BIT Numer Math (2013) 53:3–27
DOI 10.1007/s10543-012-0401-5

Multilevel Monte Carlo method for parabolic stochastic partial differential equations

Andrea Barth · Annika Lang · Christoph Schwab

Received: 18 May 2011 / Accepted: 9 August 2012 / Published online: 18 September 2012

Abstract We analyze the convergence and complexity of multilevel Monte Carlo discretizations of a class of abstract stochastic, parabolic equations driven by square integrable martingales. We show under low regularity assumptions on the solution that the judicious combination of low order Galerkin discretizations in space and an Euler–Maruyama discretization in time yields mean square convergence of order one in space and of order $1/2$ in time to the expected value of the mild solution. The complexity of the multilevel estimator is shown to scale log-linearly with respect to the corresponding work to generate a single path of the solution on the finest mesh, resp. of the corresponding deterministic parabolic problem on the finest mesh.

Keywords Multilevel Monte Carlo · Stochastic partial differential equations · Stochastic Finite Element Methods · Stochastic parabolic equation · Multilevel approximations

Mathematics Subject Classification (2010) 60H15 · 60H35 · 65C30 · 41A25 · 65C05 · 65N30

1 Introduction

Stochastic partial differential equations are increasingly used as models in engineering and the sciences. We mention only the pricing of energy derivative contracts,

---

Communicated by Desmond Higham.

A. Barth · A. Lang (✉) · C. Schwab
Seminar für Angewandte Mathematik, ETH, Rämistrasse 101, 8092 Zürich, Switzerland
e-mail: annika.lang@math.ethz.ch

A. Barth
e-mail: andrea.barth@math.ethz.ch

C. Schwab
e-mail: schwab@math.ethz.ch

porous media flows, filtering, and interest rate models. Accordingly, their efficient numerical solution has received more and more attention in recent years. Most of the numerical analysis of solution methods for stochastic partial differential equations has been devoted to identifying sufficient smoothness conditions on the data (among other parameters, covariance spectrum and smoothness of initial data) for certain discretization schemes to achieve a certain order of strong resp. weak or even pathwise convergence. See, e.g., the survey [15] and the references therein, and also [2–4, 11, 20] for recent results.

Except for results on pathwise convergence as, e.g., in [3, 4, 7], motivated by applications in numerical stochastic optimal control, commonly, moments of the solution such as first and second moments of the random solution are of main interest in applications. Such moments are approximated by Monte Carlo simulation, i.e., by averaging possibly large ensembles of approximated (in physical space and time) solution paths. We thus distinguish three principal sources of discretization errors in such moment approximations: spatial discretization, e.g., by Finite Elements or Finite Differences, time stepping errors due to e.g., Euler–Maruyama or Milstein time stepping and, finally, the sampling error incurred by replacing the mathematical expectations with finite ensemble averages. The convergence rate $1/2$ in mean square of Monte Carlo estimates (which is not improvable as can be seen in the proof of Lemma 4.1) combined with the high cost of approximating sample paths (which is due to the low spatial and temporal regularity of partial differential equations driven by noise) results in costly approximations of moments of solutions of stochastic partial differential equations.

This effect is, to a lesser extent, already present in the context of Itô stochastic (ordinary) differential equations. Recently, it was observed in [12, 13] that substantial efficiency gains in numerical simulation can be achieved by the use of so-called *multilevel path simulation techniques* in connection with Monte Carlo sampling. However, in [14] the authors show that multilevel Monte Carlo does not converge for equations with superlinearly growing coefficients if an inappropriate building block is used.

Analogous multilevel ideas have been proven successfully in the context of partial differential equations with random input data. We mention only [6] for elliptic partial differential equations with random coefficients, and [23] for Finite Volume solvers of scalar hyperbolic conservation laws with random initial data.

The analysis of a multilevel Monte Carlo discretization technique for parabolic stochastic partial differential equations driven by square integrable martingales is the purpose of the present paper. The multilevel Monte Carlo approach uses hierarchic meshes for the space and time approximation. A combination of a low number of Monte Carlo samples on very fine grids and increasing sample sizes on coarser meshes guarantees an optimal balance between the computational effort for sampling on one hand and solving the corresponding partial differential equation on the other. Using low-order Euler–Maruyama time stepping and space discretizations of low regularity, we give a-priori estimates on the mentioned three error contributions. We bound the strong error in mean square in Theorem 4.1 for the singlelevel discretization, i.e., the difference of the mild and the approximate solution for a fixed discretization mesh. This bound is the basis of the multilevel Monte Carlo error bound in Theorem 4.3 which explicitly contains error bounds in terms of the discretization parameters time step, space step, and the Monte Carlo sample size. Importantly, in the error analysis close to minimal assumptions on the spectrum of the covariance operator of the driving noise and on the initial data are imposed. The resulting multilevel Monte Carlo error bound in Theorem 4.3 is used to minimize the number of Monte Carlo samples at each discretization level in order to balance the statistical sampling with the spatial and temporal discretization errors. In Theorem 4.4, the resulting number of samples and a corresponding estimate of the asymptotic total work are given. The work estimate is shown to be superior to the corresponding work estimate for the singlelevel Monte Carlo method.

In the concluding remarks, we state, based on [4], multilevel Monte Carlo results for the Milstein scheme, in connection with a low order spatial Finite Element discretization. This discretization is shown to yield twice the (strong) convergence rate of the method based on the Euler–Maruyama time stepping scheme, however under stronger assumptions on the smoothness of the initial data and on the decay of the spectrum of the covariance operator. The approach towards the analysis of discretization schemes can be generalized to higher order discretizations. Since more sophisticated spatial and temporal discretizations exhibit higher convergence rates only under stronger regularity assumptions on the data, and since multilevel Monte Carlo variants of such discretizations are dominated by the sampling, the overall computational work for such higher order schemes is dominated once more by the work for sample path generation.

2 Preliminaries

We consider stochastic processes with values in a separable Hilbert space $(U, (\cdot,\cdot)_U)$ defined on a filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the "usual conditions". The space of all càdlàg, square integrable martingales taking values in U with respect to $(\mathcal{F}_t)_{t\geq 0}$ is denoted by $\mathcal{M}^2(U)$. We restrict ourselves to the following class of martingales

$$\begin{aligned}\mathcal{M}_b^2(U) = \big\{ M \in \mathcal{M}^2(U), \exists Q \in L_1^+(U) \text{ s.t. } \forall t \geq s \geq 0, \\ \big\langle\langle M, M\rangle\big\rangle_t - \big\langle\langle M, M\rangle\big\rangle_s \leq (t-s)Q \big\},\end{aligned}$$

where $L_1^+(U)$ denotes the space of all nuclear, symmetric, nonnegative definite operators. The operator angle bracket process $(\langle\langle M, M\rangle\rangle_t, t \geq 0)$ is defined as

$$\langle\langle M, M\rangle\rangle_t = \int_0^t Q_s \, d\langle M, M\rangle_s$$

for $t \geq 0$, where $(\langle M, M\rangle_t, t \geq 0)$ is the unique angle bracket process from the Doob–Meyer decomposition. The process $(Q_s, s \geq 0)$ is often referred to as the *martingale covariance* of the process M. Examples of such processes are Q-Wiener processes and square integrable Lévy martingales, i.e., those Lévy martingales with Lévy measure ν that satisfies

$$\int_U \|\varphi\|_U^2 \nu(d\varphi) < +\infty.$$

Since $Q \in L_1^+(U)$, there exists an orthonormal basis $(e_n, n \in \mathbb{N})$ of U consisting of eigenvectors of Q. Therefore, for $n \in \mathbb{N}$, we have the representation $Qe_n = \gamma_n e_n$, where $\gamma_n \geq 0$ is the eigenvalue corresponding to e_n. Then the square root of Q is defined as

