfrak{d}t A_{N_f}) P_{N_f} \left(Y_{m+1/2}^f + F(Y_{m+1/2}^f) \mathfrak{d}t + G(Y_{m+1/2}^f) \mathfrak{d}_{m+1/2} W_{K_f} + \mathcal{G}(Y_{m+1/2}^f) \mathfrak{d}_{m+1/2} \mathcal{W}_{m,K_f} \right). \quad (21)$$

The *antithetic counterpart* $Y^a : \Omega \times \{0, 1/2, 1, \dots, M-1/2, M\} \rightarrow V_{N_f}$ to Y^f is defined via $Y_0^a := P_{N_f} X_0$,

$$Y_{m+1/2}^a = r(\mathfrak{d}t A_{N_f}) P_{N_f} \left(Y_m^a + F(Y_m^a) \mathfrak{d}t + G(Y_m^a) \mathfrak{d}_{m+1/2} W_{K_f} + \mathcal{G}(Y_m^a) \mathfrak{d}_{m+1/2} \mathcal{W}_{m,K_f} \right), \quad (22)$$

and

$$Y_{m+1}^a = r(\mathfrak{d}t A_{N_f}) P_{N_f} \left(Y_{m+1/2}^a + F(Y_{m+1/2}^a) \mathfrak{d}t + G(Y_{m+1/2}^a) \mathfrak{d}_m W_{K_f} + \mathcal{G}(Y_{m+1/2}^a) \mathfrak{d}_m \mathcal{W}_{m,K_f} \right). \quad (23)$$

The foundation of our MLMC Milstein approach is to show that the difference of the *antithetic average*

$$\bar{Y}_m := \frac{Y_m^f + Y_m^a}{2}, \quad m = 0, \dots, M, \quad (24)$$

to the coarse scale Y_m^c approximations exhibits a rapid decay in mean-square. This property is established in our main result:

Theorem 4.1. *Let Assumptions 2.4 and 3.2 hold for some $\alpha \geq 1$, and let $M, N_f, N, K_f, K \in \mathbb{N}$ be such that $N_f \geq N$ and $K_f \geq K$. Further, let Y^c be as in (18) and let \bar{Y}^c be the antithetic average of the fine approximations as in (24). Then, there is a constant $C > 0$, independent of M, N , and K such that*

$$\max_{m=0, \dots, M} \mathbb{E} \left(\|\bar{Y}_m - Y_m^c\|_H^2 \right) \leq C \left(M^{-\min(\alpha, 2)} + N^{-2\bar{\alpha}} + K^{-2\beta} \right). \quad (25)$$

Proof. See Appendix C. □

Remark 4.2. For the truncated Milstein scheme without antithetic correction, we have

$$\max_{m=0, \dots, M} \mathbb{E} \left(\|Y_m^f - Y_m^c\|_H^2 \right) \leq C \left(M^{-1} + N^{-2\bar{\alpha}} + K^{-2\beta} \right), \quad (26)$$

and therefore a slower variance decay with respect to the time step $\Delta t = T/M$.

5 Multilevel Monte Carlo Approximation

Let $Z : \Omega \rightarrow \mathbb{R}$ be a real-valued, integrable random variable, and let $(Z^{(i)}, i \in \mathbb{N})$ be a sequence of independent copies of Z . For any finite number of samples $\mathfrak{N} \in \mathbb{N}$ we define the singlelevel Monte Carlo estimator of $\mathbb{E}(Z)$ by

$$E_{\mathfrak{N}}(Z) := \frac{1}{\mathfrak{N}} \sum_{i=1}^{\mathfrak{N}} Z^{(i)}. \quad (27)$$

We aim to estimate $\mathbb{E}(\Psi(X_T))$ for a given functional $\Psi : H \rightarrow \mathbb{R}$ by multilevel Monte Carlo (MLMC) methods. To this end, let $M_0, L \in \mathbb{N}$ and let $M_\ell = M_0 2^\ell$ for $\ell = 1, \dots, L$. Based on Theorem 4.1, we balance the error contributions in (25) on all levels by setting the remaining approximation parameters as

$$N_\ell := \lceil M_\ell^{\min(\alpha, 2)/2\tilde{\alpha}} \rceil \quad \text{and} \quad K_\ell := \lceil M_\ell^{\min(\alpha, 2)/2\beta} \rceil, \quad \ell = 1, \dots, L. \quad (28)$$

We denote for $\ell = 2, \dots, L$ by $Y^{c, \ell-1}$ the coarse step approximation in (18) with discretization parameters given by $M_{\ell-1}, N_{\ell-1}, K_{\ell-1}$. For $\ell = 1, \dots, L$, we let denote by $Y^{f, \ell}, Y^{a, \ell}$ the fine step discretization and its antithetic counterpart, respectively, both with discretization parameters $M_f = M_\ell = 2M_{\ell-1}$ and $N_f = N_\ell, K_f = K_\ell$. Furthermore, we define $\bar{Y}^\ell = 1/2(Y^{f, \ell} + Y^{a, \ell})$,

$$\Psi_0^c := 0, \quad \Psi_\ell^c := \Psi(Y_{M_\ell}^{c, \ell}), \quad \text{and} \quad \bar{\Psi}_\ell := \frac{\Psi(Y_{M_\ell}^{f, \ell}) + \Psi(Y_{M_\ell}^{a, \ell})}{2}, \quad \text{for } \ell = 1, \dots, L. \quad (29)$$

We introduce the *antithetic multilevel Monte Carlo estimator* as

$$E_L^{\text{ML}}(\Psi) := \sum_{\ell=1}^L E_{\mathfrak{N}_\ell}(\bar{\Psi}_\ell - \Psi_\ell^c), \quad (30)$$

where $\mathfrak{N}_1, \dots, \mathfrak{N}_L \in \mathbb{N}$ are level-dependent numbers of samples. Since $Y_M^{f, \ell} \stackrel{d}{=} Y_M^{a, \ell}$, it holds that

$$\mathbb{E}(E_L^{\text{ML}}(\Psi)) = \mathbb{E}\left(\Psi(Y_{M_L}^{f, L})\right).$$

To analyze the mean-squared error (MSE) and computational complexity of the estimator in (30), we formulate the following assumptions on the sample complexity and the weak error.

Assumption 5.1. For fixed $M_0 \in \mathbb{N}$ and any $\ell \in \mathbb{N}$, let $M_\ell = M_0 2^\ell$ and $N_\ell, K_\ell \in \mathbb{N}$ be as in (28).

- (i) **Sample complexity:** Denote by \mathcal{C}_ℓ the cost of generating one sample of $\bar{\Psi}^\ell$ on any a given refinement level $\ell \in \mathbb{N}$. There are constants $C > 0$ and $\gamma > 0$ such that for any $\ell \in \mathbb{N}$ there holds

$$\mathcal{C}_\ell \leq CM_\ell^{1+\gamma}.$$

- (ii) **Weak convergence:** Let $\tilde{\alpha}$ and β be as in Assumption 3.2 (iii), let $\Psi : H \rightarrow \mathbb{R}$ be Fréchet differentiable with bounded derivative, and let $\delta \in (0, 1)$ be arbitrary small. There is a constant $C = C(\Psi, \delta) > 0$ such that for $\ell \in \mathbb{N}$ there holds

$$\left| \mathbb{E}(\Psi(X(T))) - \mathbb{E}(\Psi(Y_{M_\ell}^{f, \ell})) \right| \leq CM_\ell^{-1+\delta}.$$

Remark 5.2. The parameter γ in Item (i) essentially depends on the cost of evaluating G . In case there is some sparsity to exploit in G , the cost of one evaluation may be as low as $\mathcal{O}(\max(N_\ell, K_\ell))$, in which case (28) yields that $\gamma = 1/2 \cdot \min(\alpha, 2)/\min(\tilde{\alpha}, \beta)$, see for instance the numerical example in Section 6. On the other hand, the cost of one evaluation may be as large as $\mathcal{O}(N_\ell^2 K_\ell^2)$, if evaluating G entails full matrices and nested summations in the discretization scheme, which makes each sample significantly more expensive.

Assumption 5.1 (ii) on the weak approximation error is natural, one often recovers (almost) twice the strong rates for semi-linear, parabolic SPDEs, see e.g. [25, Theorem 5.12] or [24, Theorem 4.5]. In other words, the weak error with respect to N_ℓ and K_ℓ is of order $\mathcal{O}(N_\ell^{-2\tilde{\alpha}+\delta} + K_\ell^{-2\beta+\delta})$ for any arbitrary small $\delta > 0$, and the balancing in (28) yields with $\alpha \geq 1$ that

$$\left| \mathbb{E}(\Psi(X(T))) - \mathbb{E}(\Psi(Y_{M_\ell}^{f, \ell})) \right| = \mathcal{O}(M_\ell^{-1+\delta} + N_\ell^{-2\tilde{\alpha}+\delta} + K_\ell^{-2\beta+\delta}) = \mathcal{O}(M_\ell^{-1+\delta}).$$

Theorem 5.3. *Let Assumptions 2.4, 3.2 and 5.1 hold. For any $\varepsilon \in (0, e^{-1})$, there exists an antithetic multilevel Monte Carlo-Milstein estimator $E_L(\Psi(Y_M))$ of $\mathbb{E}(\Psi(X(T)))$ such that there holds*

$$\mathbb{E} \left(|E_L^{ML}(\Psi) - \mathbb{E}(\Psi(X(T)))|^2 \right) < \varepsilon^2,$$

and the computational complexity \mathcal{C}_{ML} of $E_L(\Psi(Y_M))$ is bounded by

$$\mathcal{C}_{ML} \leq \begin{cases} C\varepsilon^{-2}, & \min(\alpha, 2) > 1 + \gamma, \\ C\varepsilon^{-2} |\log(\varepsilon)|^2, & \min(\alpha, 2) = 1 + \gamma, \\ C\varepsilon^{-2 - \frac{1+\gamma - \min(\alpha, 2)}{1-\delta}}, & \min(\alpha, 2) < 1 + \gamma. \end{cases} \quad (31)$$

Proof. See Appendix D. □

Example 5.4. To show that all three cases in (31) are conceivable, recall Example 3.4, where $\tilde{\alpha} = \min(\alpha, 2)/d$ (FEM) or $\tilde{\alpha} = \alpha/d$ (spectral Galerkin method). Assuming for simplicity that $\alpha \in [1, 2]$, $\tilde{\alpha} = \alpha/d$, and that G may be evaluated with complexity $\mathcal{O}(\max(N_\ell, K_\ell))$. The error balancing (28) then yields

$$\max(N_\ell, K_\ell) = M_\ell^{\alpha \max(d/2\alpha, 1/2\beta)},$$

and thus Item (i) holds with

$$\gamma = \alpha \max(d/2\alpha, 1/2\beta) < \alpha - 1 \iff \max(d/2, \alpha/2\beta) < \alpha - 1.$$

6 Numerics

Let $\mathcal{D} = [0, 1]^d$ for $d \in \{1, 2, 3\}$, let $H := L^2(\mathcal{D})$ and denote by $A := \Delta$ the Laplace-operator with homogeneous Dirichlet boundary conditions. We further assume that $U = H$ and denote by $(e_k, k \in \mathbb{N})$ and $(\lambda_k, k \in \mathbb{N})$ the eigenfunctions and eigenvalues of $-A$, respectively. By Weyl's law $\lambda_k = \mathcal{O}(k^{2/d})$ for $k \in \mathbb{N}$ and for rectangular and circular domains the precise eigenfunctions and eigenvalues of Δ are given in closed form, see e.g. [18, Section 3]. We consider the stochastic heat equation given by

$$dX(t) = \Delta X(t)dt + G(X(t))dW(t), \quad t \in [0, 1], \quad X(0) = X_0, \quad (32)$$

for $X_0 \in L^8(\Omega; \dot{H}^2)$. The driving noise is modeled by a Q -Wiener process $W : \Omega \times [0, 1] \rightarrow H$ with covariance operator $Q := ((-\Delta)^{-s})$ for a smoothness parameter $s > 0$. Since $\lambda_k = \mathcal{O}(k^{2/d})$ for $k \in \mathbb{N}$, Q is trace-class for $s > d/2$, in which case W admits the Karhunen-Loève expansion

$$W(t) := \sum_{k \in \mathbb{N}} \eta_k^{1/2} w_k(t) e_k, \quad (33)$$

where $\eta_k = \lambda_k^{-s}$ and w_1, w_2, \dots are independent one-dimensional Brownian motions. Hence, truncating the expansion (33) after K terms yields an error of order $\mathcal{O}(K^{1-2s/d})$ with respect to $\|\cdot\|_H^2$, uniform in $[0, T]$, and implies that Assumption 3.2 holds with $\beta = (2s/d - 1)/2$. Alternatively, we could define the diffusion part of Equation (32) as $\widehat{G}(X_t)d\widehat{W}_t$, where \widehat{W} is Gaussian white noise (i.e. has covariance operator $\widehat{Q} = \mathcal{I}$) and $\widehat{G}(v)e_k := \eta_k^{1/2} G(v)e_k$ for all $v \in H$ and $k \in \mathbb{N}$. By [25, Theorem 2.27] it then follows that Equation (32) admits a unique mild solution X such that $X(t) \in L^8(\Omega; \dot{H}^\alpha)$ for $\alpha \in [1, \min(1 + s, 2)]$ and all $t \in [0, 1]$. We fix the diffusion coefficient G to be linear and given by

$$G(v)u := \sum_{k=1}^{\infty} (v, e_k)_H e_{k+1}(u, \sqrt{\eta_{k+1}} e_{k+1})u + (v, e_k)_H e_k(u, \sqrt{\eta_k} e_k)u$$

for all $v \in H, u \in \mathcal{U} = Q^{1/2}H$ and some fixed $g \in H$. We use a spectral Galerkin approach and expand Y^c, Y^f and Y^a in the same basis, for example

$$Y_m^c = \sum_{k=1}^N y_{m,k}^c e_k$$

for $\{y_{m,k}^c\}_{k=1}^N \in \mathbb{R}^N$ and $m = 1, \dots, M$. Recall from Example 3.4 that Assumption 3.2 then holds with $\tilde{\alpha} = \alpha/d$. The scheme in (18) simplifies to

$$\begin{aligned} y_{m+1,1}^c &= r(\Delta t \lambda_1) (y_{m,1} + (g, e_1)_H \Delta_m w_1) \\ y_{m+1,2}^c &= r(\Delta t \lambda_2) \left(y_{m,2} + (y_{m,1} + (g, e_2)_H) \Delta_m w_2 + \frac{1}{2} (g, e_1)_H \Delta_m w_2 \Delta_m w_1 \right). \\ y_{m+1,k}^c &= r(\Delta t \lambda_k) (y_{m,k} + \chi_{k \leq K} (y_{m,k-1} + (g, e_k)_H) \Delta_m w_k) \\ &\quad + \chi_{k \leq K} \frac{r(\Delta t \lambda_k)}{2} (y_{m,k-2} + (g, e_{k-1})_H) \Delta_m w_k \Delta_m w_{k-1}. \end{aligned}$$

for $K \geq 2, k = 3, \dots, N$, and for χ being the characteristic function. The schemes in (20) and (21) for $\{y_{m,k}^f\}_{k=1}^{N_f}$ and (22) and (23) for $\{y_{m,k}^a\}_{k=1}^{N_f}$ simplify similarly. The cost of evaluating the previous scheme is $\mathcal{O}(N)$ for every time step since only $\mathcal{O}(\min(N, K))$ independent Brownian increments $\Delta_m w_1, \dots, \Delta_m w_N$ are needed. We set $(g, e_k)_H = k^{-1/2-\varepsilon}$ and $(X_0, e_k)_H = k^{-1/2-2/d-\varepsilon}$ for $k \in \mathbb{N}$ and some $\varepsilon > 0$. Note that with this choice $G(v) \in \mathcal{L}_{HS}(\mathcal{U}, H)$ for all $v \in H$ and $X_0 \in L^8(\Omega; \dot{H}^2)$. In Figure 1, we fix M and K to some sufficiently large values and plot estimates of the difference $\max_m \|Y_m^{N,K} - Y_m^{\lceil \sqrt{2N} \rceil, K}\|_{L^2(\Omega; H)}$ for several values of N . The plot verifies the convergence order with respect to N in Assumption 3.2 as $\mathcal{O}(N^{-\min(s+1, 2)/d})$. Next, we choose N and K in terms of M as in (28). In this case the cost per sample is $\mathcal{O}(M^{1+\gamma})$ where

$$\gamma = \max(d/2, \alpha/2\beta) = d \max(1/2, \min(1+s, 2)/(2s-d)).$$

We plot in Figure 2 estimates of the left-hand sides of (25) and (26) which verifies the claim of Theorem 4.1 and the improved convergence order of the variance for the antithetic estimator. In particular, the variance convergence order is $\mathcal{O}(M^{-\min(\alpha, 2)}) = \mathcal{O}(M^{-\min(s+1, 2)})$ for the antithetic estimator and $\mathcal{O}(M^{-1})$ for the truncated Milstein scheme without antithetic correction. Both figures also clearly showcase the reduced convergence rates in terms of N and M when $d = 1$ and as the smoothness parameter s decreases.