$$Q^{1/2}\psi = \sum_{n\in\mathbb{N}} (\psi, e_n)_U \gamma_n^{1/2} e_n$$

for $\psi \in U$, and $Q^{-1/2}$ denotes the pseudo inverse of $Q^{1/2}$. Let us denote by $(\mathcal{H}, (\cdot,\cdot)_{\mathcal{H}})$ the Hilbert space defined by $\mathcal{H} = Q^{1/2}(U)$ endowed with the inner product $(\psi, \phi)_{\mathcal{H}} = (Q^{-1/2}\psi, Q^{-1/2}\phi)_U$ for $\psi, \phi \in \mathcal{H}$. We refer to the space of all Hilbert–Schmidt operators from $\mathcal{H}$ to a separable Hilbert space $(H, (\cdot,\cdot)_H)$ as $L_{HS}(\mathcal{H}; H)$, and by $\|\cdot\|_{L_{HS}(\mathcal{H};H)}$ we denote the corresponding norm. By Proposition 8.16 in [25], we have

$$\mathbb{E}\left[\left\|\int_0^t \Psi(s)\, dM(s)\right\|_H^2\right] \leq \mathbb{E}\left[\int_0^t \|\Psi(s)\|_{L_{HS}(\mathcal{H};H)}^2\, ds\right] \tag{2.1}$$

for $t \in [0, T]$, $M \in \mathcal{M}_b^2(U)$, and a locally bounded, predictable process $\Psi : [0, T] \to L_{HS}(\mathcal{H}, H)$ with

$$\mathbb{E}\left[\int_0^T \|\Psi(s)\|_{L_{HS}(\mathcal{H};H)}^2\, ds\right] < +\infty.$$

For a separable Hilbert space $(H, (\cdot,\cdot)_H)$ with induced norm $\|\cdot\|_H$, we denote the set of strongly measurable, square summable mappings $Y : \Omega \to H$ by

$$L^2(\Omega; H) = \left\{Y : \Omega \to H, Y \text{ strongly measurable, } \|Y\|_{L^2(\Omega;H)} < +\infty\right\},$$

where

$$\|Y\|_{L^2(\Omega;H)} = \mathbb{E}\left[\|Y\|_H^2\right]^{1/2}.$$

On H we consider the stochastic partial differential equation

$$dX(t) = \big(AX(t) + F\big(X(t)\big)\big)\, dt + G\big(X(t)\big)\, dM(t) \tag{2.2}$$

for $t \in \mathbb{T} = [0, T]$, $T < +\infty$, subject to the initial condition $X(0) = X_0 \in L^2(\Omega; H)$, which is $\mathcal{F}_0$-measurable. The operator A with densely defined domain $\mathcal{D}(A) \subset H$ is assumed to be the generator of an analytic semigroup S on H and zero is in the resolvent set of $-A$. We further assume that A is boundedly invertible on $\mathcal{D}(A)$, and that $(-A)^{-1} : H \to \mathcal{D}(A)$ is a bounded linear operator. Then, for $0 < \alpha < 1$, the interpolation operators $A_\alpha = (-A)^\alpha$ of index α between the linear operator $-A$ and the identity operator I on H are well-defined (see e.g., Theorem 6.13 in [24]). We set $V = \mathcal{D}(A_{1/2})$. Let us define the continuous bilinear form $B_A : V \times V \to \mathbb{R}$ by

$$B_A(\varphi, \psi) = (A_{1/2}\varphi, A_{1/2}\psi)_H$$

for $\varphi, \psi \in V$. We set

$$\|\varphi\|_V = \|A_{1/2}\varphi\|_H$$

for $\varphi \in V$ and define the norm on $L^2(\Omega; V)$ accordingly. Furthermore, by Theorem 6.13 in [24], there exists a constant $C > 0$ such that for all $t \in \mathbb{T}$ and $\varphi \in V$

$$\big\| \big(S(t) - I\big)\varphi \big\|_H \leq C t^{1/2} \|A_{1/2}\varphi\|_H = C t^{1/2} \|\varphi\|_V, \tag{2.3}$$

and since S is strongly continuous, there exists a constant $C \geq 1$ such that

$$\big\| S(t)\varphi \big\|_H \leq C \|\varphi\|_H \tag{2.4}$$

for all $t \in \mathbb{T}$ and $\varphi \in H$. This implies that for $0 \leq s < t \leq T$ and $\varphi \in V$

$$\big\| \big(S(t) - S(s)\big)\varphi \big\|_H \leq C \sqrt{t-s} \|\varphi\|_V. \tag{2.5}$$

Following up on the properties of Eq. (2.2), we consider the operator F as a mapping from H to H and G as a mapping from H to the linear operators from U into H. We assume that the stochastic process M is in $\mathcal{M}_b^2(U)$.

Next, we introduce a diffusion problem on a bounded domain of $\mathbb{R}^d$, $d \in \mathbb{N}$, as an example of our abstract framework.

Example 2.1 Let $D \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ be a convex polygon and $H = L^2(D)$. The operator A with domain $\mathcal{D}(A) \subset L^2(D)$ is the unique self-adjoint extension of the differential operator

$$\sum_{i,j=1}^{d} \partial_i (a_{ij} \partial_j u)$$

for $u \in C_c^2(D)$ such that u satisfies Dirichlet boundary conditions, i.e., $u(x) = 0$, for $x \in \Gamma$ on the boundary $\Gamma = \partial D$. Here, $C_c^2(D)$ denotes the twice continuously differentiable functions on D with compact support. The functions a_{ij} are supposed to be continuously differentiable on D with continuous extension to the closure $\bar{D}$. We assume that there exists $\delta > 0$ such that for all $x \in D$ and $\xi \in \mathbb{R}^d$

$$\sum_{i,j=1}^{d} a_{ij}(x) \xi_i \xi_j \geq \delta \|\xi\|_{\mathbb{R}^d}^2.$$

This implies that the operator A is dissipative, see e.g., [18]. Then by the Lumer–Phillips theorem, see e.g., [10], A generates a strongly continuous contraction semigroup S on H. Furthermore, by Corollary 2 in [17], S is analytic in the right half-plane. Therefore, fractional powers of $-A$ are well defined, cf. [10] and Eq. (2.3) holds. By results in [1], $V = H_0^1(D)$ and the norm $\|A_{1/2} \cdot\|_{L^2(D)}$ is equivalent to $\|\cdot\|_{H^1}$ which is known as Kato's conjecture.

The bilinear form corresponding to A is given by

$$B_A(\varphi, \psi) = \sum_{i,j=1}^{d} (a_{ij} \partial_j \varphi, \partial_i \psi)_H$$

for $\varphi, \psi \in V$.

An example of a nonlinear F is

$$F(\varphi)(x) = \frac{\varphi(x)^3}{1+\varphi(x)^2}$$

for $\varphi \in V$ and $x \in D$. This example also satisfies the conditions specified in Assumption 2.1. To give an explicit choice for G, we first fix U and Q. Therefore, let $U = H$ and $(e_n, n \in \mathbb{N})$ be an eigenbasis of Q with corresponding eigenvalues $(\gamma_n, n \in \mathbb{N})$ such that $\sum_n \gamma_n \|e_n\|^2_{W^{1,\infty}(D)} \le C$ for some finite constant C, where $W^{1,\infty}(D)$ denotes the Sobolev space of order 1 with (weak) derivatives in $L^\infty(D)$. Then one possible choice of the operator G is

$$G(\psi)\varphi(x) = g(x)\big(1+\psi(x)\big)\varphi(x)$$

for some boundedly differentiable function $g \in C^1_0(D)$ and $\psi, \varphi \in H$.

Next, we make assumptions such that Eq. (2.2) has a mild solution. Therefore, we impose linear growth and Lipschitz conditions on the operators $F : H \to H$ and $G : H \to L(U; H)$:

Assumption 2.1 Let $B = H, V$. Assume that there exist constants $C_1, C_2 > 0$ such that for all $\phi \in B$, $\varphi_1, \varphi_2 \in H$ it holds that

$$\big\|F(\phi)\big\|_B \le C_1\big(1+\|\phi\|_B\big),$$
$$\big\|F(\varphi_1) - F(\varphi_2)\big\|_H \le C_1\|\varphi_1 - \varphi_2\|_H,$$

and

$$\big\|G(\phi)\big\|_{L_{HS}(\mathcal{H};B)} \le C_2\big(1+\|\phi\|_B\big),$$
$$\big\|G(\varphi_1) - G(\varphi_2)\big\|_{L_{HS}(\mathcal{H};H)} \le C_2\|\varphi_1 - \varphi_2\|_H.$$

We note that under Assumption 2.1 in the special case of a stochastic partial differential equation with additive Wiener noise, i.e., the last part of the equation reads $G\,dW(t)$, we can assume without loss of generality that $U = H$ since GW defines a Wiener process with covariance operator GQG^*.

Assumption 2.1 implies that Eq. (2.2) has a unique mild solution in H by results in Chap. 9 in [25] and that the predictable process $X : \mathbb{T} \times \Omega \to H$ is given by

$$X(t) = S(t)X_0 + \int_0^t S(t-s)F\big(X(s)\big)\,ds + \int_0^t S(t-s)G\big(X(s)\big)\,dM(s) \quad (2.6)$$

for $t \in \mathbb{T}$. For further discussions on stochastic differential equations in infinite dimensions, the reader is referred to [9] and [25] and the references therein.

A certain regularity on the initial condition causes the regularity of the mild solution $X = (X(t), t \in \mathbb{T})$, which is specified in the following lemma.

Lemma 2.1 *If Assumption 2.1 holds and* $\|X_0\|_{L^2(\Omega;V)} < +\infty$, *then the solution* X *defined in Eq.* (2.6) *is in* $L^2(\Omega; V)$. *In particular, for all* $t \in \mathbb{T}$ *it holds*

$$\big\|X(t)\big\|_{L^2(\Omega;V)} \le C(T)\big(1 + \|X_0\|_{L^2(\Omega;V)}\big).$$

Proof With Assumption 2.1, Eq. (2.1), Eq. (2.4), Hölder's inequality, and Gronwall's inequality, we have for Eq. (2.6)

$$\begin{aligned}
&\big\|X(t)\big\|^2_{L^2(\Omega;V)}\\
&\quad = \left\|S(t)X_0 + \int_0^t S(t-s)F\big(X(s)\big)\,ds + \int_0^t S(t-s)G\big(X(s)\big)\,dM(s)\right\|^2_{L^2(\Omega;V)}\\
&\quad \le 3\bigg(C\|X_0\|^2_{L^2(\Omega;V)} + C(T)\int_0^t \big\|S(t-s)F\big(X(s)\big)\big\|^2_{L^2(\Omega;V)}\,ds\\
&\qquad\qquad + \int_0^t \big\|S(t-s)G\big(X(s)\big)\big\|^2_{L_{HS}(\mathcal{H};V)} ds\bigg)\\
&\quad \le C\|X_0\|_{L^2(\Omega;V)} + 2\cdot C(T)\bigg(1 + \int_0^t \big\|X(s)\big\|^2_{L^2(\Omega;V)}\,ds\bigg)\\
&\quad \le C(T)\big(1 + \|X_0\|^2_{L^2(\Omega;V)}\big)\\
&\quad < +\infty,
\end{aligned}$$

where $C(T)$ denotes a varying constant that depends on T. □

For later proofs we need a lemma on the regularity of the mild solution X in time. It is mainly based on Eq. (2.5). Related results can be found in [2] and [22]. We include the proof for completeness of exposition.

Lemma 2.2 *If Assumption 2.1 holds and* $\|X_0\|_{L^2(\Omega;V)} < +\infty$, *then there exists a constant* $C(T)$ *such that the mild solution* X *in Eq.* (2.6) *satisfies*

$$\big\|X(t) - X(s)\big\|_{L^2(\Omega;H)} \le C(T)\sqrt{t-s}\big(1 + \|X_0\|_{L^2(\Omega;V)}\big)$$

for $0 \le s < t \le T$.

Proof The regularity is provided by

$$\begin{aligned}
&\big\|X(t) - X(s)\big\|_{L^2(\Omega;H)}\\
&\quad \le \big\|\big(S(t) - S(s)\big)X_0\big\|_{L^2(\Omega;H)}\\
&\qquad + \left\|\int_0^s \big(S(t-r) - S(s-r)\big)F\big(X(r)\big)\,dr\right\|_{L^2(\Omega;H)}\\
&\qquad + \left\|\int_0^s \big(S(t-r) - S(s-r)\big)G\big(X(r)\big)\,dM(r)\right\|_{L^2(\Omega;H)}
\end{aligned}$$

$$+\left\| \int_s^t S(t-r)F\big(X(r)\big)\,dr \right\|_{L^2(\Omega;H)}$$
$$+\left\| \int_s^t S(t-r)G\big(X(r)\big)\,dM(r) \right\|_{L^2(\Omega;H)}.$$

Equation (2.5) implies for the first term that

$$\big\|\big(S(t)-S(s)\big)X_0\big\|_{L^2(\Omega;H)} \le C\sqrt{t-s}\,\|X_0\|_{L^2(\Omega;V)}.$$

To bound the second term, we use Eq. (2.5), Assumption 2.1, and Lemma 2.1:

$$\begin{aligned}
&\left\| \int_0^s \big(S(t-r)-S(s-r)\big)F\big(X(r)\big)\,dr \right\|_{L^2(\Omega;H)} \\
&\quad\le C\sqrt{t-s}\int_0^s \big\|F\big(X(r)\big)\big\|_{L^2(\Omega;V)}\,dr \\
&\quad\le C\sqrt{t-s}\left(1+\int_0^s \|X(r)\|_{L^2(\Omega;V)}\,dr\right) \\
&\quad\le C(T)\sqrt{t-s}\big(1+\|X_0\|_{L^2(\Omega;V)}\big).
\end{aligned}$$

Similarly, we have for the third term using Eq. (2.1) additionally

$$\begin{aligned}
&\left\| \int_0^s \big(S(t-r)-S(s-r)\big)G\big(X(r)\big)\,dM(r) \right\|_{L^2(\Omega;H)} \\
&\quad\le C\sqrt{t-s}\,\mathbb{E}\left[\int_0^s \big\|G\big(X(r)\big)\big\|^2_{L_{HS}(\mathcal{H};V)}dr\right]^{1/2} \\
&\quad\le C(T)\sqrt{t-s}\big(1+\|X_0\|_{L^2(\Omega;V)}\big).
\end{aligned}$$

For the fourth term, it holds with Eq. (2.4), Assumption 2.1, and Lemma 2.1 that

$$\begin{aligned}
\left\| \int_s^t S(t-r)F\big(X(r)\big)\,dr \right\|_{L^2(\Omega;H)} &\le C\int_s^t \big\|F\big(X(r)\big)\big\|_{L^2(\Omega;H)}\,dr \\
&\le C(T)(t-s)\big(1+\|X_0\|_{L^2(\Omega;H)}\big).
\end{aligned}$$