7 Conclusions

We have developed an antithetic MLMC-Milstein scheme for parabolic SPDEs, which offers a significant improvement in computational efficiency for estimating quantities of interest in SPDE models. This scheme circumvents the need to simulate intractable Lévy area terms, making it particularly advantageous for SPDEs with multiplicative noise and non-commutative diffusion terms. In our study, we have derived precise variance decay bounds for a fully discrete scheme that incorporates antithetic time stepping, spatial approximations, and noise approximations. Furthermore, we have bounded the computational effort by considering the cost associated with evaluating the noise term. These results provide valuable insights into the efficiency and accuracy of our proposed scheme.

There are several possible extensions to the current work that could be explored. A further step to enhance efficiency would be to develop a higher-order noise approximation that achieves a better rate than $\mathcal{O}(K^{-\gamma})$ in relation to the truncation index K (cf. Theorem 4.1). Additionally, the methodology employed in this study could be extended to incorporate discontinuous Lévy driving noise, provided that it possesses a sufficient number of moments. While this extension may initially seem straightforward, it is important to emphasize that our results heavily rely on the continuous version of the Burkholder-Davis-Gundy inequality (Eq. 36), while only a weaker version (Eq. 35) is available for discontinuous martingales.

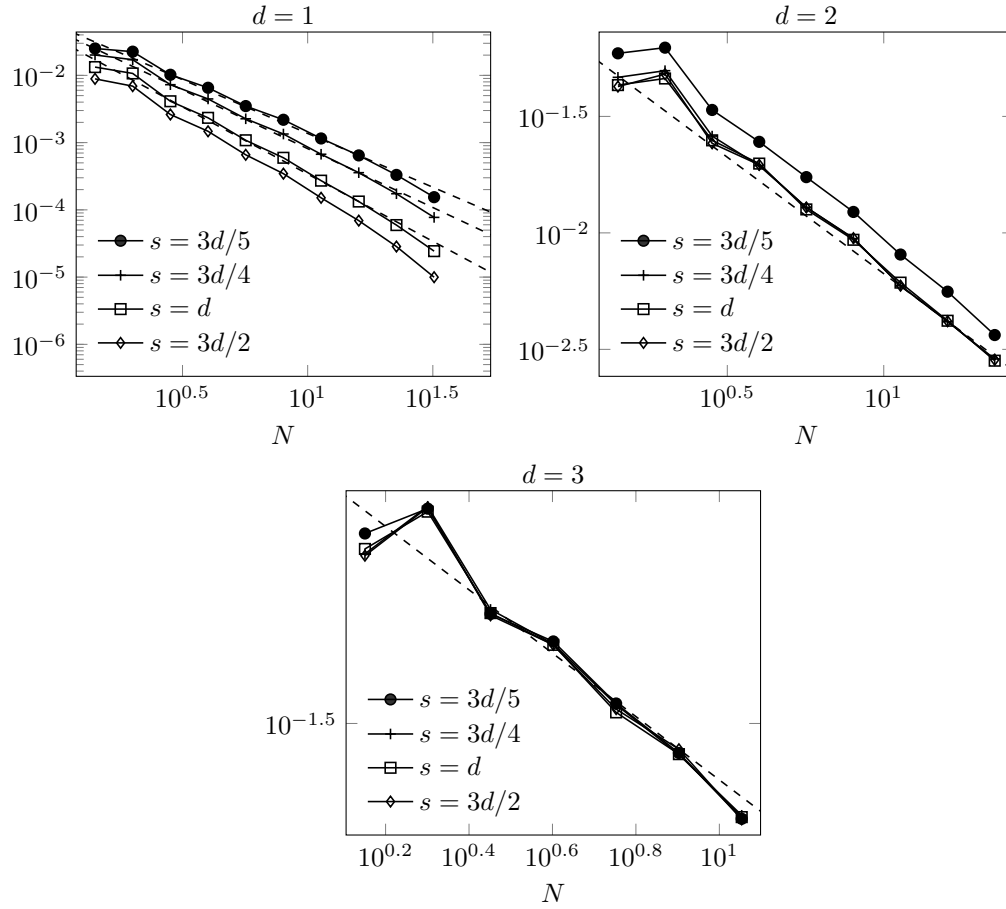


Figure 1: Estimates of the $L^2(\Omega; H)$ difference $\max_m \|Y_m^{N,K} - Y_m^{\lceil \sqrt{2}N \rceil, K}\|_{L^2(\Omega; H)}$ for the numerical example in Section 6 and when using the Galerkin method for different number of terms, N , in the spatial approximation. The estimates were obtained using Monte Carlo sampling with at least 4000 samples. The dashed reference lines are $\propto N^{-\min(1+s, 2)/d}$ and verify the assumed convergence rates in Assumption 3.2.

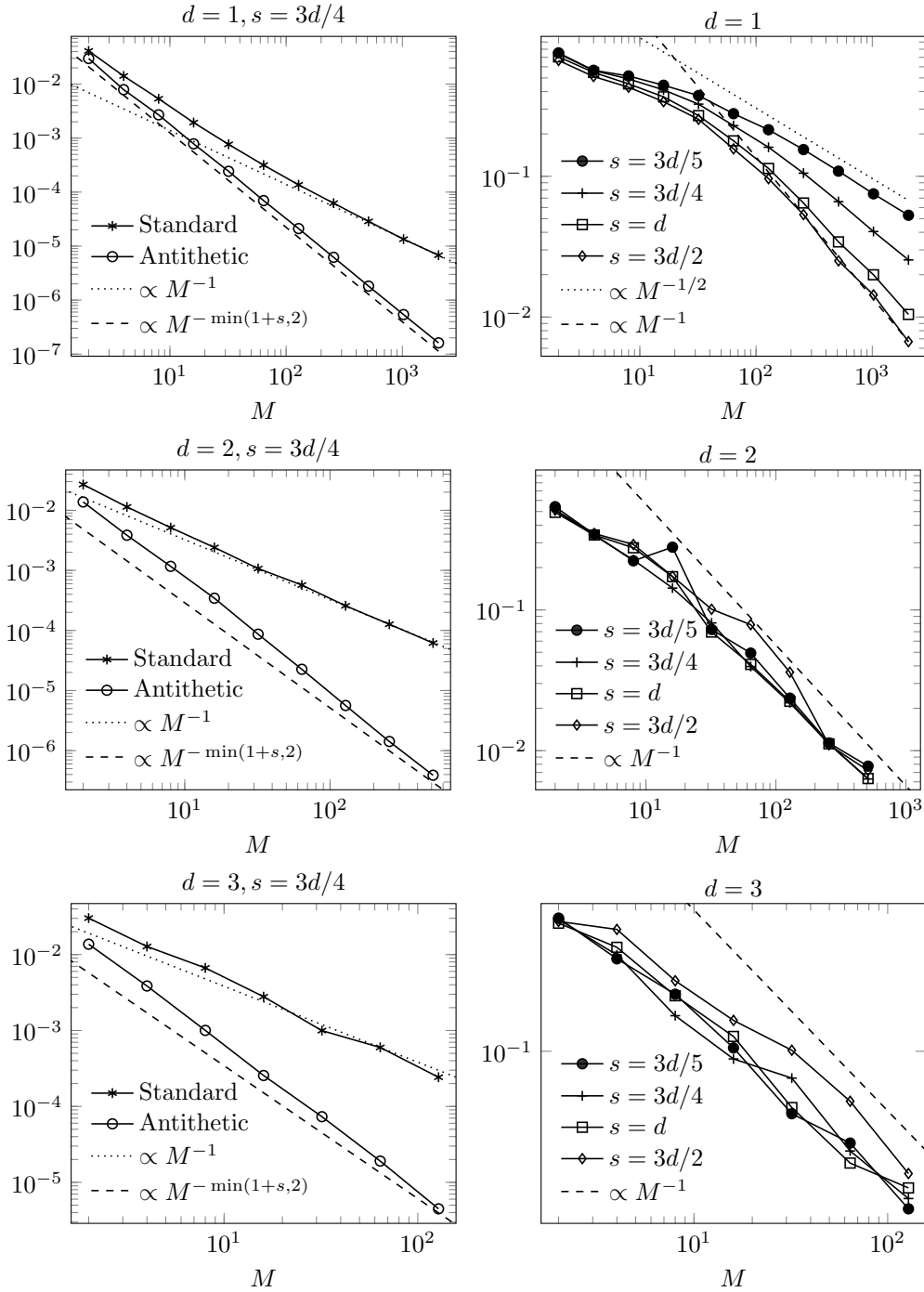


Figure 2: Results for numerical example in Section 6 and M, N and K as in (28). (left) Shows the left-hand sides of (25), the variance for the antithetic estimator, and (26), the variance for the “Standard” truncated Milstein estimator without the antithetic correction, for the smoothness parameter $s = 3d/4$. (right) Shows the relative variance decay between the two estimators, i.e., $\max_m \mathbb{E} \left(\|\bar{Y}_m - Y_m^c\|_H^2 \right) / \max_m \mathbb{E} \left(\|Y_m^f - Y_m^c\|_H^2 \right) = \mathcal{O}(M^{-\min(s, 1)})$, for different smoothness parameters s . The variance estimates were obtained using Monte Carlo sampling with at least 4000 samples.

Consequently, a completely different proof technique would be required, even for the relatively simple case of Poisson driving noise.

Another intriguing avenue for exploration would be the consideration of first-order hyperbolic SPDEs, which commonly arise in the modeling of energy forward contracts [10, 3]. In such cases, the weak formulation of SPDEs becomes essential for pathwise discretizations, see [8], as the associated semigroups lack the smoothing properties observed in the parabolic case. Furthermore, recent developments have seen the application of a modified version of the antithetic Milstein scheme to finite-dimensional stochastic differential equations with non-Lipschitz drift [28]. Extending this result to SPDEs in infinite dimensions would be both intriguing and worthwhile.

Acknowledgements

Work on this manuscript was initiated during the authors' visit to the workshop "Computational Uncertainty Quantification: Mathematical Foundations, Methodology & Data" at the Erwin-Schrödinger Institute (ESI) in Vienna. We would like to thank the organizers of this workshop and the ESI for their invitation and their hospitality during this time. AS is partly funded by ETH Foundations of Data Science (ETH-FDS), and it is gratefully acknowledged.

A Itô Isometry and Burkholder-Davis-Gundy Inequalities

We record the following Itô isometry and Burkholder-Davis-Gundy (BDG)-type inequalities for Hilbert space-valued stochastic integrals in the setting of Section 2.

Lemma A.1. *Let $Y \in \mathcal{M}^2(U)$ and let $G : \Omega \times \mathbb{T} \rightarrow \mathcal{L}(U, H)$ be a $\mathcal{P}_{\mathbb{T}}/\mathcal{B}(\mathcal{L}_{HS}(U, H))$ -measurable process such that*

$$\mathbb{E} \left(\int_0^T \left\| G(s) \mathcal{Q}_Y^{1/2}(s) \right\|_{\mathcal{L}_{HS}(U, H)}^2 d\langle Y, Y \rangle_s \right) < \infty.$$

1. *There holds the isometric formula*

$$\mathbb{E} \left(\left\| \int_0^t G(s) dY(s) \right\|_H^2 \right) = \mathbb{E} \left(\int_0^t \left\| G(s) \mathcal{Q}_Y^{1/2}(s) \right\|_{\mathcal{L}_{HS}(U, H)}^2 d\langle Y, Y \rangle_s \right), \quad t \in \mathbb{T}. \quad (34)$$

2. *If for some $p > 2$ there holds*

$$\mathbb{E} \left(\int_0^T \left\| G(s) \mathcal{Q}_Y^{1/2}(s) \right\|_{\mathcal{L}_{HS}(U, H)}^p d\langle Y, Y \rangle_s \right) < \infty,$$

then there is a $C = C(T, p) > 0$ such that

$$\mathbb{E} \left(\left\| \int_0^t G(s) dY(s) \right\|_H^p \right) \leq C \mathbb{E} \left(\int_0^t \left\| G(s) \mathcal{Q}_Y^{1/2}(s) \right\|_{\mathcal{L}_{HS}(U, H)}^p d\langle Y, Y \rangle_s \right), \quad t \in \mathbb{T}. \quad (35)$$

Moreover, if Y has continuous trajectories, then

$$\mathbb{E} \left(\left\| \int_0^t G(s) dY(s) \right\|_H^p \right) \leq C \mathbb{E} \left(\left(\int_0^t \left\| G(s) \mathcal{Q}_Y^{1/2}(s) \right\|_{\mathcal{L}_{HS}(U, H)}^2 d\langle Y, Y \rangle_s \right)^{p/2} \right), \quad t \in \mathbb{T}. \quad (36)$$

For a proof of (34) see e.g. [30, Theorem 8.7], the BDG inequality (36) for continuous martingales may for instance be found [20, Eq. (1.5)] and the references therein. The previous result simplifies for Wiener processes with constant martingale covariance $\mathcal{Q}_Y = Q \operatorname{Tr}(Q)^{-1}$ as in Section 2.4. In this case, the Itô isometry and BDG inequality from Lemma A.1 admit the following form.

Corollary A.2. [30, Corollary 8.17 and Lemma 8.27] Let W be a Q -Wiener process and let $G : \Omega \times \mathbb{T} \rightarrow \mathcal{L}_{HS}(\mathcal{U}, H)$ be a $\mathcal{P}_{\mathbb{T}}/\mathcal{B}(\mathcal{L}_{HS}(\mathcal{U}, H))$ -measurable and square-integrable mapping.