The last term is bounded by

$$\begin{aligned}
\left\| \int_s^t S(t-r)G\big(X(r)\big)\,dM(r) \right\|_{L^2(\Omega;H)} &\le \mathbb{E}\left[\int_s^t \big\|S(t-r)G\big(X(r)\big)\big\|^2_{L_{HS}(\mathcal{H};H)}dr\right]^{1/2} \\
&\le C(T)\sqrt{t-s}\big(1+\|X_0\|_{L^2(\Omega;H)}\big)
\end{aligned}$$

with similar calculations as in the fourth term and the application of Eq. (2.1). This concludes the proof, since

$$\begin{aligned}\big\|X(t)-X(s)\big\|_{L^2(\Omega;H)} &\le C(T)\sqrt{t-s}\|X_0\|_{L^2(\Omega;V)}\\&\quad+2\cdot C\sqrt{t-s}\big(1+\|X_0\|_{L^2(\Omega;V)}\big)\\&\quad+C(T)(t-s)\big(1+\|X_0\|_{L^2(\Omega;H)}\big)\\&\quad+C(T)\sqrt{t-s}\big(1+\|X_0\|_{L^2(\Omega;H)}\big)\\&\le C(T)\sqrt{t-s}\big(1+\|X_0\|_{L^2(\Omega;V)}\big).\end{aligned}$$ □

In the next section, we introduce an approximation scheme for Eq. (2.6). We present a discretization in time and space, which we combine in the subsequent section with a multilevel Monte Carlo estimator.

3 Approximation scheme

Let $\mathcal{V}=(V_\ell,\ell\in\mathbb{N}_0)$ be a nested family of finite dimensional subspaces of V with refinement level $\ell>0$, refinement sizes $(h_\ell,\ell\in\mathbb{N}_0)$, associated H-orthogonal projections $(P_\ell,\ell\in\mathbb{N}_0)$, and norm induced by H. The sequence $\mathcal{V}$ is supposed to be dense in H in the sense that for all $\phi\in H$

$$\lim_{\ell\to+\infty}\|\phi-P_\ell\phi\|_H=0.$$

We define the approximate operator $A_\ell:V_\ell\to V_\ell$ through the bilinear form

$$(-A_\ell\varphi_\ell,\psi_\ell)_H=B_A(\varphi_\ell,\psi_\ell)$$

for all $\varphi_\ell,\psi_\ell\in V_\ell$. The operator A_ℓ is the generator of an analytic semigroup $S_\ell=(S_\ell(t),t\ge 0)$ defined formally by $S_\ell(t)=\exp(tA_\ell)$ for $t\ge 0$. Then the semidiscrete problem is given by

$$d\tilde{X}_\ell(t)=\big(A_\ell\tilde{X}_\ell(t)+P_\ell F\big(\tilde{X}_\ell(t)\big)\big)\,dt+P_\ell G\big(\tilde{X}_\ell(t)\big)\,dM(t)$$

for $t\in\mathbb{T}$ with initial condition $\tilde{X}_\ell(0)=P_\ell X_0$. The semidiscrete problem has a mild solution which reads

$$\tilde{X}_\ell(t)=S_\ell(t)\tilde{X}_\ell(0)+\int_0^t S_\ell(t-s)P_\ell F\big(\tilde{X}_\ell(s)\big)\,ds+\int_0^t S_\ell(t-s)P_\ell G\big(\tilde{X}_\ell(s)\big)\,dM(s)$$

for $t\in\mathbb{T}$. We shall remark here that we do not approximate the noise. For a stochastic partial differential equation with additive noise M with covariance $Q\in L_1^+(U)$, where $G(\varphi)=I$ for all $\varphi\in H$ and $U=H$, the noise is automatically finite dimensional if V_ℓ contains a finite subset of the eigenbasis of Q since P_ℓ cuts off the Karhunen–Loève expansion of M (see e.g., [19]). Otherwise, this approximation might not be suitable for simulations. In this case it is possible to truncate—if existent—the series representation of M depending on the level ℓ. For example for Lévy processes it is shown in [3] which properties especially of the eigenvalues of M

and the chosen number of eigenbasis elements imply that the overall order of convergence is preserved.

Next, we introduce a fully discrete approximation. Therefore, let $(\Theta^n, n \in \mathbb{N}_0)$ be a sequence of equidistant time discretizations with step sizes $\delta t^n = T\, 2^{-n}$, i.e., for $n \in \mathbb{N}_0$

$$\Theta^n = \{t_k^n = T2^{-n}k = \delta t^n k, k = 0, \ldots, 2^n\}.$$

For $t_k^n \in \Theta^n$, we approximate the semigroup $S_\ell(t_k^n)$ by a rational approximation $r(\delta t^n A_\ell)^k$ that satisfies the following assumption, which can for example be realized by a backward Euler scheme.

Assumption 3.1 The rational approximation r of the semigroup is stable and there exists a constant $C > 0$ such that r satisfies for $\ell, n, k \in \mathbb{N}_0$, $k \le 2^n$, and $\varphi \in V$ the error bound

$$\big\|\big(S(t_k^n) - r\big(\delta t^n A_\ell\big)^k P_\ell\big)\varphi\big\|_H \le C\big(h_\ell + \sqrt{\delta t^n}\big)\|\varphi\|_V.$$

We consider the fully discrete approximation of Euler–Maruyama type

$$\begin{aligned} X_{\ell,n}\big(t_k^n\big) &= r\big(\delta t^n A_\ell\big) X_{\ell,n}\big(t_{k-1}^n\big) + r\big(\delta t^n A_\ell\big) P_\ell F\big(X_{\ell,n}\big(t_{k-1}^n\big)\big)\delta t^n \\ &\quad + r\big(\delta t^n A_\ell\big) P_\ell G\big(X_{\ell,n}\big(t_{k-1}^n\big)\big)\big(M\big(t_k^n\big) - M\big(t_{k-1}^n\big)\big) \end{aligned}$$

for $\ell, n \in \mathbb{N}_0$, $0 < k \le 2^n$, which may be rewritten as

$$\begin{aligned} X_{\ell,n}\big(t_k^n\big) &= r\big(\delta t^n A_\ell\big)^k P_\ell X_0 + \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} r\big(\delta t^n A_\ell\big)^{k-j+1} P_\ell F\big(X_{\ell,n}\big(t_{j-1}^n\big)\big)\, ds \\ &\quad + \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} r\big(\delta t^n A_\ell\big)^{k-j+1} P_\ell G\big(X_{\ell,n}\big(t_{j-1}^n\big)\big)\, dM(s). \end{aligned} \tag{3.1}$$

In the following, we give an approximation for Example 2.1 that meets Assumption 3.1.

Example 3.1 (Diffusion problem on a bounded domain $D \subset \mathbb{R}^d$) Here, we introduce an approximation of Example 2.1 according to [26]. In the physical domain $D \subset \mathbb{R}^d$ with $d \in \mathbb{N}$, we approximate the mild solution X, given in Eq. (2.6), with a Finite Element discretization in D. The Finite Element method which we consider is based on nested sequences of simplicial triangulations $\mathcal{T}_\ell$, $\ell \in \mathbb{N}_0$ of the polygonal domain D. For any $\ell \ge 0$, we denote the mesh width of $\mathcal{T}_\ell$ by

$$h_\ell = \max_{K \in \mathcal{T}_\ell} \{\operatorname{diam}(K)\}.$$

The uniform refinement of the mesh is achieved by regular subdivision of the initial mesh $\mathcal{T}_0$ with maximal diameter h_0. This results in the mesh width $h_\ell = 2^{-\ell} h_0$ for $\ell \in \mathbb{N}$, since $h_{\ell+1} = 2^{-1} h_\ell$. On $\mathcal{T}_\ell$, we define the Finite Element spaces

$$V_\ell = \mathcal{S}_0^1(D, \mathcal{T}_\ell) = \{v \in H_0^1(D), v|_K \in \mathcal{P}_1, K \in \mathcal{T}_\ell\},$$

where $\mathcal{P}_1 = \text{span}\{x^\alpha, |\alpha| \le 1\}$ denotes the space of polynomials of total degree not exceeding 1. In this framework P_ℓ denotes the $L^2(D)$ projection onto V_ℓ. The bilinear form on V_ℓ and the corresponding approximate operator A_ℓ read

$$(-A_\ell \varphi_\ell, \psi_\ell)_H = B_A(\varphi_\ell, \psi_\ell) = \sum_{i,j=1}^{d} (a_{ij} \partial_j \varphi_\ell, \partial_i \psi_\ell)_H$$

for $\varphi_\ell, \psi_\ell \in V_\ell$. Assumption 3.1 is fulfilled by a rational approximation of the semigroup which is stable and accurate at least of order $q = 1/2$, see for example Theorem 7.1 in [26]. This means

$$r(\lambda) = \exp(-\lambda) + \mathrm{O}(\lambda^{q+1})$$

for $|\lambda| \to 0$ and $\sup_{\lambda \in \sigma(\delta t^n A_\ell)} |r(\lambda)| \le 1$ for $\ell \in \mathbb{N}_0$, where $\sigma(\delta t^n A_\ell)$ denotes the spectrum of $\delta t^n A_\ell$. The approximation of the semigroup by a backward Euler type time stepping is of order 1 (cf. [26]) and therefore meets the assumption.

4 Rate of strong convergence of a multilevel Monte Carlo approximation

In this section we derive a strong convergence result of a multilevel Monte Carlo approximation of the expectation of Eq. (3.1), i.e., a convergence result for the difference

$$\big\| \mathbb{E}\big[X\big(t_k^n\big)\big] - E^L\big[X_{L,n}\big(t_k^n\big)\big]\big\|_{L^2(\Omega;H)}, \tag{4.1}$$

where E^L is a multilevel estimator for the expectation which is introduced in the following. Therefore, let $(Y_\ell, \ell \in \mathbb{N}_0)$ be a sequence of V-valued random variables such that $Y_\ell \in V_\ell$ for all $\ell \in \mathbb{N}_0$. Then, for $L \in \mathbb{N}_0$, Y_L can be written as

$$Y_L = \sum_{\ell=0}^{L} (Y_\ell - Y_{\ell-1}),$$

where $Y_{-1} = 0$. By linearity of the expectation, it holds that

$$\mathbb{E}[Y_L] = \mathbb{E}\left[\sum_{\ell=0}^{L} (Y_\ell - Y_{\ell-1})\right] = \sum_{\ell=0}^{L} \mathbb{E}[Y_\ell - Y_{\ell-1}].$$

To derive a multilevel estimator for the expectation from this expression, we approximate $\mathbb{E}[Y_\ell - Y_{\ell-1}]$ by a Monte Carlo method with a level dependent number N_ℓ of samples, which implies that we may estimate $\mathbb{E}[Y_L]$ by

$$E^L[Y_L] = \sum_{\ell=0}^{L} E_{N_\ell}[Y_\ell - Y_{\ell-1}].$$

The terms in the sum on the right hand side are Monte Carlo estimators for the expectation of the difference of Y_ℓ and $Y_{\ell-1}$, defined by

$$E_N[Y] = \frac{1}{N}\sum_{i=1}^{N}\hat{Y}^i \tag{4.2}$$

for $N \in \mathbb{N}$, where $(\hat{Y}^i, i = 1, \dots, N)$ is a sequence of independent, identically distributed copies of a random variable Y.

We give a detailed analysis of the (singlelevel) Monte Carlo estimator next, before we prove a multilevel Monte Carlo error bound for Eq. (4.1). The strong error bound that we derive in the next section is needed for both estimates. Further, this enables us to compare the singlelevel Monte Carlo method with the multilevel approach.

4.1 Singlelevel Monte Carlo approximation

In this section we derive a result on the convergence of Monte Carlo estimators of random variables. Further, we prove a mean square convergence rate of the approximation in space and time of the stochastic partial differential equation (2.6). We combine both results to an error bound for the singlelevel Monte Carlo method. First, we consider the convergence of the Monte Carlo estimator (4.2), which cannot be improved for a given random variable with fixed variance.

Lemma 4.1 *For any $N \in \mathbb{N}$ and for $Y \in L^2(\Omega; H)$, it holds that*

$$\big\|\mathbb{E}[Y] - E_N[Y]\big\|_{L^2(\Omega;H)} = \frac{1}{\sqrt{N}}\operatorname{Var}[Y]^{1/2} \le \frac{1}{\sqrt{N}}\|Y\|_{L^2(\Omega;H)}.$$

Proof With the independence of the identically distributed samples it follows that

$$\begin{aligned}
\big\|\mathbb{E}[Y] - E_N[Y]\big\|^2_{L^2(\Omega;H)} &= \mathbb{E}\bigg[\bigg\|\mathbb{E}[Y] - \frac{1}{N}\sum_{i=1}^{N}\hat{Y}^i\bigg\|^2_H\bigg] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}\big[\big\|\mathbb{E}[Y] - \hat{Y}^i\big\|^2_H\big] \\
&= \frac{1}{N}\mathbb{E}\big[\big\|\mathbb{E}[Y] - Y\big\|^2_H\big] = \frac{1}{N}\big(\mathbb{E}\big[\|Y\|^2_H\big] - \big\|\mathbb{E}[Y]\big\|^2_H\big) \\
&\le \frac{1}{N}\|Y\|^2_{L^2(\Omega;H)},
\end{aligned}$$

where $\operatorname{Var}[Y] = \mathbb{E}[\|\mathbb{E}[Y] - Y\|^2_H]$. □

Remark 4.1 Lemma 4.1 is formulated for an arbitrary random variable $Y \in L^2(\Omega; H)$. In the subsequent proofs for $\ell, n \in \mathbb{N}_0$ and for $t \in \Theta^n$, we estimate the

Monte Carlo error of the discrete mild solution, which is bounded with Lemma 4.1 by

$$\left\|\mathbb{E}\left[X_{\ell,n}(t)\right] - E_N\left[X_{\ell,n}(t)\right]\right\|_{L^2(\Omega;H)} \leq \frac{1}{\sqrt{N}}\left\|X_{\ell,n}(t)\right\|_{L^2(\Omega;H)}.$$