1. There holds the isometric formula

$$\mathbb{E} \left(\left\| \int_0^t G(s) dW(s) \right\|_H^2 \right) = \int_0^t \mathbb{E} \left(\|G(s)\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 \right) ds = \int_0^t \sum_{k \in \mathbb{N}} \eta_k \mathbb{E} \left(\|G(s)e_k\|_H^2 \right) ds. \quad (37)$$

2. If, in addition, for some $p > 2$ there holds

$$\mathbb{E} \left(\int_0^t \|G(s)\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^p ds \right) < \infty,$$

then there is a $C = C(p) > 0$ such that for $t \in \mathbb{T}$ there holds

$$\mathbb{E} \left(\left\| \int_0^t G(s) dW(s) \right\|_H^p \right) \leq C \mathbb{E} \left(\left(\int_0^t \|G(s)\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 ds \right)^{p/2} \right). \quad (38)$$

B Proofs of Section 3

Proof of Proposition 3.1. For any fixed $X \in H$, define

$$\tilde{G}_X : U \times U \mapsto H, \quad (\phi, \varphi) \mapsto \frac{1}{2} G'(X) (P_N G(X) \phi) \varphi.$$

As $G'(X) \in \mathcal{L}(H, \mathcal{L}(\mathcal{U}, H))$ and $G(X) \in \mathcal{L}(\mathcal{U}, H)$, it readily follows that \tilde{G}_X is bilinear. Thus, there exists a unique $\mathcal{G}_X \in \mathcal{L}(U \otimes U, H) \simeq \mathcal{L}(\mathcal{L}_{HS}(\mathcal{U}), H)$ such that

$$\tilde{G}_X(\phi, \varphi) = \mathcal{G}_X(\phi \otimes \varphi), \quad (\phi, \varphi) \in U \times U. \quad (39)$$

We thus define $\mathcal{G} : H \rightarrow \mathcal{L}(\mathcal{L}_{HS}(\mathcal{U}), H)$, $X \mapsto \mathcal{G}_X$, and (13) follows by the linearity of \mathcal{G}_X together with

$$\begin{aligned} \Delta_m \mathcal{W}_{m,K} &= (W_K(t_{m+1}) - W_K(t_m)) \otimes (W_K(t_{m+1}) - W_K(t_m)) - \Delta t \sum_{k=1}^K \eta_k e_k \otimes e_k \\ &= \sum_{k,l=1}^K (\Delta_m w_k \Delta_m w_l - \delta_{k,l} \eta_k \Delta t) e_k \otimes e_l. \end{aligned}$$

To show that $\mathcal{G}(X) \in \mathcal{L}_{HS}(\mathcal{L}_{HS}(\mathcal{U}), H)$ for all $X \in H$, we use (39) and that $(\sqrt{\eta_k} e_k \otimes \sqrt{\eta_l} e_l, (k, l) \in \mathbb{N}^2)$ is an orthonormal basis of $\mathcal{L}_{HS}(\mathcal{U})$ to obtain

$$\begin{aligned} \|\mathcal{G}(X)\|_{\mathcal{L}_{HS}(\mathcal{L}_{HS}(\mathcal{U}), H)}^2 &= \sum_{k,l \in \mathbb{N}} \|\mathcal{G}(X)(\sqrt{\eta_k} e_k \otimes \sqrt{\eta_l} e_l)\|_H^2 \\ &= \frac{1}{4} \sum_{k,l \in \mathbb{N}} \|G'(X)(P_N G(X) \sqrt{\eta_k} e_k) \sqrt{\eta_l} e_l\|_H^2 \\ &= \frac{1}{4} \sum_{k \in \mathbb{N}} \|G'(X)(P_N G(X) \sqrt{\eta_k} e_k)\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 \\ &\leq \frac{1}{4} \|G'(X)\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))}^2 \sum_{k \in \mathbb{N}} \|G(X) \sqrt{\eta_k} e_k\|_H^2 \\ &= \frac{1}{4} \|G'(X)\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))}^2 \|G(X)\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2. \end{aligned}$$

Using that G is twice differentiable with bounded derivatives from Assumption 2.4 (iii), we obtain

$$\|\mathcal{G}(X)\|_{\mathcal{L}_{HS}(\mathcal{L}_{HS}(\mathcal{U}), H)}^2 \leq C(1 + \|X\|_H)^2.$$

The bound in (15) follows analogously, by using Assumption 2.4 (iv) instead.

The Fréchet derivative $\mathcal{G}' \in \mathcal{L}(H \times H, \mathcal{L}_{HS}(\mathcal{L}_{HS}(\mathcal{U}), H))$ of \mathcal{G} is for all $(X, Y) \in H \times H$ given by

$$\mathcal{G}'(X)(Y)(\phi \otimes \varphi) = \frac{1}{2} (G''(X)[P_N G(X)\phi, Y]\varphi + G'(X)(P_N G'(X)(Y)\phi)\varphi), \quad \phi, \varphi \in \mathcal{U}.$$

The last estimate then follows since G' and G'' are globally bounded by Assumption 2.4 ((iii)), since

$$\begin{aligned} & \|\mathcal{G}'(X)(Y)(\phi \otimes \varphi)\|_{\mathcal{L}_{HS}(\mathcal{L}_{HS}(\mathcal{U}), H)} \\ & \leq \frac{1}{2} \left(\|G''(X)[P_N G(X)\phi, Y]\|_{\mathcal{L}_{HS}(\mathcal{U}, H)} + \|G'(X)(P_N G'(X)(Y)\phi)\|_{\mathcal{L}_{HS}(\mathcal{U}, H)} \right) \|\varphi\|_{\mathcal{U}} \\ & \leq \frac{1}{2} \left(\|G''(X)\|_{\mathcal{L}(H \times H, \mathcal{L}_{HS}(\mathcal{U}, H))} \|G(X)\phi\|_H \|Y\|_H + \|G'(X)\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))} \|G'(X)(Y)\phi\|_H \right) \|\varphi\|_{\mathcal{U}} \\ & \leq \frac{1}{2} \left(\|G''(X)\|_{\mathcal{L}(H \times H, \mathcal{L}_{HS}(\mathcal{U}, H))} \|G(X)\|_{\mathcal{L}_{HS}(\mathcal{U}, H)} \|Y\|_H \right. \\ & \quad \left. + \|G'(X)\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))} \|G'(X)\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))} \|Y\|_H \right) \|\phi\|_{\mathcal{U}} \|\varphi\|_{\mathcal{U}} \\ & \leq C(1 + \|X\|_H) \|Y\|_H \|\phi \otimes \varphi\|_{\mathcal{L}_1(\mathcal{U})}. \end{aligned}$$

□

We next record some stability and error estimates on the rational approximation $r(\Delta t A_N) \approx S_N(\Delta t)$ to prove Theorem 3.5.

Lemma B.1. *Let Assumption 3.2 (i) and (ii) hold.*

1. *For any $N \in \mathbb{N}$, $\Delta t > 0$ and $\alpha \geq 0$ there holds*

$$\|r(\Delta t A_N) P_N\|_{\mathcal{L}(H)} \leq 1 \quad \text{and} \quad \|r(\Delta t A)\|_{\mathcal{L}(\dot{H}^\alpha)} \leq 1.$$

2. *For any $\alpha \in [0, 2(q+1)]$ there exists $C > 0$ such that for any $\Delta t > 0$ there holds*

$$\|r(\Delta t A) - I\|_{\mathcal{L}(\dot{H}^\alpha, H)} \leq C \Delta t^{\alpha/2}.$$

3. *For any $\alpha \in [0, 2(q+1)]$ there exists C such that for any $N, \Delta t > 0$, $j \in \mathbb{N}$ and $v \in \dot{H}^\alpha$ it holds*

$$\|(r(\Delta t A_N)^j - S_N(j\Delta t)) P_N v\|_H \leq C \Delta t^{\min(\alpha, 2)/2} \|v\|_{\dot{H}^{\min(\alpha, 2)}}.$$

For $\mathfrak{d}t := \frac{\Delta t}{2}$ and $j \in \mathbb{N}$ there holds

$$\|(r(\mathfrak{d}t A_N)^{2j} - r(\Delta t A_N))^j P_N v\|_H \leq C \Delta t^{\min(\alpha, 2)/2} \|v\|_{\dot{H}^{\min(\alpha, 2)}}.$$

Furthermore, there exists $\tilde{\alpha} > 0$ such that

$$\|(r(\Delta t A_N) P_N - I)v\|_H \leq C(\Delta t^{\min(\alpha, 2)/2} + N^{-\tilde{\alpha}}) \|v\|_{\dot{H}^\alpha}.$$

Proof.

1. These stability estimates are well-known and may be found for instance in the proof of [32, Theorem 7.1]. We give a short proof here for the reader's convenience.

Let $(\tilde{f}_1, \dots, \tilde{f}_N)$ denote the eigenbasis of $(-A_N)$, and recall that the corresponding eigenvalues satisfy $\tilde{\lambda}_n > 0$ for all $n = 1, \dots, N$. Since $r(z) < 1$ for all $z \geq 0$ by the first part of Assumption 3.2 (i), we have for all $v \in H$ that

$$\|r(\Delta t A_N) P_N v\|_H^2 = \left\| \sum_{n=1}^N r(\Delta t \tilde{\lambda}_n) (P_N v, \tilde{f}_n)_H \tilde{f}_n \right\|_H^2 \leq \|P_N v\|_H^2 \leq \|v\|_H^2.$$

For the second part, let $\lambda_n > 0$ and $f_n \in H$ denote for $n \in \mathbb{N}$ the eigenvalues and eigenfunctions of $(-A)$. Similar as for the first part, we have

$$\|r(\Delta t A) v\|_{\dot{H}^\alpha}^2 = \sum_{n \in \mathbb{N}} \lambda_n^\alpha |r(\Delta t \lambda_n)|^2 |(v, f_n)_H|^2 \leq \sum_{n \in \mathbb{N}} \lambda_n^\alpha |(v, f_n)_H|^2 \leq \|v\|_{\dot{H}^\alpha}^2.$$

2. The triangle inequality yields for any $v \in \dot{H}^\alpha$ that

$$\|r(\Delta t A) - I\|_{\mathcal{L}(\dot{H}^\alpha, H)} \leq \|r(\Delta t A) - S(\Delta t)\|_{\mathcal{L}(\dot{H}^\alpha, H)} + \|S(\Delta t) - I\|_{\mathcal{L}(\dot{H}^\alpha, H)} \leq C \Delta t^{\alpha/2}.$$

The first term on the right hand side is bounded by [32, Theorem 7.1], the second term by [29, Theorem 6.13, part d)].

3. The first part is again given in [32, Theorem 7.1] together with Assumption 3.2 (ii) that yields

$$\left\| A_N^{\min(\alpha, 2)/2} P_N v \right\|_H \leq C \|v\|_{\dot{H}^{\min(\alpha, 2)}}.$$

The second part then follows immediately by the triangle inequality, since $S_N(2j\delta t) = S_N(j\Delta t)$.

The final estimate follows by

$$\begin{aligned} \|(r(\Delta t A_N) P_N - I)v\|_H &\leq \|(r(\Delta t A_N) - I)P_N v\|_H + \|(P_N - I)v\|_H \\ &\leq C \Delta t^{\min(\alpha, 2)/2} \left\| A_N^{\min(\alpha, 2)/2} P_N v \right\|_H + \|(P_N - I)v\|_H \\ &\leq C \left(\Delta t^{\min(\alpha, 2)/2} + N^{-\tilde{\alpha}} \right) \|v\|_{\dot{H}^\alpha}, \end{aligned}$$

where the second estimate is derived in the same fashion as part 2.), and we have used Assumption 3.2 (ii) in the last step. □

Proof of Theorem 3.5. 1. For $m = 0, \dots, M-1$, we re-iterate the representation in (16) to obtain

$$\begin{aligned} Y_{m+1}^{N,K} &= r(\Delta t A_N)^m P_N X_0 + \sum_{j=0}^m r(\Delta t A_{N_f})^j P_N F(Y_{m-j}^{N,K}) \Delta t + \sum_{j=0}^m r(\Delta t A_{N_f})^j P_N G(Y_{m-j}^{N,K}) \Delta_{m-j} W_K \\ &\quad + \sum_{j=0}^m r(\Delta t A_N) P_N \mathcal{G}(Y_{m-j}^{N,K}) \Delta_{m-j} W_{m-j,K} \\ &=: r(\Delta t A_N)^m P_N X_0 + \text{I} + \text{II} + \text{III}. \end{aligned} \tag{40}$$

The first term is bounded in $L^p(\Omega; H)$ by Jensen's inequality and the first part of Lemma B.1

$$\mathbb{E}(\|\text{I}\|_H^p) \leq \Delta t^p m^{p-1} \sum_{j=0}^m \mathbb{E} \left(\left\| r(\Delta t A_{N_f})^j P_N F(Y_{m-j}^{N,K}) \right\|_H^p \right) \leq C \Delta t \sum_{j=0}^m \mathbb{E} \left(1 + \left\| Y_{m-j}^{N,K} \right\|_H^p \right),$$

where the last bound holds since F is of linear growth.

For the second term, we use the BDG inequality in (38) together with Jensen's inequality and the linear growth of G to obtain

$$\begin{aligned}\mathbb{E}(\|\text{II}\|_H^p) &\leq C\mathbb{E}\left(\left(\sum_{j=0}^m \Delta t \left\|r(\Delta t A_{N_f})^j P_N G(Y_{m-j}^{N,K})\right\|_{\mathcal{L}_{HS}(\mathcal{U},H)}^2\right)^{p/2}\right) \\ &\leq C\Delta t^{p/2} m^{p/2-1} \sum_{j=0}^m \mathbb{E}\left(\left\|r(\Delta t A_{N_f})^j P_N G(Y_{m-j}^{N,K})\right\|_{\mathcal{L}_{HS}(\mathcal{U},H)}^p\right) \\ &\leq C\Delta t \sum_{j=0}^m \mathbb{E}\left(1 + \left\|Y_{m-j}^{N,K}\right\|_H^p\right).\end{aligned}$$

To bound the last term, recall the bound in (12), which shows with the BDG inequality from (36) that

$$\begin{aligned}\mathbb{E}(\|\text{III}\|_H^p) &\leq C\mathbb{E}\left(\left(\sum_{j=0}^m \int_{t_j}^{t_{j+1}} \left\|r(\Delta t A_{N_f})^j P_N \mathcal{G}(Y_{m-j}^{N,K}) Q_{\mathcal{W}_{m-j,K}}^{1/2}(s)\right\|_{\mathcal{L}_{HS}(U\otimes U,H)}^2 d\langle \mathcal{W}_{m-j,K}, \mathcal{W}_{m-j,K} \rangle_s\right)^{p/2}\right) \\ &\leq C\mathbb{E}\left(\left(\sum_{j=0}^m \left\|\mathcal{G}(Y_{m-j}^{N,K})\right\|_{\mathcal{L}_{HS}(U\otimes U,H)}^2 \Delta t^2\right)^{p/2}\right) \\ &\leq C\Delta t^p m^{p/2-1} \sum_{j=0}^m \mathbb{E}\left(\left\|\mathcal{G}(Y_{m-j}^{N,K})\right\|_{\mathcal{L}_{HS}(\mathcal{L}_1(U),H)}^p\right) \\ &\leq C\Delta t^{p/2+1} \sum_{j=0}^m \mathbb{E}\left(1 + \left\|Y_{m-j}^{N,K}\right\|_H^p\right),\end{aligned}$$

where the last bound holds due to the linear growth bound (14).

Substituting the estimate for I–III into (40) and taking expectations yields

$$\mathbb{E}\left(\left\|Y_{m+1}^{N,K}\right\|_H^p\right) \leq C(1 + \mathbb{E}(\|X_0\|_H^p)) + \Delta t \sum_{j=0}^m \mathbb{E}\left(\left\|Y_{m-j}^{N,K}\right\|_H^p\right),$$

and the first part of Theorem 3.5 follows by the discrete Grönwall inequality.