Furthermore, for $t_k^n = t$ it holds that

$$\begin{aligned}
&\left\|X_{\ell,n}\left(t_k^n\right)\right\|^2_{L^2(\Omega;H)} \\
&\quad\leq 3\Bigg(\left\|r\left(\delta t^n A_\ell\right)^k P_\ell X_0\right\|^2_{L^2(\Omega;H)} \\
&\qquad + \left\|\sum_{j=1}^{k}\int_{t_{j-1}^n}^{t_j^n} r\left(\delta t^n A_\ell\right)^{k-j+1} P_\ell F\left(X_{\ell,n}\left(t_{j-1}^n\right)\right) ds\right\|^2_{L^2(\Omega;H)} \\
&\qquad + \left\|\sum_{j=1}^{k}\int_{t_{j-1}^n}^{t_j^n} r\left(\delta t^n A_\ell\right)^{k-j+1} P_\ell G\left(X_{\ell,n}\left(t_{j-1}^n\right)\right) dM(s)\right\|^2_{L^2(\Omega;H)}\Bigg) \\
&\quad\leq C\Bigg(\|X_0\|^2_{L^2(\Omega;H)} \\
&\qquad + \left(\sum_{j=1}^{k}\int_{t_{j-1}^n}^{t_j^n}\left(1 + \left\|X_{\ell,n}\left(t_{j-1}^n\right)\right\|_{L^2(\Omega;H)}\right) ds\right)^2 \\
&\qquad + \sum_{j=1}^{k}\int_{t_{j-1}^n}^{t_j^n}\left(1 + \left\|X_{\ell,n}\left(t_{j-1}^n\right)\right\|_{L^2(\Omega;H)}\right)^2 ds\Bigg),
\end{aligned}$$

where we used the stability of the rational approximation of the semigroup, Eq. (2.1), and Assumption 2.1. Hölder's inequality and a discrete Gronwall inequality (see [8]) lead to

$$\begin{aligned}
\left\|X_{\ell,n}\left(t_k^n\right)\right\|^2_{L^2(\Omega;H)} &\leq C(T)\left(\|X_0\|^2_{L^2(\Omega;H)} + 1 + \delta t^n\sum_{j=1}^{k}\left\|X_{\ell,n}\left(t_{j-1}^n\right)\right\|^2_{L^2(\Omega;H)}\right) \\
&\leq C(T)\left(1 + \|X_0\|^2_{L^2(\Omega;H)}\right).
\end{aligned}$$

This estimate implies

$$\sup_{t\in\Theta^n}\left\|\mathbb{E}\left[X_{\ell,n}(t)\right] - E_N\left[X_{\ell,n}(t)\right]\right\|_{L^2(\Omega;H)} \leq \frac{1}{\sqrt{N}}C(T)\left(1 + \|X_0\|_{L^2(\Omega;H)}\right).$$

The error bound in Lemma 4.1 is of limited practical value, since any implementation of the estimator $E_N[X(t)]$ for $t \in \mathbb{T}$ of the mild solution X of Eq. (2.6) requires an approximation of the 'samples' $\hat{X}(t)^i$, incurring an additional error. We therefore now derive an a-priori error bound which includes the additional discretization error stemming from the space discretization and from time stepping along sample paths. The considered discretization scheme is introduced in Eq. (3.1) for some level $\ell \in \mathbb{N}_0$.

A strong error bound, i.e., an error bound in $L^2(\Omega; H)$ for the approximation is given in the following theorem.

Theorem 4.1 *If X is the mild solution of* (2.6) *and* $(X_{\ell,n}, \ell, n \in \mathbb{N}_0)$ *is the sequence of discrete mild solutions introduced in Eq.* (3.1), *then there exists a constant $C(T)$ such that for all $\ell, n \in \mathbb{N}_0$*

$$\sup_{t\in\Theta^n} \big\| X(t) - X_{\ell,n}(t)\big\|_{L^2(\Omega;H)} \le C(T)\big(h_\ell + \sqrt{\delta t^n}\big)\big(1 + \|X_0\|_{L^2(\Omega;V)}\big).$$

Proof For $\ell, n \in \mathbb{N}_0$ and $t_k^n \in \Theta^n$, the error is bounded by

$$\begin{aligned}
&\big\| X\big(t_k^n\big) - X_{\ell,n}\big(t_k^n\big)\big\|^2_{L^2(\Omega;H)} \\
&\quad\le 3\Bigg(\big\| \big(S\big(t_k^n\big) - r\big(\delta t^n A_\ell\big)^k P_\ell\big)X_0\big\|^2_{L^2(\Omega;H)} \\
&\qquad + \Bigg\| \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} S\big(t_k^n - s\big) F\big(X(s)\big) \\
&\qquad\qquad - r\big(\delta t^n A_\ell\big)^{k-j+1} P_\ell F\big(X_{\ell,n}\big(t_{j-1}^n\big)\big)\, ds \Bigg\|^2_{L^2(\Omega;H)} \\
&\qquad + \Bigg\| \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} S\big(t_k^n - s\big) G\big(X(s)\big) \\
&\qquad\qquad - r\big(\delta t^n A_\ell\big)^{k-j+1} P_\ell G\big(X_{\ell,n}\big(t_{j-1}^n\big)\big)\, dM(s) \Bigg\|^2_{L^2(\Omega;H)} \Bigg) \\
&\quad = 3(I + II + III).
\end{aligned}$$

Assumption 3.1 implies for the first term

$$I \le C\big(h_\ell + \sqrt{\delta t^n}\big)^2 \|X_0\|^2_{L^2(\Omega;V)}.$$

The second term is decomposed into

$$\begin{aligned}
II \le 4\Bigg(&\Bigg\| \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} \big(S\big(t_k^n - s\big) - S\big(t_k^n - t_{j-1}^n\big)\big) F\big(X(s)\big)\, ds \Bigg\|^2_{L^2(\Omega;H)} \\
&+ \Bigg\| \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} \big(S\big(t_k^n - t_{j-1}^n\big) - r\big(\delta t^n A_\ell\big)^{k-j+1} P_\ell\big) F\big(X(s)\big)\, ds \Bigg\|^2_{L^2(\Omega;H)}
\end{aligned}$$

$$
\begin{aligned}
&+ \Big\| \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} r\big(\delta t^n A_\ell\big)^{k-j+1} P_\ell\big(F\big(X(s)\big) - F\big(X\big(t_{j-1}^n\big)\big)\big)\, ds \Big\|_{L^2(\Omega;H)}^2 \\
&+ \Big\| \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} r\big(\delta t^n A_\ell\big)^{k-j+1} P_\ell\big(F\big(X\big(t_{j-1}^n\big)\big) - F\big(X_{\ell,n}\big(t_{j-1}^n\big)\big)\big)\, ds \Big\|_{L^2(\Omega;H)}^2 \Bigg) \\
&= 4(II_a + II_b + II_c + II_d).
\end{aligned}
$$

Similarly, we get for term III

$$
\begin{aligned}
III \le 4\Bigg(&\Big\| \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} \big(S\big(t_k^n - s\big) - S\big(t_k^n - t_{j-1}^n\big)\big) G\big(X(s)\big)\, dM(s) \Big\|_{L^2(\Omega;H)}^2 \\
&+ \Big\| \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} \big(S\big(t_k^n - t_{j-1}^n\big) - r\big(\delta t^n A_\ell\big)^{k-j+1} P_\ell\big) G\big(X(s)\big)\, dM(s) \Big\|_{L^2(\Omega;H)}^2 \\
&+ \Big\| \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} r\big(\delta t^n A_\ell\big)^{k-j+1} P_\ell\big(G\big(X(s)\big) - G\big(X\big(t_{j-1}^n\big)\big)\big)\, dM(s) \Big\|_{L^2(\Omega;H)}^2 \\
&+ \Big\| \sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} r\big(\delta t^n A_\ell\big)^{k-j+1} P_\ell\big(G\big(X\big(t_{j-1}^n\big)\big) \\
&\qquad - G\big(X_{\ell,n}\big(t_{j-1}^n\big)\big)\big)\, dM(s) \Big\|_{L^2(\Omega;H)}^2 \Bigg) \\
= 4(III_a &+ III_b + III_c + III_d).
\end{aligned}
$$

Next, we give estimates for these eight terms. Eq. (2.5), Assumption 2.1, as well as Lemma 2.1 imply

$$
\begin{aligned}
II_a &\le C\delta t^n \Bigg(\sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} \big\| F\big(X(s)\big) \big\|_{L^2(\Omega;V)}\, ds \Bigg)^2 \\
&\le C\delta t^n \Bigg(\sum_{j=1}^{k} \int_{t_{j-1}^n}^{t_j^n} \big(1 + \|X(s)\|_{L^2(\Omega;V)}\big)\, ds \Bigg)^2 \\
&\le C(T)\delta t^n \big(1 + \|X_0\|_{L^2(\Omega;V)}^2\big).
\end{aligned}
$$

Similarly, with an application of Eq. (2.1) in addition, we have

$$
III_a \le C(T)\delta t^n \big(1 + \|X_0\|_{L^2(\Omega;V)}^2\big).
$$

The next terms are bounded with the use of Eq. (2.1), Assumption 3.1, Assumption 2.1, and Lemma 2.1 by

$$II_b + III_b \le C(T)\big(h_\ell + \sqrt{\delta t^n}\big)^2\big(1 + \|X_0\|^2_{L^2(\Omega;V)}\big).$$

For the terms labeled with c, besides Eq. (2.1), the stability of the rational approximation of the semigroup, the Lipschitz continuity of F and G (see Assumption 2.1), and Lemma 2.2 are used to obtain

$$II_c + III_c \le C(T)\delta t^n\big(1 + \|X_0\|^2_{L^2(\Omega;V)}\big).$$

Applying again Eq. (2.1) to III_d and Hölder's inequality to II_d, the stability of the rational approximation of the semigroup, and the Lipschitz continuity of F and G (see Assumption 2.1), we get

$$II_d + III_d \le C(T)\delta t^n \sum_{j=0}^{k-1} \big\|X\big(t_j^n\big) - X_{\ell,n}\big(t_j^n\big)\big\|^2_{L^2(\Omega;H)}.$$

Overall this leads to

$$\begin{aligned}
&\big\|X\big(t_k^n\big) - X_{\ell,n}\big(t_k^n\big)\big\|^2_{L^2(\Omega;H)} \\
&\quad\le C(T)\big(h_\ell + \sqrt{\delta t^n}\big)^2\big(1 + \|X_0\|^2_{L^2(\Omega;V)}\big) \\
&\qquad + C(T)\delta t^n \sum_{j=0}^{k-1} \big\|X\big(t_j^n\big) - X_{\ell,n}\big(t_j^n\big)\big\|^2_{L^2(\Omega;H)} \\
&\quad\le C(T)\big(h_\ell + \sqrt{\delta t^n}\big)^2\big(1 + \|X_0\|^2_{L^2(\Omega;V)}\big) \prod_{j=0}^{k-1}\big(1 + C(T)\delta t^n\big) \\
&\quad\le C(T)\big(h_\ell + \sqrt{\delta t^n}\big)^2\big(1 + \|X_0\|^2_{L^2(\Omega;V)}\big) \exp\big(T \cdot C(T)\big),
\end{aligned}$$

where we applied a discrete version of Gronwall's inequality (see [8]) in the second step. This proves the theorem. □

We establish a first error estimate for an approximation in space and time in combination with the Monte Carlo method, i.e., $\mathbb{E}[X(t)]$ is approximated by $E_N[X_{\ell,n}(t)]$ for $t \in \Theta^n$ as introduced in Eq. (4.2).

Theorem 4.2 *Let X be the mild solution of (2.6) and let $(X_{\ell,n}, \ell, n \in \mathbb{N}_0)$ be the sequence of discrete mild solutions introduced in Eq. (3.1). Then there exists a constant $C(T)$ such that for all $\ell, n \in \mathbb{N}_0$ and $N \in \mathbb{N}$*

$$\begin{aligned}
&\sup_{t\in\Theta^n} \big\|\mathbb{E}\big[X(t)\big] - E_N\big[X_{\ell,n}(t)\big]\big\|_{L^2(\Omega;H)} \\
&\qquad\le C(T)\left(h_\ell + \sqrt{\delta t^n} + \frac{1}{\sqrt{N}}\right)\big(1 + \|X_0\|_{L^2(\Omega;V)}\big).
\end{aligned}$$

Proof For $\ell, n \in \mathbb{N}_0$ and $t \in \Theta^n$, we split the left hand side of the equation above as follows

$$\begin{aligned}
&\big\|\mathbb{E}\big[X(t)\big] - E_N\big[X_{\ell,n}(t)\big]\big\|_{L^2(\Omega;H)} \\
&\quad\le \big\|\mathbb{E}\big[X(t)\big] - \mathbb{E}\big[X_{\ell,n}(t)\big]\big\|_{L^2(\Omega;H)} + \big\|\mathbb{E}\big[X_{\ell,n}(t)\big] - E_N\big[X_{\ell,n}(t)\big]\big\|_{L^2(\Omega;H)} \\
&\quad\le \big\|X(t) - X_{\ell,n}(t)\big\|_{L^2(\Omega;H)} + \big\|\mathbb{E}\big[X_{\ell,n}(t)\big] - E_N\big[X_{\ell,n}(t)\big]\big\|_{L^2(\Omega;H)}.
\end{aligned}$$

The first term on the right hand side is bounded by Theorem 4.1. The assertion follows with Lemma 4.1 and Remark 4.1 for the second term. □

Theorem 4.2 raises the question of the optimal time discretization level $n \in \mathbb{N}_0$ and Monte Carlo sampling size $N_\ell = N$ for any given space discretization level $\ell \in \mathbb{N}_0$. Let $h_\ell \simeq 2^{-\ell}$. The space and time errors are equilibrated if $\delta t^n \simeq h_\ell^2$ and therefore, $n = 2\ell$. With the convergence rate shown in Theorem 4.2, it can easily be seen that we equilibrate discretization and sampling error for $\ell \in \mathbb{N}_0$ by the choices

$$(N_\ell)^{-1/2} \simeq h_\ell, \quad \text{resp.} \quad N_\ell \simeq h_\ell^{-2}.$$

Let us assume that in each (implicit) time step the linear system associated to the discretized version of the operator A can be solved numerically in linear complexity, i.e., in $\mathcal{W}_\ell^H \simeq h_\ell^{-d}$ work and memory, where $d \in \mathbb{N}$ and e.g., $d = \dim D$ in the framework of Example 2.1. Then the overall work $\mathcal{W}_\ell$ is given by

$$\mathcal{W}_\ell = \mathcal{W}_\ell^H \cdot \mathcal{W}_{2\ell}^{\mathbb{T}} \cdot N_\ell \simeq h_\ell^{-d} \cdot h_\ell^{-2} \cdot h_\ell^{-2} = h_\ell^{-(d+4)} \simeq 2^{(d+4)\ell}.$$

Here, $\mathcal{W}_{2\ell}^{\mathbb{T}}$ denotes the work in time with respect to the time discretization $\Theta^{2\ell}$. The error bound in Theorem 4.2 in terms of the overall computational work $\mathcal{W}_\ell$ reads

$$\sup_{t \in \Theta^{2\ell}} \big\|\mathbb{E}\big[X(t)\big] - E_{2^{2\ell}}\big[X_{\ell,2\ell}(t)\big]\big\|_{L^2(\Omega;H)} \le C(T) h_\ell \simeq \mathcal{W}_\ell^{-1/(d+4)}. \tag{4.3}$$

4.2 Multilevel Monte Carlo approximation

After we have established error bounds for the singlelevel Monte Carlo method, we are in position to state and prove error versus complexity bounds for the multilevel Monte Carlo discretization.