2. Let $e_m := Y_m^{N,K} - \widehat{Y}_m^{N,K}$ for $m = 0, \dots, M$ to obtain the error representation

$$\begin{aligned}\widehat{e}_{m+1} &= \sum_{j=0}^m r(\Delta t A_{N_f})^j P_N \left(F(Y_{m-j}^{N,K}) - F(\widehat{Y}_{m-j}^{N,K})\right) \Delta t \\ &\quad + \sum_{j=0}^m r(\Delta t A_{N_f})^j P_N \left(G(Y_{m-j}^{N,K}) - G(\widehat{Y}_{m-j}^{N,K})\right) \Delta_{m-j} W_M \\ &\quad + \sum_{j=0}^m r(\Delta t A_N) P_N \mathcal{G}(Y_{m-j}^{N,K}) \mathcal{W}_{m-j,K}\end{aligned}$$

Using that F and G are Lipschitz and repeating arguments from the first part of the proof yields

$$\mathbb{E}(\|e_{m+1}\|_H^p) \leq C \left(\Delta t \sum_{j=0}^m \mathbb{E}(\|e_j\|_H^p) + \Delta t^{p/2+1} \sum_{j=0}^m \mathbb{E}(\|Y_j\|_H^p) \right).$$

The claim then follows with Grönwall's inequality, since we have shown that $\mathbb{E}(\|Y_j\|_H^p) < \infty$ is uniformly bounded with respect to Δt in the first part. \square

C Proof of Theorem 4.1 – Antithetic Variance Decay

Our strategy to prove Theorem 4.1 closely follows the approach in [17]. We bound $\mathbb{E}(\|\bar{Y}_m - Y_m^c\|_H^2)$ by deriving appropriate difference equations of the antithetic average in (24) and by bounding higher-order remainder terms. We introduce the *semi-discrete temporal* fine discretizations $\tilde{Y}^f : \Omega \times \{0, 1/2, 1, \dots, M - 1/2, M\} \rightarrow H$ via $\tilde{Y}_0^f = X_0$,

$$\tilde{Y}_{m+1/2}^f = r(\mathfrak{d}tA) \left(\tilde{Y}_m^f + F(\tilde{Y}_m^f)\mathfrak{d}t + G(\tilde{Y}_m^f)\mathfrak{d}_m W_{K_f} + \mathcal{G}(\tilde{Y}_m^f)\mathfrak{d}_m \mathcal{W}_{m,K_f} \right), \quad (41)$$

and

$$\tilde{Y}_{m+1}^f = r(\mathfrak{d}tA) \left(\tilde{Y}_{m+1/2}^f + F(\tilde{Y}_{m+1/2}^f)\mathfrak{d}t + G(\tilde{Y}_{m+1/2}^f)\mathfrak{d}_{m+1/2} + \mathcal{G}(\tilde{Y}_{m+1/2}^f)\mathfrak{d}_{m+1/2} \mathcal{W}_{m,K_f} \right). \quad (42)$$

Note that we have used A instead of A_{N_f} as compared to Equations (41) and (42), hence \tilde{Y}^f involves temporal and noise discretization, but no spatial approximation. The corresponding antithetic fine semi-discretizations \tilde{Y}^a are defined analogously by replacing A_{N_f} by A and P_{N_f} by I in (22) and (23).

The next two auxiliary results establish a bound on $\tilde{Y}^f - Y^f$;

Lemma C.1. *Let Assumptions 2.4 and 3.2 (i) hold. For any $p \in (0, 8]$ there is a constant $C = C(p) > 0$ such that for all $M, N_f, K_f \in \mathbb{N}$ and $m = 0, 1/2, 1, \dots, M - 1/2$ there holds*

$$\mathbb{E} \left(\left\| Y_{m+1/2}^f - r(\mathfrak{d}tA_{N_f})Y_m^f \right\|_H^p \right) + \mathbb{E} \left(\left\| Y_{m+1/2}^a - r(\mathfrak{d}tA_{N_f})Y_m^a \right\|_H^p \right) \leq CM^{-p/2}. \quad (43)$$

Proof. We have by Equation (20), Lemma B.1 and Jensen's inequality that

$$\begin{aligned} & \mathbb{E} \left(\left\| Y_{m+1/2}^f - r(\mathfrak{d}tA_{N_f})Y_m^f \right\|_H^p \right) \\ & \leq C \left(\mathbb{E} \left(\left\| r(\mathfrak{d}tA_{N_f})P_{N_f}F(Y_m^f)\mathfrak{d}t \right\|_H^p \right) + \mathbb{E} \left(\left\| r(\mathfrak{d}tA_{N_f})P_{N_f}G(Y_m^f)\mathfrak{d}_m W_{K_f} \right\|_H^p \right) \right) \\ & \quad + C \mathbb{E} \left(\left\| \frac{r(\mathfrak{d}tA_{N_f})P_{N_f}}{2} \mathcal{G}(Y_m^f)\mathfrak{d}_m \mathcal{W}_{m,K_f} \right\|_H^p \right) \\ & \leq C \left(\mathbb{E} \left(\left\| F(Y_m^f) \right\|_H^p \right) \mathfrak{d}t^p + \mathbb{E} \left(\left\| G(Y_m^f)\mathfrak{d}_m W_{K_f} \right\|_H^p \right) \right) + C \mathbb{E} \left(\left\| \mathcal{G}(Y_m^f)\mathfrak{d}_m \mathcal{W}_{m,K_f} \right\|_H^p \right). \end{aligned}$$

The second part of Corollary A.2 shows together with (12), Assumption 2.4 (iii) and Proposition 3.1 that

$$\begin{aligned} \mathbb{E} \left(\left\| Y_{m+1/2}^f - r(\mathfrak{d}tA_{N_f})Y_m^f \right\|_H^p \right) & = C \left(\mathfrak{d}t^p \mathbb{E} \left(1 + \|Y_m^f\|_H^p \right) + \mathfrak{d}t^{p/2} \mathbb{E} \left(\|G(Y_m^f)\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^p \right) \right) \\ & \quad + C \mathfrak{d}t^p \mathbb{E} \left(\|\mathcal{G}(Y_m^f)\|_{\mathcal{L}_{HS}(\mathcal{L}_{HS}(\mathcal{U}, H))}^p \right) \\ & \leq C \mathfrak{d}t^{p/2} \mathbb{E} \left(1 + \|Y_m^f\|_H^p \right) \\ & \leq C \mathfrak{d}t^{p/2}. \end{aligned}$$

For the last step we have used that $\mathbb{E}(\|Y_m^f\|_H^p)$ is uniformly bounded by Theorem 3.5. The bound for $\mathbb{E} \left(\left\| Y_{m+1/2}^a - r(\mathfrak{d}tA_{N_f})Y_m^a \right\|_H^p \right)$ follows analogously. \square

Lemma C.2. *Let Assumptions 2.4 and 3.2 hold. Then, there is a constant $C = C(p) > 0$ such that for all $M, N_f, K_f \in \mathbb{N}$ and $m = 0, 1/2, 1, \dots, M - 1/2, M$ there holds*

$$\begin{aligned} \mathbb{E} \left(\left\| \tilde{Y}_m^f \right\|_{\dot{H}^\alpha}^p + \left\| \tilde{Y}_m^a \right\|_{\dot{H}^\alpha}^p \right) &\leq C, & \text{for } p \in (0, 8], \quad \text{and} \\ \mathbb{E} \left(\left\| Y_m^f - \tilde{Y}_m^f \right\|_H^p + \left\| Y_m^a - \tilde{Y}_m^a \right\|_H^p \right) &\leq C \left(M^{-p} + N_f^{-p\tilde{\alpha}} + K_f^{-p2\beta} \right) & \text{for } p \in (0, 4]. \end{aligned}$$

Proof. We may represent \tilde{Y}_m^f for any $m = 0, 1/2, 1, \dots, M - 1/2, M$ by the expansion

$$\tilde{Y}_m^f = r(\mathfrak{d}tA)^{2m} \tilde{Y}_0^f + \sum_{j=0}^{2m-1} r(\mathfrak{d}tA)^{2m-j} \left(F(\tilde{Y}_{j/2}^f) \mathfrak{d}t + G(\tilde{Y}_{j/2}^f) \mathfrak{d}_{j/2} W_{K_f} + \mathcal{G}(\tilde{Y}_{j/2}^f) \mathfrak{d}_{j/2} \mathcal{W}_{\lfloor j/2 \rfloor, K_f} \right). \quad (44)$$

This in turn shows with the first part of Lemma B.1, Jensen's inequality, and the second part of Corollary A.2 for $p \in [2, 8]$ that

$$\begin{aligned} \mathbb{E} \left(\left\| \tilde{Y}_m^f \right\|_{\dot{H}^\alpha}^p \right) &\leq C \left(\mathbb{E} \left(\left\| \tilde{Y}_0^f \right\|_{\dot{H}^\alpha}^p \right) + \sum_{j=0}^{2m-1} \mathbb{E} \left(\left\| F(\tilde{Y}_{j/2}^f) \right\|_{\dot{H}^\alpha}^p \right) (2m)^{p-1} \mathfrak{d}t^p \right) \\ &\quad + C \sum_{j=0}^{2m-1} \mathbb{E} \left(\left\| G(\tilde{Y}_{j/2}^f) \right\|_{\mathcal{L}_{HS}(\mathcal{U}, \dot{H}^\alpha)}^p \right) (2m)^{p/2-1} \mathfrak{d}t^{p/2} \\ &\quad + C \sum_{j=0}^{2m-1} \mathbb{E} \left(\left\| \mathcal{G}(\tilde{Y}_{j/2}^f) \right\|_{\mathcal{L}_{HS}(\mathcal{L}_1(\mathcal{U}), \dot{H}^\alpha)}^p \right) (2m)^{p/2-1} \mathfrak{d}t^p \\ &\leq C \left(\mathbb{E} \left(\left\| \tilde{Y}_0^f \right\|_{\dot{H}^\alpha}^p \right) + \mathfrak{d}t \sum_{j=0}^{2m-1} \mathbb{E} \left(1 + \left\| \tilde{Y}_{j/2}^f \right\|_{\dot{H}^\alpha}^p \right) \right) \\ &\leq C \left(\mathbb{E} \left(\left\| X_0 \right\|_{\dot{H}^\alpha}^p \right) + 1 + \mathfrak{d}t \sum_{j=0}^{2m-1} \mathbb{E} \left(\left\| \tilde{Y}_{j/2}^f \right\|_{\dot{H}^\alpha}^p \right) \right). \end{aligned}$$

We have used Assumption 2.4 (iv) and Proposition 3.1 to derive the second inequality. The discrete Grönwall inequality now shows that

$$\mathbb{E} \left(\left\| \tilde{Y}_m^f \right\|_{\dot{H}^\alpha}^p \right) \leq C \left(1 + \mathbb{E} \left(\left\| X_0 \right\|_{\dot{H}^\alpha}^p \right) \right) < \infty. \quad (45)$$

Thus, $\mathbb{E} \left(\left\| \tilde{Y}_m^f \right\|_{\dot{H}^\alpha}^p \right) < \infty$ for all $p \in (0, 8]$, and $\mathbb{E} \left(\left\| \tilde{Y}_m^a \right\|_{\dot{H}^\alpha}^p \right) < \infty$ follows analogously.

To show the second part, we use again (44) and repeat the above reasoning to find for $p \in [2, 4]$ that

$$\begin{aligned}
\mathbb{E} \left(\left\| Y_m^f - \tilde{Y}_m^f \right\|_H^p \right) &\leq C \mathbb{E} \left(\left\| Y_0^f - \tilde{Y}_0^f \right\|_H^p \right) + C \mathfrak{d} t \sum_{j=0}^{2m-1} \mathbb{E} \left(\left\| F(Y_{j/2}^f) - F(\tilde{Y}_{j/2}^f) \right\|_H^p \right) \\
&\quad + C \mathfrak{d} t \sum_{j=0}^{2m-1} \mathbb{E} \left(\left\| G(Y_{j/2}^f) - G(\tilde{Y}_{j/2}^f) \right\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^p \right) \\
&\quad + C \mathfrak{d} t^{1+p/2} \sum_{j=0}^{2m-1} \mathbb{E} \left(\left\| \mathcal{G}(Y_{j/2}^f) - \mathcal{G}(\tilde{Y}_{j/2}^f) \right\|_{\mathcal{L}_{HS}(\mathcal{L}_{HS}(\mathcal{U}, H))}^p \right) \\
&\leq C \left(\mathbb{E} \left(\left\| (P_{N_f} - I) X_0 \right\|_H^p \right) + \mathfrak{d} t \sum_{j=0}^{2m-1} \mathbb{E} \left(\left\| Y_{j/2}^f - \tilde{Y}_{j/2}^f \right\|_H^p \right) \right) \\
&\quad + C \mathfrak{d} t^{1+p/2} \sum_{j=0}^{2m-1} \mathbb{E} \left(\left\| \mathcal{G}(Y_{j/2}^f) - \mathcal{G}(\tilde{Y}_{j/2}^f) \right\|_{\mathcal{L}_{HS}(\mathcal{L}_{HS}(\mathcal{U}, H))}^p \right). \tag{46}
\end{aligned}$$

The second estimate follows since F and G are Fréchet differentiable with globally bounded derivatives. To bound the last term in (46), we recall that \mathcal{G}' satisfies the linear growth bound derived in Proposition 3.1 to obtain by Hölder's inequality and (45)

$$\begin{aligned}
\mathbb{E} \left(\left\| \mathcal{G}(Y_{j/2}^f) - \mathcal{G}(\tilde{Y}_{j/2}^f) \right\|_{\mathcal{L}_{HS}(\mathcal{L}_{HS}(\mathcal{U}, H))}^p \right) &\leq C \mathbb{E} \left(\max \left(1, \left\| Y_{j/2}^f \right\|_H, \left\| \tilde{Y}_{j/2}^f \right\|_H \right)^p \left\| Y_{j/2}^f - \tilde{Y}_{j/2}^f \right\|_H^p \right) \\
&\leq C \mathbb{E} \left(\left\| Y_{j/2}^f - \tilde{Y}_{j/2}^f \right\|_H^{2p} \right)^{1/2}.
\end{aligned}$$

Assumption 3.2 (iii) implies that the semi-discrete approximation satisfies for $p \in [2, 4]$ that

$$\mathbb{E} \left(\left\| Y \left(\frac{j\delta t}{2} \right) - \tilde{Y}_{j/2}^f \right\|_H^{2p} \right)^{1/2} \leq C \left(\mathfrak{d} t^{p/2} + K_f^{-p\beta} \right).$$

Together with the first part of Theorem 3.5 and the strong error from Assumption 3.2 (iii) this shows

$$\begin{aligned}
&\mathbb{E} \left(\left\| \mathcal{G}(Y_{j/2}^f) - \mathcal{G}(\tilde{Y}_{j/2}^f) \right\|_{\mathcal{L}_{HS}(\mathcal{L}_{HS}(\mathcal{U}, H))}^p \right) \\
&\leq C \left(\mathbb{E} \left(\left\| Y_{j/2}^f - Y \left(\frac{j\delta t}{2} \right) \right\|_H^{2p} \right)^{1/2} + \mathbb{E} \left(\left\| Y \left(\frac{j\delta t}{2} \right) - \tilde{Y}_{j/2}^f \right\|_H^{2p} \right)^{1/2} \right) \\
&\leq C \left(\mathfrak{d} t^{p/2} + N_f^{-p\tilde{\alpha}} + K_f^{-p\beta} \right). \tag{47}
\end{aligned}$$

Substituting (47) back to (46) now yields with Assumption 3.2 (ii)

$$\begin{aligned}
\mathbb{E} \left(\left\| Y_m^f - \tilde{Y}_m^f \right\|_H^p \right) &\leq C \left(N_f^{-p\tilde{\alpha}} + \mathfrak{d} t \sum_{j=0}^{2m-1} \mathbb{E} \left(\left\| Y_{j/2}^f - \tilde{Y}_{j/2}^f \right\|_H^p \right) + \mathfrak{d} t^{p/2} \left(\mathfrak{d} t^{p/2} + N_f^{-p\tilde{\alpha}} + K_f^{-p\beta} \right) \right) \\
&\leq C \left(\mathfrak{d} t^p + N_f^{-p\tilde{\alpha}} + K_f^{-p2\beta} + \mathfrak{d} t \sum_{j=0}^{2m-1} \mathbb{E} \left(\left\| Y_{j/2}^f - \tilde{Y}_{j/2}^f \right\|_H^p \right) \right).
\end{aligned}$$

The claim for $p \in [2, 4]$ follows by applying the discrete Grönwall inequality, and the estimate for $p \in (0, 2)$ is then obtained immediately by Hölder's inequality. \square

Lemma C.2 enables us to derive a difference equation on the fine approximation Y^f .

Proposition C.3. *Let Assumptions 2.4 and 3.2 hold. Then, for any $m = 0, \dots, M-1$ there holds that*

$$Y_{m+1}^f =: r(\partial_t A_{N_f})^2 P_{N_f} (Y_m^f + F(Y_m^f)\Delta t + G(Y_m^f)\Delta_m W_{K_f} + \mathcal{G}(Y_m^f)\Delta_m \mathcal{W}_{n,K_f}) - r(\partial_t A_{N_f})^2 P_{N_f} \mathcal{G}(Y_m^f) (\mathfrak{d}_{m+1/2} W_{K_f} \otimes \mathfrak{d}_m W_{K_f} - \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f}) + \Xi_m^f + O_m^f, \quad (48)$$

where $\Xi_m^f, O_m^f : \Omega \rightarrow H$ are random variables such that

$$\begin{aligned} \mathbb{E} \left(\|\Xi_m^f\|_H^2 \right) &\leq C\mathfrak{d}t^2 \left(M^{-\min(\alpha,2)} + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} \right), \\ \mathbb{E} \left(O_m^f | \mathcal{F}_{t_m} \right) &= 0 \quad \text{and} \quad \mathbb{E} \left(\|O_m^f\|_H^2 \right) \leq C\mathfrak{d}t \left(M^{-\min(\alpha,2)} + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} \right). \end{aligned}$$

The constant C is independent of M, N_f and K_f .

Proof. Assume for simplicity that Assumptions 2.4 and 3.2 hold with $\alpha \in [1, 2]$. Equations (21) and (20) show that

$$Y_{m+1}^f = r(\partial_t A_{N_f})^2 P_{N_f} (Y_m^f + F(Y_m^f)\Delta t + G(Y_m^f)\mathfrak{d}_m W_{K_f} + \mathcal{G}(Y_m^f)\mathfrak{d}_m \mathcal{W}_{m,K_f}) + r(\partial_t A_{N_f}) P_{N_f} \left(F(Y_{m+1/2}^f)\Delta t + G(Y_{m+1/2}^f)\mathfrak{d}_{m+1/2} W_{K_f} + \mathcal{G}(Y_{m+1/2}^f)\mathfrak{d}_{m+1/2} \mathcal{W}_{m,K_f} \right),$$

and by Equation (19) we have

$$\mathfrak{d}_m \mathcal{W}_{m,K_f} = \Delta_m \mathcal{W}_{m,K_f} - \mathfrak{d}_{m+1/2} \mathcal{W}_{m,K_f} - \mathfrak{d}_{m+1/2} W_{K_f} \otimes \mathfrak{d}_m W_{K_f} - \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f}.$$

Rearranging some terms yields

$$\begin{aligned} Y_{m+1}^f &= r(\partial_t A_{N_f})^2 P_{N_f} (Y_m^f + F(Y_m^f)\Delta t + G(Y_m^f)\Delta_m W_{K_f} + \mathcal{G}(Y_m^f)\Delta_m \mathcal{W}_{n,K_f}) \\ &\quad - r(\partial_t A_{N_f})^2 P_{N_f} \mathcal{G}(Y_m^f) (\mathfrak{d}_{m+1/2} W_{K_f} \otimes \mathfrak{d}_m W_{K_f} - \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f}) \\ &\quad + r(\partial_t A_{N_f}) P_{N_f} [F(Y_{m+1/2}^f) - r(\partial_t A_{N_f}) P_{N_f} F(Y_m^f)] \mathfrak{d}t \\ &\quad + r(\partial_t A_{N_f}) P_{N_f} \left[G(Y_{m+1/2}^f) - r(\partial_t A_{N_f}) P_{N_f} G(Y_m^f) \right] \mathfrak{d}_{m+1/2} W_{K_f} \\ &\quad - 2r(\partial_t A_{N_f}) P_{N_f} r(\partial_t A_{N_f}) P_{N_f} \mathcal{G}(Y_m^f) \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f} \\ &\quad + r(\partial_t A_{N_f}) P_{N_f} \left[\mathcal{G}(Y_{m+1/2}^f) - r(\partial_t A_{N_f}) P_{N_f} \mathcal{G}(Y_m^f) \right] \mathfrak{d}_{m+1/2} \mathcal{W}_{m,K_f}. \end{aligned}$$

The first two lines in the above equation correspond to the first two terms on the right hand side in (48), and we label the remaining terms via

$$\begin{aligned} \text{I}_m^f &:= [F(Y_{m+1/2}^f) - r(\partial_t A_{N_f}) P_{N_f} F(Y_m^f)] \mathfrak{d}t, \\ \text{II}_m^f &:= \left[G(Y_{m+1/2}^f) - r(\partial_t A_{N_f}) P_{N_f} G(Y_m^f) \right] \mathfrak{d}_{m+1/2} W_{K_f} - 2r(\partial_t A_{N_f}) P_{N_f} \mathcal{G}(Y_m^f) \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f} \\ \text{III}_m^f &:= \left[\mathcal{G}(Y_{m+1/2}^f) - r(\partial_t A_{N_f}) P_{N_f} \mathcal{G}(Y_m^f) \right] \mathfrak{d}_{m+1/2} \mathcal{W}_{m,K_f}, \end{aligned}$$

to obtain

$$Y_{m+1}^f = r(\partial_t A_{N_f})^2 P_{N_f} (Y_m^f + F(Y_m^f)\Delta t + G(Y_m^f)\Delta_m W_{K_f} + \mathcal{G}(Y_m^f)\Delta_m \mathcal{W}_{n,K_f}) - r(\partial_t A_{N_f})^2 P_{N_f} \mathcal{G}(Y_m^f) (\mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f} - \mathfrak{d}_{m+1/2} W_{K_f} \otimes \mathfrak{d}_m W_{K_f}) + r(\partial_t A_{N_f}) P_{N_f} \left[\text{I}_m^f + \text{II}_m^f + \text{III}_m^f \right]. \quad (49)$$

We split the first term \mathbf{I}_m^f further into

$$\mathbf{I}_m^f = \left[F(Y_{m+1/2}^f) - F(r(\partial t A_{N_f}) Y_m^f) + F(r(\partial t A_{N_f}) \tilde{Y}_m^f) - r(\partial t A_{N_f}) P_{N_f} F(Y_m^f) \right] \partial t =: \left(\mathbf{I}_m^{f,1} + \mathbf{I}_m^{f,2} \right) \partial t.$$

A first order Taylor expansion of $\mathbf{I}_m^{f,1}$ then yields for some $\xi_m^1 \in H$

$$\begin{aligned} \mathbf{I}_m^{f,1} &= F(Y_{m+1/2}^f) - F(r(\partial t A_{N_f}) Y_m^f) \\ &= F'(\xi_m^1) \left(Y_{m+1/2}^f - r(\partial t A_{N_f}) Y_m^f \right) \\ &= F'(\xi_m^1) r(\partial t A_{N_f}) \left(F(Y_m^f) \Delta t + G(Y_m^f) \mathfrak{d}_m W_{K_f} + \mathcal{G}(Y_m^f) \mathfrak{d}_m \mathcal{W}_{m, K_f} \right). \end{aligned} \quad (50)$$

As F, G and \mathcal{G} are of linear growth and $\mathbb{E}(\|Y_m^f\|_H^2) < \infty$ is uniformly bounded by Theorem 3.5, we have

$$\mathbf{I}_m^{f,1} = \Xi_m^{f,1} + O_m^{f,1},$$

where $\Xi_m^{f,1}, O_m^{f,1} : \Omega \rightarrow H$ are random variables such that $\mathbb{E}(\|\Xi_m^{f,1}\|_H^2) \leq C \partial t^2$, $\mathbb{E}(O_m^{f,1} | \mathcal{F}_{t_m}) = 0$ and $\mathbb{E}(\|O_m^{f,1}\|_H^2) \leq C \partial t$ holds for an independent constant $C > 0$. To bound $\mathbf{I}_m^{f,2}$, we use first order Taylor expansions around Y_m^f and \tilde{Y}_m^f to show that for some $\xi_m^2, \tilde{\xi}_m^2 \in H$ there holds

$$\begin{aligned} \mathbf{I}_m^{f,2} &= F(r(\partial t A_{N_f}) Y_m^f) - r(\partial t A_{N_f}) P_{N_f} F(Y_m^f) \\ &= F(Y_m^f) + F'(\xi_m^2) \left[r(\partial t A_{N_f}) Y_m^f - Y_m^f \right] - r(\partial t A_{N_f}) P_{N_f} F(Y_m^f) \\ &= \left[I - r(\partial t A_{N_f}) P_{N_f} \right] F(Y_m^f) + F'(\xi_m^2) \left[r(\partial t A_{N_f}) - I \right] Y_m^f \\ &= \left[I - r(\partial t A_{N_f}) P_{N_f} \right] \left(F(\tilde{Y}_m^f) + F'(\tilde{\xi}_m^2) (Y_m^f - \tilde{Y}_m^f) \right) + F'(\xi_m^2) \left[r(\partial t A_{N_f}) - I \right] \left(\tilde{Y}_m^f + Y_m^f - \tilde{Y}_m^f \right). \end{aligned} \quad (51)$$

Lemmas B.1 and C.2 thus show together with Items (iii) and (iv) in Assumption 2.4 that

$$\mathbb{E}(\|\mathbf{I}_m^{f,2}\|_H^2) \leq C \left(\partial t^\alpha + \partial t^2 + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} \right) \leq C \left(\partial t^\alpha + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} \right).$$

As $\|r(\partial t A_{N_f}) P_{N_f}\|_{\mathcal{L}(H)} \leq 1$ by Lemma B.1 this in turn shows that

$$r(\partial t A_{N_f}) P_{N_f} \mathbf{I}_m^f = r(\partial t A_{N_f}) P_{N_f} \left(\mathbf{I}_m^{f,1} + \mathbf{I}_m^{f,2} \right) \partial t = \Xi_m^{f,1} + O_m^{f,1}, \quad (52)$$

where $\Xi_m^{f,1}, O_m^{f,1} : \Omega \rightarrow H$ are random variables such that for an independent $C > 0$ there holds

$$\mathbb{E} \left(\|\Xi_m^{f,1}\|_H^2 \right) \leq C \partial t^2 \left(\partial t^\alpha + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} \right), \quad \mathbb{E} \left(O_m^{f,1} | \mathcal{F}_{t_m} \right) = 0 \quad \text{and} \quad \mathbb{E} \left(\|O_m^{f,1}\|_H^2 \right) \leq C \partial t^3.$$

We expand the second term \mathbf{II}_m^f in (49) via

$$\begin{aligned} \mathbf{II}_m^f &= \left[G(Y_{m+1/2}^f) - G(r(\partial t A_{N_f}) Y_m^f) \right] \mathfrak{d}_{m+1/2} W_{K_f} - 2r(\partial t A_{N_f}) P_{N_f} \mathcal{G}(Y_m^f) \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f} \\ &\quad + \left[G(r(\partial t A_{N_f}) Y_m^f) - r(\partial t A_{N_f}) P_{N_f} G(Y_m^f) \right] \mathfrak{d}_{m+1/2} W_{K_f} \\ &=: \mathbf{II}_m^{f,1} + \mathbf{II}_m^{f,2}. \end{aligned} \quad (53)$$

We observe that $\mathbb{E}(\mathbf{II}_m^{f,1} | \mathcal{F}_{t_m}) = 0$ and obtain by a second order Taylor expansion of G around $r(\partial t A_{N_f}) Y_m^f$

that for some $\xi_m^{\text{II}} \in H$ it holds

$$\begin{aligned}
\Pi_m^{f,1} &= G'(r(\partial t A_{N_f}) Y_m^f) \left(Y_{m+1/2}^f - r(\partial t A_{N_f}) Y_m^f \right) \mathfrak{d}_{m+1/2} W_{K_f} \\
&\quad + \frac{1}{2} G''(\xi_m^{\text{II}}) \left(Y_{m+1/2}^f - r(\partial t A_{N_f}) Y_m^f, Y_{m+1/2}^f - r(\partial t A_{N_f}) Y_m^f \right) \mathfrak{d}_{m+1/2} W_{K_f} \\
&\quad - 2r(\partial t A_{N_f}) P_{N_f} \mathcal{G}(Y_m^f) \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f} \\
&= G'(r(\partial t A_{N_f}) Y_m^f) r(\partial t A_{N_f}) P_{N_f} (F(Y_m^f) \partial t + G(Y_m^f) \mathfrak{d}_m W_{K_f} + \mathcal{G}(Y_m^f) \mathfrak{d}_m \mathcal{W}_{m,K_f}) \mathfrak{d}_{m+1/2} W_{K_f} \\
&\quad + \frac{1}{2} G''(\xi_m^{\text{II}}) \left(Y_{m+1/2}^f - r(\partial t A_{N_f}) Y_m^f, Y_{m+1/2}^f - r(\partial t A_{N_f}) Y_m^f \right) \mathfrak{d}_{m+1/2} W_{K_f} \\
&\quad - r(\partial t A_{N_f}) P_{N_f} G'(Y_m^f) (P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}) \mathfrak{d}_{m+1/2} W_{K_f} \\
&= G'(r(\partial t A_{N_f}) Y_m^f) r(\partial t A_{N_f}) P_{N_f} (F(Y_m^f) \partial t + \mathcal{G}(Y_m^f) \mathfrak{d}_m \mathcal{W}_{m,K_f}) \mathfrak{d}_{m+1/2} W_{K_f} \\
&\quad + \frac{1}{2} G''(\xi_m^{\text{II}}) \left(Y_{m+1/2}^f - r(\partial t A_{N_f}) Y_m^f, Y_{m+1/2}^f - r(\partial t A_{N_f}) Y_m^f \right) \mathfrak{d}_{m+1/2} W_{K_f} \\
&\quad + G'(r(\partial t A_{N_f}) Y_m^f) (r(\partial t A_{N_f}) P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}) \mathfrak{d}_{m+1/2} W_{K_f} \\
&\quad - r(\partial t A_{N_f}) P_{N_f} G'(Y_m^f) (P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}) \mathfrak{d}_{m+1/2} W_{K_f} \\
&=: \widetilde{\text{I}}_m^{f,1} + \widetilde{\text{II}}_m^{f,1} + \widetilde{\text{III}}_m^{f,1}.
\end{aligned}$$

The second identity follows by Proposition 3.1 since

$$\mathcal{G}(Y_m^f) \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f} = G'(Y_m^f) (P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}) \mathfrak{d}_{m+1/2} W_{K_f}.$$

As F and \mathcal{G} are of linear growth (see Proposition 3.1) and G' is bounded, it follows by the independence of Y_m , $\mathfrak{d}_m \mathcal{W}_{m,K_f}$ and $\mathfrak{d}_{m+1/2} W_{K_f}$, Lemma A.1 and Theorem 3.5 that

$$\mathbb{E} \left(\left\| \widetilde{\text{I}}_m^{f,1} \right\|_H^2 \right) \leq C \left(1 + \mathbb{E} \left(\|Y_m^f\|_H^2 \right) \right) \mathfrak{d}t^3 \leq C \mathfrak{d}t^3.$$

Similarly, Lemma C.1 yields with the boundedness of G'' that

$$\mathbb{E} \left(\left\| \widetilde{\text{II}}_m^{f,1} \right\|_H^2 \right) \leq C \mathbb{E} \left(\left\| Y_{m+1/2}^f - r(\partial t A_{N_f}) Y_m^f \right\|_H^4 \right) \mathfrak{d}t \leq C \mathfrak{d}t^3.$$

We use Taylor expansion to split the integrand in $\widetilde{\text{III}}_m^{f,1}$ for some $\widetilde{\xi}_m^{\text{III}} \in H$ into

$$\begin{aligned}
&G'(r(\partial t A_{N_f}) Y_m^f) (r(\partial t A_{N_f}) P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}) - r(\partial t A_{N_f}) P_{N_f} G'(Y_m^f) (P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}) \\
&= [G'(r(\partial t A_{N_f}) Y_m^f) - G'(Y_m^f)] (r(\partial t A_{N_f}) P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}) \\
&\quad + G'(Y_m^f) (r(\partial t A_{N_f}) P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}) - G'(Y_m^f) (P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}) \\
&\quad + G'(Y_m^f) (P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}) - r(\partial t A_{N_f}) P_{N_f} G'(Y_m^f) (P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}) \\
&= G''(\widetilde{\xi}_m^{\text{III}}) [(r(\partial t A_{N_f}) - I) Y_m^f, r(\partial t A_{N_f}) P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}] \\
&\quad + G'(Y_m^f) [(r(\partial t A_{N_f}) - I) P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}] \\
&\quad + (I - r(\partial t A_{N_f}) P_{N_f}) G'(Y_m^f) (P_{N_f} G(Y_m^f) \mathfrak{d}_m W_{K_f}).
\end{aligned}$$

We then use Assumption 2.4 (iii) together with Corollary A.2 and Theorem 3.5 to estimate

$$\begin{aligned}
\mathbb{E} \left(\left\| \widetilde{\Pi}_m^{f,1} \right\|_H^2 \right) &\leq 3\mathbb{E} \left(\left\| G''(\widetilde{\xi}_m^{\text{III}}) [(r(\partial t A_{N_f}) - I)Y_m^f, r(\partial t A_{N_f})P_{N_f}G(Y_m^f)\mathfrak{d}_m W_{K_f}] \right\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 \right) \mathfrak{d}t \\
&\quad + 3\mathbb{E} \left(\left\| G'(Y_m^f) [(r(\partial t A_{N_f}) - I)P_{N_f}G(Y_m^f)\mathfrak{d}_m W_{K_f}] \right\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 \right) \mathfrak{d}t \\
&\quad + 3\mathbb{E} \left(\left\| (I - r(\partial t A_{N_f})P_{N_f})G'(Y_m^f)(P_{N_f}G(Y_m^f)\mathfrak{d}_m W_{K_f}) \right\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 \right) \mathfrak{d}t \\
&\leq C\mathbb{E} \left(\left\| (r(\partial t A_{N_f}) - I)Y_m^f \right\|_H^2 \left\| (r(\partial t A_{N_f})P_{N_f}G(Y_m^f)\mathfrak{d}_m W_{K_f}) \right\|_H^2 \right) \mathfrak{d}t \\
&\quad + C\mathbb{E} \left(\left\| (r(\partial t A_{N_f}) - I)P_{N_f}G(Y_m^f) \right\|_H^2 \right) \mathfrak{d}t^2 \\
&\quad + C\mathbb{E} \left(\left\| (r(\partial t A_{N_f})P_{N_f} - I)G'(Y_m^f) \right\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))}^2 \left\| G(Y_m^f)\mathfrak{d}_m W_{K_f} \right\|_H^2 \right) \mathfrak{d}t \\
&\leq C\mathbb{E} \left(\left\| (r(\partial t A_{N_f}) - I)Y_m^f \right\|_H^4 \right)^{1/2} \mathbb{E} \left(\left\| r(\partial t A_{N_f})P_{N_f}G(Y_m^f)\mathfrak{d}_m W_{K_f} \right\|_H^4 \right)^{1/2} \mathfrak{d}t \\
&\quad + C\mathbb{E} \left(\left\| (r(\partial t A_{N_f})P_{N_f} - I)G'(Y_m^f) \right\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 \right) \mathfrak{d}t^2 \\
&\quad + C\mathbb{E} \left(\left\| (r(\partial t A_{N_f})P_{N_f} - I)G'(Y_m^f) \right\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))}^4 \right)^{1/2} \mathbb{E} \left(\left\| G(Y_m^f)\mathfrak{d}_m W_{K_f} \right\|_H^4 \right)^{1/2} \mathfrak{d}t \\
&\leq C\mathbb{E} \left(\left\| (r(\partial t A_{N_f}) - I)(\widetilde{Y}_m^f + Y_m^f - \widetilde{Y}_m^f) \right\|_H^4 \right)^{1/2} \mathfrak{d}t^2 \\
&\quad + C\mathbb{E} \left(\left\| (r(\partial t A_{N_f})P_{N_f} - I)(G(\widetilde{Y}_m^f) + G(Y_m^f) - G(\widetilde{Y}_m^f)) \right\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 \right) \mathfrak{d}t^2 \\
&\quad + C\mathbb{E} \left(\left\| (r(\partial t A_{N_f})P_{N_f} - I)(G'(\widetilde{Y}_m^f) + G'(Y_m^f) - G'(\widetilde{Y}_m^f)) \right\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))}^4 \right)^{1/2} \mathfrak{d}t^2.
\end{aligned}$$

In the last step we have used that G is of linear growth together with Theorem 3.5.

Assumption 2.4 (iv) together with Lemmas B.1 and C.2 then yields

$$\begin{aligned}
&\mathbb{E} \left(\left\| \widetilde{\Pi}_m^{f,1} \right\|_H^2 \right) \\
&\leq C\mathfrak{d}t^2 \left(\mathbb{E} \left(\left\| (r(\partial t A_{N_f}) - I)\widetilde{Y}_m^f \right\|_H^4 \right)^{1/2} + \mathbb{E} \left(\left\| Y_m^f - \widetilde{Y}_m^f \right\|_H^4 \right)^{1/2} \right) \\
&\quad + C\mathfrak{d}t^2 \left(\mathbb{E} \left(\left\| (r(\partial t A_{N_f}) - I)G(\widetilde{Y}_m^f) \right\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 \right) + \mathbb{E} \left(\left\| G(Y_m^f) - G(\widetilde{Y}_m^f) \right\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 \right) \right) \\
&\quad + C\mathfrak{d}t^2 \left(\mathbb{E} \left(\left\| (r(\partial t A_{N_f}) - I)G'(\widetilde{Y}_m^f) \right\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))}^4 \right)^{1/2} + \mathbb{E} \left(\left\| G'(Y_m^f) - G'(\widetilde{Y}_m^f) \right\|_{\mathcal{L}(H, \mathcal{L}_{HS}(\mathcal{U}, H))}^4 \right)^{1/2} \right) \\
&\leq C\mathfrak{d}t^2 \left(\mathfrak{d}t^\alpha + N_f^{-2\tilde{\alpha}} \right) \mathbb{E} \left(\left\| \widetilde{Y}_m^f \right\|_{\dot{H}^\alpha}^4 + \left\| G(\widetilde{Y}_m^f) \right\|_{\mathcal{L}_{HS}(\mathcal{U}, \dot{H}^\alpha)}^4 + \left\| G'(\widetilde{Y}_m^f) \right\|_{\mathcal{L}(\dot{H}^\alpha, \mathcal{L}_{HS}(\mathcal{U}, \dot{H}^\alpha))}^4 + \left\| Y_m^f - \widetilde{Y}_m^f \right\|_H^4 \right)^{1/2} \\
&\leq C\mathfrak{d}t^2 \left(\mathfrak{d}t^\alpha + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} \right).
\end{aligned}$$

Thus, we obtain for $\Pi_m^{f,1} : \Omega \rightarrow H$ that $\mathbb{E}(\Pi_m^{f,1} | \mathcal{F}_{t_m}) = 0$ and $\mathbb{E}(\|\Pi_m^{f,1}\|_H^2) \leq C\mathfrak{d}t^2 \left(\mathfrak{d}t^\alpha + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} \right)$.

For the remaining term $\Pi_m^{f,2}$ in (53), we have that $\mathbb{E}(\Pi_m^{f,2} | \mathcal{F}_{t_m}) = 0$. Moreover, with the Itô isometry

and analogous calculations as in (50) and (51) we obtain the bound

$$\begin{aligned}\mathbb{E}(\|\mathbb{II}_m^{f,2}\|_H) &= \mathbb{E}\left(\left\|\left[G(r(\mathfrak{d}tA_{N_f})Y_m^f) - r(\mathfrak{d}tA_{N_f})P_{N_f}G(Y_m^f)\right] \mathfrak{d}_{m+1/2}W_{K_f}\right\|_H^2\right) \\ &\leq C\mathfrak{d}t\left(\mathfrak{d}t^\alpha + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta}\right).\end{aligned}$$

To bound the last term \mathbb{III}_m^f in (49), we first observe again that $\mathbb{E}(\mathbb{III}_m^f|\mathcal{F}_n) = 0$. We further use the BDG inequality in Lemma A.1 to obtain for some $\xi_m^{1,\text{III}}, \xi_m^{2,\text{III}} \in H$ that

$$\begin{aligned}\mathbb{E}\left(\left\|\mathbb{III}_m^f\right\|_H^2\right) &= \mathbb{E}\left(\left\|\left[\mathcal{G}(Y_{m+1/2}^f) - r(\mathfrak{d}tA_{N_f})P_{N_f}\mathcal{G}(Y_m^f)\right] \mathfrak{d}_{m+1/2}\mathcal{W}_{m,K_f}\right\|_H^2\right) \\ &\leq C\mathbb{E}\left(\left\|\mathcal{G}(Y_{m+1/2}^f) - r(\mathfrak{d}tA_{N_f})P_{N_f}\mathcal{G}(Y_m^f)\right\|_{\mathcal{L}_{HS}(\mathcal{L}_1(\mathcal{U}),H)}^2\right)\mathfrak{d}t^2 \\ &\leq C\mathbb{E}\left(\left\|\mathcal{G}(Y_{m+1/2}^f) - \mathcal{G}(r(\mathfrak{d}tA_{N_f})P_{N_f}Y_m^f)\right\|_{\mathcal{L}_{HS}(\mathcal{L}_1(\mathcal{U}),H)}^2\right)\mathfrak{d}t^2 \\ &\quad + C\mathbb{E}\left(\left\|\mathcal{G}(r(\mathfrak{d}tA_{N_f})P_{N_f}Y_m^f) - \mathcal{G}(Y_m^f)\right\|_{\mathcal{L}_{HS}(\mathcal{L}_1(\mathcal{U}),H)}^2\right)\mathfrak{d}t^2 \\ &\quad + C\mathbb{E}\left(\left\|\mathcal{G}(Y_m^f) - r(\mathfrak{d}tA_{N_f})P_{N_f}\mathcal{G}(Y_m^f)\right\|_{\mathcal{L}_{HS}(\mathcal{L}_1(\mathcal{U}),H)}^2\right)\mathfrak{d}t^2 \\ &\leq C\mathbb{E}\left(\left\|\mathcal{G}'(\xi_m^{1,\text{III}})(Y_{m+1/2}^f - r(\mathfrak{d}tA_{N_f})Y_m^f)\right\|_{\mathcal{L}_{HS}(\mathcal{L}_1(\mathcal{U}),H)}^2\right)\mathfrak{d}t^2 \\ &\quad + C\mathbb{E}\left(\left\|\mathcal{G}'(\xi_m^{2,\text{III}})(I - r(\mathfrak{d}tA_{N_f}))Y_m^f\right\|_{\mathcal{L}_{HS}(\mathcal{L}_1(\mathcal{U}),H)}^2\right)\mathfrak{d}t^2 \\ &\quad + C\mathbb{E}\left(\left\|(I - r(\mathfrak{d}tA_{N_f}))\mathcal{G}(Y_m^f)\right\|_{\mathcal{L}_{HS}(\mathcal{L}_1(\mathcal{U}),H)}^2\right)\mathfrak{d}t^2.\end{aligned}$$

Hölder's inequality, Proposition 3.1 and now with similar calculations as for $\widetilde{\mathbb{III}}_m^{f,1}$ that

$$\begin{aligned}\mathbb{E}\left(\left\|\mathbb{III}_m^f\right\|_H^2\right) &\leq C\mathbb{E}\left(1 + \|\xi_m^{1,\text{III}}\|_H^4\right)^{1/2}\mathbb{E}\left(\left\|Y_{m+1/2}^f - r(\mathfrak{d}tA_{N_f})Y_m^f\right\|_H^4\right)^{1/2}\mathfrak{d}t^2 \\ &\quad + C\mathbb{E}\left(1 + \|\xi_m^{2,\text{III}}\|_H^4\right)^{1/2}\mathbb{E}\left(\left\|(I - r(\mathfrak{d}tA_{N_f}))(\tilde{Y}_m^f + Y_m^f - \tilde{Y}_m^f)\right\|_H^4\right)^{1/2}\mathfrak{d}t^2 \\ &\quad + C\mathbb{E}\left(\left\|(I - r(\mathfrak{d}tA_{N_f}))\left(\mathcal{G}(\tilde{Y}_m^f) + \mathcal{G}(Y_m^f) - \mathcal{G}(\tilde{Y}_m^f)\right)\right\|_{\mathcal{L}_{HS}(\mathcal{L}_1(\mathcal{U}),H)}^2\right)\mathfrak{d}t^2 \\ &\leq C\mathfrak{d}t^2\left(\mathfrak{d}t^\alpha + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta}\right).\end{aligned}$$

□

A similar result to Proposition C.3 also holds for the antithetic fine approximation Y^a .

Corollary C.4. *Let Assumptions 2.4 and 3.2 hold. Then, for any $m = 0, \dots, M-1$ there holds that*

$$\begin{aligned}Y_{m+1}^a &=: r(\mathfrak{d}tA_{N_f})^2P_{N_f}\left(Y_m^a + F(Y_m^a)\Delta t + G(Y_m^a)\Delta_m W_{K_f} + \mathcal{G}(Y_m^a)\Delta_m \mathcal{W}_{m,K_f}\right) \\ &\quad + r(\mathfrak{d}tA_{N_f})^2P_{N_f}\mathcal{G}(Y_m^a)\left(\mathfrak{d}_{m+1/2}W_{K_f} \otimes \mathfrak{d}_m W_{K_f} - \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2}W_{K_f}\right) \\ &\quad + \Xi_m^a + O_m^a,\end{aligned}\tag{54}$$

where $\Xi_m^a, O_m^a : \Omega \rightarrow H$ are random variables such that

$$\begin{aligned}\mathbb{E}\left(\left\|\Xi_m^a\right\|_H^2\right) &\leq C\mathfrak{d}t^2\left(M^{-\min(\alpha,2)} + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta}\right), \\ \mathbb{E}\left(O_m^a \mid \mathcal{F}_{t_m}\right) &= 0 \quad \text{and} \quad \mathbb{E}\left(\left\|O_m^a\right\|_H^2\right) \leq C\mathfrak{d}t\left(M^{-\min(\alpha,2)} + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta}\right).\end{aligned}$$

The constant C is independent of $\delta t = T/2M, N_f$ and K_f .

Proof. Equations (22) and (23) show that

$$\begin{aligned} Y_{m+1}^a &= r(\partial_t A_{N_f})^2 P_{N_f} (Y_m^a + F(Y_m^a)\Delta t + G(Y_m^a)\Delta_m W_{K_f} + \mathcal{G}(Y_m^a)\Delta_m \mathcal{W}_{n,K_f}) \\ &\quad + r(\partial_t A_{N_f})^2 P_{N_f} \mathcal{G}(Y_m^a) (\mathfrak{d}_{m+1/2} W_{K_f} \otimes \mathfrak{d}_m W_{K_f} - \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f}) \\ &\quad + r(\partial_t A_{N_f}) P_{N_f} [I_m^a + \text{II}_m^a + \text{III}_m^a], \end{aligned}$$

where sign change in the third line is due to the swapping of the increments $\mathfrak{d}_{m+1/2} W_{K_f}$ and $\delta_n W_{K_f}$ in the antithetic estimator. The remainder terms are given by

$$\begin{aligned} I_m^a &:= [F(Y_{m+1/2}^a) - r(\partial_t A_{N_f}) P_{N_f} F(Y_m^a)] \partial t, \\ \text{II}_m^a &:= \left[G(Y_{m+1/2}^a) - r(\partial_t A_{N_f}) P_{N_f} G(Y_m^a) \right] \mathfrak{d}_m W_{K_f} - 2r(\partial_t A_{N_f}) P_{N_f} \mathcal{G}(Y_m^a) \mathfrak{d}_{m+1/2} W_{K_f} \otimes \mathfrak{d}_m W_{K_f}, \\ \text{III}_m^a &:= \left[\mathcal{G}(Y_{m+1/2}^f) - r(\partial_t A_{N_f}) P_{N_f} \mathcal{G}(Y_m^a) \right] \mathfrak{d}_m \mathcal{W}_{m,K_f}. \end{aligned}$$

The claim follows analogously to Proposition C.3. \square

Proposition C.3 and Corollary C.4 are now combined to derive a similar difference equation for the (antithetic) average $\bar{Y}_m := 1/2(Y_m^f + Y_m^a)$ for $m = 0, \dots, M$.

Proposition C.5. *Let Assumptions 2.4 and 3.2. Then, for any $m = 0, \dots, M - 1$ there holds that*

$$\bar{Y}_{m+1} = r(\partial_t A_{N_f})^2 P_{N_f} (\bar{Y}_m + F(\bar{Y}_m)\Delta t + G(\bar{Y}_m)\Delta_m W_{K_f} + \mathcal{G}(\bar{Y}_m)\Delta_m \mathcal{W}_{m,K_f}) + \bar{\Xi}_m + \bar{O}_m,$$

where $\bar{\Xi}_m, \bar{O}_m : \Omega \rightarrow H$ are random variables such that

$$\begin{aligned} \mathbb{E} \left(\|\bar{\Xi}_m\|_H^2 \right) &\leq C\Delta t^2 \left(M^{-\min(\alpha,2)} + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} \right), \\ \mathbb{E} (\bar{O}_m | \mathcal{F}_{t_m}) &= 0 \quad \text{and} \quad \mathbb{E} \left(\|\bar{O}_m\|_H^2 \right) \leq C\Delta t \left(M^{-\min(\alpha,2)} + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} \right). \end{aligned}$$

The constant C is independent of $\Delta t = T/M, N_f$ and K_f .

Proof. Lemma C.3 and Corollary C.4 show that

$$\begin{aligned} \bar{Y}_{m+1} &= r(\partial_t A_{N_f})^2 P_{N_f} (\bar{Y}_m + F(\bar{Y}_m)\Delta t + G(\bar{Y}_m)\Delta_m W_{K_f} + \mathcal{G}(\bar{Y}_m)\Delta_m \mathcal{W}_{m,K_f}) \\ &\quad + r(\partial_t A_{N_f})^2 P_{N_f} \left(\frac{F(Y_m^f) + F(Y_m^a)}{2} - F(\bar{Y}_m) \right) \Delta t \\ &\quad + r(\partial_t A_{N_f})^2 P_{N_f} \left(\frac{G(Y_m^f) + G(Y_m^a)}{2} - G(\bar{Y}_m) \right) \Delta_m W_{K_f} \\ &\quad + r(\partial_t A_{N_f})^2 P_{N_f} \left(\frac{\mathcal{G}(Y_m^f) + \mathcal{G}(Y_m^a)}{2} - \mathcal{G}(\bar{Y}_m) \right) \Delta_m \mathcal{W}_{m,K_f} \\ &\quad + \frac{r(\partial_t A_{N_f})^2 P_{N_f}}{2} (\mathcal{G}(Y_m^a) - \mathcal{G}(Y_m^f)) (\mathfrak{d}_{m+1/2} W_{K_f} \otimes \mathfrak{d}_m W_{K_f} - \mathfrak{d}_m W_{K_f} \otimes \mathfrak{d}_{m+1/2} W_{K_f}) \\ &\quad + \frac{1}{2} (\Xi_m^f + \Xi_m^a + O_m^f + O_m^a) \\ &=: r(\partial_t A_{N_f})^2 P_{N_f} (\bar{Y}_m + F(\bar{Y}_m)\Delta t + G(\bar{Y}_m)\Delta_m W_{K_f} + \mathcal{G}(\bar{Y}_m)\Delta_m \mathcal{W}_{m,K_f}) \\ &\quad + \bar{\text{I}}_m + \bar{\text{II}}_m + \bar{\text{III}}_m + \bar{\text{IV}}_m + \frac{1}{2} (\Xi_m^f + \Xi_m^a + O_m^f + O_m^a) \end{aligned}$$

To bound the first term $\bar{\text{I}}_m$, we use a second order Taylor expansion of F around \bar{Y}_m together with $\bar{Y}_m = \frac{1}{2}(Y_m^f + Y_m^a)$ to obtain for some $\xi_m^f, \xi_m^a \in H$

$$\begin{aligned} \frac{F(Y_m^f) + F(Y_m^a)}{2} - F(\bar{Y}_m) &= \frac{F''(\xi_m^f) - F''(\xi_m^a)}{4} (Y_m^f - \bar{Y}_m, Y_m^f - \bar{Y}_m) \\ &= \frac{F''(\xi_m^f) - F''(\xi_m^a)}{16} (Y_m^f - Y_m^a, Y_m^f - Y_m^a). \end{aligned}$$

Assumption 3.2 (iii) and the triangle inequality further show that

$$\mathbb{E} \left(\|Y_m^f - Y_m^a\|_H^4 \right) \leq C \mathbb{E} \left(\|Y_m^f - X(t_m)\|_H^4 \right) \leq C \left(\delta t^2 + N_f^{-4\tilde{\alpha}} + K_f^{-4\beta} \right).$$

We then use the global bound on F'' and Jensen's inequality together with Assumption 3.2 (iii) to derive

$$\begin{aligned} \mathbb{E} \left(\|\bar{\mathbb{I}}_m\|_H^2 \right) &\leq \mathbb{E} \left(\left\| \frac{F(Y_m^f) + F(Y_m^a)}{2} - F(\bar{Y}_m) \right\|_H^2 \right) \Delta t^2 \\ &\leq C \mathbb{E} \left(\|Y_m^f - X(m\Delta t) + X(m\Delta t) - Y_m^a\|_H^4 \right) \Delta t^2 \\ &\leq C \Delta t^2 \left(\Delta t^2 + N_f^{-4\tilde{\alpha}} + K_f^{-4\beta} \right). \end{aligned}$$

We observe that $\mathbb{E}(\bar{\mathbb{I}}_m | \mathcal{F}_{t_m}) = 0$ holds for the second term, and arrive with Itô's isometry and similar calculations as for $\bar{\mathbb{I}}_m$ at

$$\mathbb{E} \left(\|\bar{\mathbb{I}}_m\|_H^2 \right) \leq C \Delta t \left(\Delta t^2 + N_f^{-4\tilde{\alpha}} + K_f^{-4\beta} \right).$$

For the next, we first note that $\mathbb{E}(\bar{\mathbb{I}}\bar{\mathbb{I}}_m | \mathcal{F}_{t_m}) = 0$. To bound $\bar{\mathbb{I}}\bar{\mathbb{I}}_m$ in mean-square, we use Proposition 3.1 and first order expansion of \mathcal{G} to obtain for some $\tilde{\xi}_m^f, \tilde{\xi}_m^a \in H$

$$\frac{\mathcal{G}(Y_m^f) + \mathcal{G}(Y_m^a)}{2} - \mathcal{G}(\bar{Y}_m) = \frac{\mathcal{G}'(\tilde{\xi}_m^f) - \mathcal{G}'(\tilde{\xi}_m^a)}{2} (Y_m^f - \bar{Y}_m) = \frac{\mathcal{G}'(\tilde{\xi}_m^f) - \mathcal{G}'(\tilde{\xi}_m^a)}{4} (Y_m^f - Y_m^a).$$

Since the intermediate points $\tilde{\xi}_m^f, \tilde{\xi}_m^a \in H$ are convex combinations of Y_m^f and Y_m^a there holds by Lemma A.1, Proposition 3.1 and Theorem 3.5

$$\begin{aligned} \mathbb{E} \left(\|\bar{\mathbb{I}}\bar{\mathbb{I}}_m\|_H^2 \right) &\leq C \mathbb{E} \left(\left\| (\mathcal{G}'(\tilde{\xi}_m^f) - \mathcal{G}'(\tilde{\xi}_m^a))(Y_m^f - Y_m^a) \right\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 \right) \Delta t^2 \\ &\leq C \mathbb{E} \left(\left(1 + \|Y_m^f\|_H^2 + \|Y_m^a\|_H^2 \right) \|Y_m^f - Y_m^a\|_H^2 \right) \Delta t^2 \\ &\leq C \mathbb{E} \left(\|Y_m^f - Y_m^a\|_H^4 \right)^{1/2} \Delta t^2 \\ &\leq C \Delta t^2 \left(\Delta t + N_f^{-2\tilde{\alpha}} + K_f^{-2\beta} \right) \\ &\leq C \Delta t \left(\Delta t^2 + \Delta t N_f^{-2\tilde{\alpha}} + \Delta t K_f^{-2\beta} \right) \\ &\leq C \Delta t \left(\Delta t^2 + N_f^{-4\tilde{\alpha}} + K_f^{-4\beta} \right), \end{aligned}$$

where we have used Young's inequality for the final estimate.

As \mathcal{G} is of linear growth by Proposition 3.1, we obtain analogously that $\mathbb{E}(\bar{\mathbb{I}}\bar{\mathbb{V}}_m | \mathcal{F}_{t_m}) = 0$ and

$$\mathbb{E} \left(\|\bar{\mathbb{I}}\bar{\mathbb{V}}_m\|_H^2 \right) \leq C \Delta t \left(\Delta t^2 + N_f^{-4\tilde{\alpha}} + K_f^{-4\beta} \right).$$

The dominating remainder terms in the expansion of \bar{Y}_{m+1} are thus $\Xi_m^f, \Xi_m^a, O_m^f, O_m^a$, and the claim follows from Proposition C.3 and Corollary C.4. \square

We are now ready to proof our main result.

Proof of Theorem 4.1. Define $e_{m+1} := \bar{Y}_{m+1} - Y_{m+1}^c$ for any $m = 0, \dots, M-1$ and assume again without loss of generality that $\alpha \in [1, 2]$. By Proposition C.5 it holds that

$$\begin{aligned}
e_{m+1} &= r(\mathfrak{d}tA_{N_f})^2 \bar{Y}_m - r(\Delta tA_N) Y_m^c \\
&\quad + (r(\mathfrak{d}tA_{N_f})^2 P_{N_f} F(\bar{Y}_m) - r(\Delta tA_N) P_N F(Y_m^c)) \Delta t \\
&\quad + (r(\mathfrak{d}tA_{N_f})^2 P_{N_f} G(\bar{Y}_m) - r(\Delta tA_N) P_N G(Y_m^c)) \Delta_m W_{K_f} \\
&\quad + (r(\mathfrak{d}tA_{N_f})^2 P_{N_f} \mathcal{G}(\bar{Y}_m) - r(\Delta tA_N) P_N \mathcal{G}(Y_m^c)) \Delta_m \mathcal{W}_{m, K_f} \\
&\quad + r(\Delta tA_N) P_N G(Y_m^c) (\Delta_m W_{K_f} - \Delta_m W_M) \\
&\quad + r(\Delta tA_N) P_N \mathcal{G}(Y_m^c) (\Delta_m \mathcal{W}_{m, K_f} - \Delta_m \mathcal{W}_{m, K}) \\
&\quad + \bar{\Xi}_m + \bar{O}_m.
\end{aligned} \tag{55}$$

We now re-iterate the representation of \bar{Y}_m and Y_m^c to obtain

$$\begin{aligned}
e_{m+1} &= r(\mathfrak{d}tA_{N_f})^{2m} \bar{Y}_0 - r(\Delta tA_N)^m Y_0^c \\
&\quad + \sum_{j=0}^m \left((r(\mathfrak{d}tA_{N_f})^{2(m-j+1)} P_{N_f} F(\bar{Y}_j) - r(\Delta tA_N)^{m-j+1} P_N F(Y_j^c)) \Delta t \right. \\
&\quad + \sum_{j=0}^m \left(r(\mathfrak{d}tA_{N_f})^{2(m-j+1)} P_{N_f} G(\bar{Y}_j) - r(\Delta tA_N)^{m-j+1} P_N G(Y_j^c) \right) \Delta_j W_{K_f} \\
&\quad + \sum_{j=0}^m \left(r(\mathfrak{d}tA_{N_f})^{2(m-j+1)} P_{N_f} \mathcal{G}(\bar{Y}_j) - r(\Delta tA_N)^{m-j+1} P_N \mathcal{G}(Y_j^c) \right) \Delta_j \mathcal{W}_{m, K_f} \\
&\quad + \sum_{j=0}^m r(\Delta tA_N)^{m-j+1} P_N G(Y_j^c) (\Delta_j W_{K_f} - \Delta_j W_M) \\
&\quad + \sum_{j=0}^m r(\mathfrak{d}tA_{N_f})^{m-j+1} P_N \mathcal{G}(Y_j^c) (\Delta_j \mathcal{W}_{m, K_f} - \Delta_j \mathcal{W}_{m, K}) \\
&\quad + \sum_{j=0}^m r(\mathfrak{d}tA_{N_f})^{2(n-j)} (\bar{\Xi}_j + \bar{O}_j) \\
&=: \text{I} + \sum_{j=0}^m \text{II}_j + \text{III}_j + \text{IV}_j + \text{V}_j + \text{VI}_j + r(\mathfrak{d}tA_{N_f})^{2(j-1)} (\bar{\Xi}_j + \bar{O}_j).
\end{aligned} \tag{56}$$

The first term I is bounded Lemma B.1 and Assumption 3.2 (ii) since $X_0 \in L^8(\Omega, \dot{H}^\alpha)$ by

$$\begin{aligned}
\mathbb{E} \left(\|\text{I}\|_H^2 \right) &\leq 3\mathbb{E} \left(\left\| (r(\mathfrak{d}tA_{N_f})^{2m} - r(\Delta tA_{N_f})^m) \bar{Y}_0 \right\|_H^2 \right) + 3\mathbb{E} \left(\left\| r(\Delta tA_{N_f})^m - r(\Delta tA_N)^m \right\|_H^2 \right) \\
&\quad + 3\mathbb{E} \left(\left\| r(\Delta tA_N)^m (\bar{Y}_0 - Y_0^c) \right\|_H^2 \right) \\
&\leq C\mathbb{E} \left(\left\| (r(\mathfrak{d}tA_{N_f})^{2m} - r(\Delta tA_{N_f})^m) P_{N_f} X_0 \right\|_H^2 \right) \\
&\quad + C\mathbb{E} \left(\left\| (r(\Delta tA_{N_f})^m - r(\Delta tA_N)^m) P_{N_f} X_0 \right\|_H^2 \right) \\
&\quad + C \left(\mathbb{E} \left(\left\| P_{N_f} X_0 - X_0 \right\|_H^2 \right) + \mathbb{E} \left(\left\| X_0 - P_N X_0 \right\|_H^2 \right) \right) \\
&\leq C \left(M^{-\alpha} + N^{-2\bar{\alpha}} \right).
\end{aligned}$$

To bound the terms II_j , we define the semi-discrete averages $\tilde{Y}_j := \frac{1}{2} \left(\tilde{Y}_j^f + \tilde{Y}_j^a \right)$ for $j = 0, \dots, m$. We

then obtain for any $j = 0, \dots, n$ by a first order Taylor expansion for some $\tilde{\xi}_j, \xi_j^c \in H$ that

$$\begin{aligned}
\Pi_j &= \left(r(\mathfrak{d}tA_{N_f})^{2(m-j+1)}P_{N_f} - r(\Delta tA_{N_f})^{m-j+1}P_{N_f} \right) \left(F(\bar{Y}_j) - F(\tilde{Y}_j) + F(\tilde{Y}_j) \right) \Delta t \\
&\quad + \left(r(\Delta tA_{N_f})^{m-j+1}P_{N_f} - r(\Delta tA_N)^{m-j+1}P_N \right) \left(F(\bar{Y}_j) - F(\tilde{Y}_j) + F(\tilde{Y}_j) \right) \Delta t \\
&\quad + r(\Delta tA_N)^{m-j+1}P_N \left(F(\bar{Y}_m) - F(Y_m^c) \right) \Delta t \\
&= \left(r(\mathfrak{d}tA_{N_f})^{2(m-j+1)}P_{N_f} - r(\Delta tA_{N_f})^{m-j+1}P_{N_f} \right) \left(F'(\tilde{\xi}_j)(\bar{Y}_j - \tilde{Y}_j) + F(\tilde{Y}_j) \right) \Delta t \\
&\quad + \left(r(\Delta tA_{N_f})^{m-j+1}P_{N_f} - r(\Delta tA_N)^{m-j+1}P_N \right) \left(F'(\tilde{\xi}_j)(\bar{Y}_j - \tilde{Y}_j) + F(\tilde{Y}_j) \right) \Delta t \\
&\quad + r(\Delta tA_N)^{m-j+1}P_N \left(F'(\xi_j^c)(\bar{Y}_j - Y_j^c) \right) \Delta t.
\end{aligned}$$

Lemmas B.1 and C.2 then show together with Assumption 3.2 (iv) that

$$\begin{aligned}
\mathbb{E} \left(\|\Pi_j\|_H^2 \right) &\leq C\Delta t^2 \mathbb{E} \left(\|\bar{Y}_j - \tilde{Y}_j\|_H^2 \right) \\
&\quad + C\Delta t^2 \mathbb{E} \left(\left\| \left(r(\mathfrak{d}tA_{N_f})^{2(m-j+1)}P_{N_f} - r(\Delta tA_{N_f})^{m-j+1}P_{N_f} \right) F(\tilde{Y}_j) \right\|_H^2 \right) \\
&\quad + C\Delta t^2 \mathbb{E} \left(\left\| \left(r(\Delta tA_{N_f})^{m-j+1}P_{N_f} - S_{N_f}((m-j+1)\Delta t) \right) F(\tilde{Y}_j) \right\|_H^2 \right) \\
&\quad + C\Delta t^2 \mathbb{E} \left(\left\| \left(S_{N_f}((m-j+1)\Delta t) - r(\Delta tA_N)^{m-j+1}P_N \right) F(\tilde{Y}_j) \right\|_H^2 + \|e_j\|_H^2 \right) \\
&\leq C\Delta t^2 \mathbb{E} \left(\Delta t^2 + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} + (\Delta t^\alpha + N_f^{-2\tilde{\alpha}}) \|F(\tilde{Y}_j)\|_{H^\alpha} + \|e_j\|_H^2 \right) \\
&\leq C\Delta t^2 \left(M^{-\alpha} + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} + \mathbb{E} \left(\|e_j\|_H^2 \right) \right).
\end{aligned}$$

By Lemma A.1 and with similar calculations as for Π_m^e we further obtain

$$\mathbb{E} \left(\|\text{III}_j\|_H^2 \right) + \mathbb{E} \left(\|\text{IV}_j\|_H^2 \right) \leq C\Delta t \left(M^{-\alpha} + N_f^{-2\tilde{\alpha}} + K_f^{-4\beta} + \mathbb{E} \left(\|e_j\|_H^2 \right) \right).$$

The fifth term V_j is bounded by Corollary A.2 and Theorem 3.5 via

$$\mathbb{E} \left(\|V_j\|_H^2 \right) = \Delta t \mathbb{E} \left(\|G(Y_j^c)\|_{\mathcal{L}_{HS}(\mathcal{U}, H)}^2 \right) \sum_{j=K+1}^{K_f} \eta_j \leq C\Delta t \mathbb{E} \left(1 + \|Y_j^c\|_H^2 \right) \sum_{j=K+1}^{K_f} \eta_j \leq C\Delta t K^{-2\beta},$$

where we have used that $\eta_j = \mathcal{O}(j^{-(1+\varepsilon)-2\beta})$ for arbitrary small $\varepsilon > 0$ in the last step, cf. Remark 3.3. Similarly, Lemma A.1, Proposition 3.1 and Theorem 3.5 show that

$$\mathbb{E} \left(\|\text{VI}_j\|_H^2 \right) \leq C\Delta t^2 \mathbb{E} \left(\|\mathcal{G}(Y_j^c)\|_{\mathcal{L}_{HS}(\mathcal{L}_1(\mathcal{U}, H))}^2 \right) \sum_{j=K+1}^{K_f} \eta_j^2 \leq C\Delta t^2 K^{-4\beta}.$$

Now we finally observe that $\mathbb{E}(Z | \mathcal{F}_j) = 0$ for $Z \in \{\text{III}_j, \dots, \text{VI}_j\}$ and every $j = 0, \dots, n$, and thus

obtain with the estimates on $\mathbf{I}, \mathbf{II}_j, \dots, \mathbf{VI}_j$ that

$$\begin{aligned} \mathbb{E} \left(\|e_{m+1}\|_H^2 \right) &\leq C \mathbb{E} \left(\|\mathbf{I}\|_H^2 \right) + Cm \left(\sum_{j=1}^m \mathbb{E} \left(\|\mathbf{II}_j\|_H^2 \right) + \mathbb{E} \left(\|\bar{\Xi}_j\|_H^2 \right) \right) \\ &\quad + C \sum_{j=1}^m \left(\mathbb{E} \left(\|\mathbf{III}_j\|_H^2 \right) + \mathbb{E} \left(\|\mathbf{IV}_j\|_H^2 \right) + \mathbb{E} \left(\|\mathbf{V}_j\|_H^2 \right) + \mathbb{E} \left(\|\mathbf{VI}_j\|_H^2 \right) + \mathbb{E} \left(\|\bar{O}_j\|_H^2 \right) \right) \\ &\leq C \left(M^{-\alpha} + N^{-2\tilde{\alpha}} + \Delta t \left(\sum_{j=1}^m M^{-\alpha} + N_f^{-2\tilde{\alpha}} + K^{-2\beta} + \mathbb{E} \left(\|e_j\|_H^2 \right) \right) \right). \end{aligned}$$

The claim now follows by the discrete Grönwall inequality. \square

D Proof of Theorem 5.3 – Multilevel Monte Carlo Complexity

Proof of Theorem 5.3. Fix $\ell = 1, \dots, L$. By Theorem 4.1 and (28) there holds that

$$\max_{m=0, \dots, M} \mathbb{E} \left(\left\| \bar{Y}_m^\ell - Y_m^{c, \ell-1} \right\|_H^2 \right) \leq C \left(M_{\ell-1}^{\min(\alpha, 2)} + N_{\ell-1}^{-2\tilde{\alpha}} + K_{\ell-1}^{-2\beta} \right) = CM_{\ell-1}^{\min(\alpha, 2)}.$$

Now let $Y^{f, L}$ and $Y^{a, L}$ denote the fine approximation and its antithetic counterpart, respectively, on the finest level L . The bias of the MLMC estimator is then bounded due to Assumption 5.1 (ii) by

$$|\mathbb{E}(\Psi(X_T) - \bar{\Psi}_L)| \leq CM_L^{-1+\delta}.$$

Using a first order Taylor expansion of $\Psi \in C_b^2(H, \mathbb{R})$ around \bar{Y}_M^ℓ and Theorem 4.1 shows that for some $\xi^{f, \ell}, \xi^{a, \ell} \in H$ the variance decay on each level may be bounded by

$$\begin{aligned} \text{Var}(\bar{\Psi}_\ell - \Psi_{\ell-1}^c) &\leq \frac{1}{4} \mathbb{E} \left(\left(\Psi(Y_{M_\ell}^{f, \ell}) + \Psi(Y_{M_\ell}^{a, \ell}) - 2\Psi(Y_{M_{\ell-1}}^{c, \ell-1}) \right)^2 \right) \\ &\leq \mathbb{E} \left(\left(\Psi(\bar{Y}_{M_\ell}^\ell) - \Psi(Y_{M_{\ell-1}}^{c, \ell-1}) + \frac{\Psi'(\xi^{f, \ell})}{2} (Y_{M_\ell}^{f, \ell} - \bar{Y}_{M_\ell}^\ell) + \frac{\Psi'(\xi^{a, \ell})}{2} (Y_{M_\ell}^{a, \ell} - \bar{Y}_{M_\ell}^\ell) \right)^2 \right) \\ &\leq C \mathbb{E} \left(\left\| \bar{Y}_{M_\ell}^\ell - Y_{M_{\ell-1}}^{c, \ell-1} \right\|_H^2 \right) \\ &\leq CM_\ell^{-\min(\alpha, 2)}. \end{aligned}$$

Finally, Assumption 5.1 (i) yields a cost per sample of $\bar{\Psi}^\ell$ given by $\mathcal{C}_\ell \leq CM_\ell^{1+\gamma}$. As $M_\ell = M_0 2^\ell$, [15, Theorem 2.1] yields the claim. \square

References

- [1] R. Anton, D. Cohen, S. Larsson, and X. Wang. Full discretization of semilinear stochastic wave equations driven by multiplicative noise. *SIAM Journal on Numerical Analysis*, 54(2):1093–1119, 2016.
- [2] A. Barth. A finite element method for martingale-driven stochastic partial differential equations. *Communications on Stochastic Analysis*, 4(3):355–375, 2010.

- [3] A. Barth and F. E. Benth. The forward dynamics in energy markets – infinite-dimensional modelling and simulation. *Stochastics: An International Journal of Probability and Stochastic Processes*, 86(6):932–966, 2014.
- [4] A. Barth and A. Lang. Milstein approximation for advection-diffusion equations driven by multiplicative noncontinuous martingale noises. *Applied Mathematics and Optimization*, 66(3):387–413, 2012.
- [5] A. Barth and A. Lang. Multilevel Monte Carlo method with applications to stochastic partial differential equations. *International Journal of Computer Mathematics*, 89(18):2479–2498, 2012.
- [6] A. Barth, A. Lang, and C. Schwab. Multilevel Monte Carlo method for parabolic stochastic partial differential equations. *BIT Numerical Mathematics*, 53(1):3–27, 2013.
- [7] A. Barth and A. Stein. Approximation and simulation of infinite-dimensional Lévy processes. *Stochastics and Partial Differential Equations: Analysis and Computations*, 6(2):286–334, 2018.
- [8] A. Barth and A. Stein. Stochastic transport with Lévy noise – Fully discrete numerical approximation. *arXiv preprint arXiv:1910.14657*, 2019.
- [9] A. Barth and T. Stüwe. Weak convergence of Galerkin approximations of stochastic partial differential equations driven by additive Lévy noise. *Mathematics and Computers in Simulation*, 143:215–225, 2018.
- [10] F. E. Benth and S. Koekebakker. Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3):1116–1157, 2008.
- [11] A. Bonito and J. Pasciak. Numerical approximation of fractional powers of elliptic operators. *Mathematics of Computation*, 84(295):2083–2110, 2015.
- [12] R. Carmona and M. R. Tehranchi. *Interest Rate Models: An Infinite Dimensional Stochastic Analysis Perspective*. Springer Science & Business Media, 2007.
- [13] A. Debussche and J. Printems. Weak order for the discretization of the stochastic heat equation. *Mathematics of Computation*, 78(266):845–863, 2009.
- [14] T. Dunst, E. Hausenblas, and A. Prohl. Approximate Euler method for parabolic stochastic partial differential equations driven by space-time Lévy noise. *SIAM Journal on Numerical Analysis*, 50(6):2873–2896, 2012.
- [15] M. B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015.
- [16] M. B. Giles and C. Reisinger. Stochastic finite differences and multilevel Monte Carlo for a class of spdes in finance. *SIAM Journal on Financial Mathematics*, 3(1):572–592, 2012.
- [17] M. B. Giles and L. Szpruch. Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. *The Annals of Applied Probability*, 24(4):1585–1620, 2014.
- [18] D. S. Grebenkov and B.-T. Nguyen. Geometrical structure of Laplacian eigenfunctions. *SIAM Review*, 55(4):601–667, 2013.
- [19] I. Gyöngy, S. Sabanis, and D. Šiška. Convergence of tamed Euler schemes for a class of stochastic evolution equations. *Stochastics and Partial Differential Equations: Analysis and Computations*, 4(2):225–245, 2016.
- [20] E. Hausenblas and J. Seidler. Stochastic convolutions driven by martingales: maximal inequalities and exponential integrability. *Stochastic Analysis and Applications*, 26(1):98–119, 2007.

- [21] A. Jentzen and P. E. Kloeden. The numerical approximation of stochastic partial differential equations. *Milan Journal of Mathematics*, 77(1):205–244, 2009.
- [22] P. E. Kloeden, G. J. Lord, A. Neuenkirch, and T. Shardlow. The exponential integrator scheme for stochastic partial differential equations: Pathwise error bounds. *Journal of Computational and Applied Mathematics*, 235(5):1245–1260, 2011.
- [23] M. Kovács, S. Larsson, and F. Lindgren. On the backward Euler approximation of the stochastic Allen-Cahn equation. *Journal of Applied Probability*, 52(2):323–338, 2015.
- [24] M. Kovács, F. Lindner, and R. L. Schilling. Weak convergence of finite element approximations of linear stochastic evolution equations with additive Lévy noise. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1159–1199, 2015.
- [25] R. Kruse. *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*. Springer, 2014.
- [26] A. Lang. A note on the importance of weak convergence rates for spde approximations in multilevel Monte Carlo schemes. In *Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Leuven, Belgium, April 2014*, pages 489–505. Springer, 2016.
- [27] D. H. Nguyen, N. N. Nguyen, and G. Yin. Analysis of a spatially inhomogeneous stochastic partial differential equation epidemic model. *Journal of Applied Probability*, 57(2):613–636, 2020.
- [28] C. Pang and X. Wang. Antithetic multilevel Monte Carlo method for approximations of SDEs with non-globally lipschitz continuous coefficients. *arXiv preprint arXiv:2305.12992*, 2023.
- [29] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*, volume 44. Springer Science & Business Media, Applied Mathematical Sciences, 2 edition, 1983.
- [30] S. Peszat and J. Zabczyk. *Stochastic Partial Differential Equations with Lévy Noise*. Cambridge University Press, 2007.
- [31] C. Prévôt and M. Röckner. *A Concise Course on Stochastic Partial Differential Equations*. Springer, 2007.
- [32] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*, volume 25. Springer Science & Business Media, 2007.