The previous results on the convergence of the singlelevel Monte Carlo method and equilibration of the various error contributions suggest the use of sets of equidistant partitions $(\Theta^\ell, \ell \in \mathbb{N}_0)$ of the time interval $\mathbb{T} = [0, T]$ defined by

$$\Theta^\ell = \big\{t_{k(\ell)}^\ell = T 2^{-2\ell} k(\ell) = \delta t^\ell k(\ell), k(\ell) = 0, \ldots, 2^{2\ell}\big\},$$

i.e., with the notation of the previous section we have $n = 2\ell$. Here and in what follows, we denote the (constant) time steps in Θ^ℓ by $\delta t^\ell = T/2^{2\ell}$ for $\ell \in \mathbb{N}_0$. We set $h_\ell \simeq 2^{-\ell}$, $\ell \in \mathbb{N}_0$, and relate in the error analysis of the multilevel Monte Carlo discretization the spatial discretization level to the temporal discretization level by $n = 2\ell$, which explains the redefinition of the time grids. We abbreviate $X_{\ell,n}$ by X_ℓ.

For $L, k(L) \in \mathbb{N}_0$, $k(L) \leq 2^{2L}$, we recollect the multilevel Monte Carlo estimator for $\mathbb{E}[X(t^L_{k(L)})]$ introduced in the beginning of the section

$$E^L\big[X_L\big(t^L_{k(L)}\big)\big] = \sum_{\ell=0}^{L} E_{N_\ell}\big[X_\ell\big(t^L_{k(L)}\big) - X_{\ell-1}\big(t^L_{k(L)}\big)\big].$$

For each summand on the right hand side, we choose a level dependent number of samples N_ℓ for $\ell = 0, \ldots, L$ and we subtract the simulated solutions on two consecutive discretization levels (V_ℓ, Θ^ℓ) and $(V_{\ell-1}, \Theta^{\ell-1})$ *generated with the same random samples* of X_0 and M.

For a given time in Θ^ℓ, $\ell \in \{0, \ldots, L\}$, we linearly interpolate the solution on the next coarser grid to that time and define it as follows:

$$X_{\ell-1}\big(t^\ell_{k(\ell)}\big) = a_\ell X_{\ell-1}\big(t^{\ell-1}_{k(\ell-1)}\big) + b_\ell X_{\ell-1}\big(t^{\ell-1}_{k(\ell-1)+1}\big), \tag{4.4}$$

where

$$a_\ell = 1 - \left(\frac{k(\ell)}{4} - k(\ell-1)\right) \quad \text{and} \quad b_\ell = \frac{k(\ell)}{4} - k(\ell-1),$$

further, $k(\ell-1)$ is recursively defined by

$$k(\ell-1) = \left\lfloor \frac{k(\ell)}{4} \right\rfloor.$$

Here, for any number $\lambda > 0$, $\lfloor \lambda \rfloor$ denotes the next smaller integer below λ. Iterating Eq. (4.4), we may write for $\ell < L$

$$X_\ell\big(t^L_{k(L)}\big) = a_{\ell:L} X_\ell\big(t^\ell_{k(\ell)}\big) + b_{\ell:L} X_\ell\big(t^\ell_{k(\ell)+1}\big) \tag{4.5}$$

with

$$a_{\ell:L} = a_{\ell+1} - \sum_{i=\ell+2}^{L} \frac{1}{2^{2(i-(\ell+2)+1)}} b_i \quad \text{and} \quad b_{\ell:L} = b_{\ell+1} + \sum_{i=\ell+2}^{L} \frac{1}{2^{2(i-(\ell+2)+1)}} b_i.$$

We remark for further use that $a_{\ell:L} + b_{\ell:L} = 1$ for all $\ell \in \{0, \ldots L-1\}$.

With this linear interpolation, we have the following theorem on the convergence of the multilevel Monte Carlo estimator.

Theorem 4.3 *Let X be the mild solution of* (2.6) *and let $(X_\ell, \ell \in \mathbb{N}_0)$ be the sequence of discrete mild solutions introduced in Eq.* (3.1). *Then there exists a constant $C(T)$ such that for all $L \in \mathbb{N}_0$*

$$\sup_{t \in \Theta^L} \big\|\mathbb{E}\big[X(t)\big] - E^L\big[X_L(t)\big]\big\|_{L^2(\Omega;H)} \leq C(T)\left(h_L + \frac{1}{\sqrt{N_0}} + \sum_{\ell=0}^{L} \frac{1}{\sqrt{N_\ell}} h_\ell\right)\big(1 + \|X_0\|_{L^2(\Omega;V)}\big).$$

Proof For fixed $L \in \mathbb{N}_0$, we choose any $t^L_{k(L)} \in \Theta^L$. Similarly to the proof of Theorem 4.2, we split the error into

$$\begin{aligned}
&\left\|\mathbb{E}\left[X\left(t^L_{k(L)}\right)\right] - E^L\left[X_L\left(t^L_{k(L)}\right)\right]\right\|_{L^2(\Omega;H)} \\
&\quad \leq \left\|X\left(t^L_{k(L)}\right) - X_L\left(t^L_{k(L)}\right)\right\|_{L^2(\Omega;H)} \\
&\qquad + \sum_{\ell=0}^{L}\left\|(\mathbb{E} - E_{N_\ell})\left[X_\ell\left(t^L_{k(L)}\right) - X_{\ell-1}\left(t^L_{k(L)}\right)\right]\right\|_{L^2(\Omega;H)} \\
&\quad = I + II.
\end{aligned}$$

With Theorem 4.1 and the assumption that $\delta t^\ell \simeq h_\ell^2$, I is bounded by

$$\begin{aligned}
\left\|X\left(t^L_{k(L)}\right) - X_L\left(t^L_{k(L)}\right)\right\|_{L^2(\Omega;H)} &\leq C(T)\left(h_L + \sqrt{\delta t^L}\right)\left(1 + \|X_0\|_{L^2(\Omega;V)}\right) \\
&\leq C(T)h_L\left(1 + \|X_0\|_{L^2(\Omega;V)}\right).
\end{aligned}$$

Applying Lemma 4.1, we get for all summands in II

$$\begin{aligned}
&\left\|(\mathbb{E} - E_{N_\ell})\left[X_\ell\left(t^L_{k(L)}\right) - X_{\ell-1}\left(t^L_{k(L)}\right)\right]\right\|_{L^2(\Omega;H)} \\
&\quad \leq \frac{1}{\sqrt{N_\ell}}\left\|X_\ell\left(t^L_{k(L)}\right) - X_{\ell-1}\left(t^L_{k(L)}\right)\right\|_{L^2(\Omega;H)} \\
&\quad \leq \frac{1}{\sqrt{N_\ell}}\left(\left\|X_\ell\left(t^L_{k(L)}\right) - X\left(t^L_{k(L)}\right)\right\|_{L^2(\Omega;H)} + \left\|X\left(t^L_{k(L)}\right) - X_{\ell-1}\left(t^L_{k(L)}\right)\right\|_{L^2(\Omega;H)}\right) \\
&\quad \leq \frac{1}{\sqrt{N_\ell}}(II_a + II_b).
\end{aligned}$$

Equation (4.5), together with Theorem 4.1 and Lemma 2.2 gives the approximation bound

$$\begin{aligned}
II_a &\leq \left\|a_{\ell:L}X_\ell\left(t^\ell_{k(\ell)}\right) + b_{\ell:L}X_\ell\left(t^\ell_{k(\ell)+1}\right) - \left(a_{\ell:L}X\left(t^L_{k(L)}\right) + b_{\ell:L}X\left(t^L_{k(L)}\right)\right)\right\|_{L^2(\Omega;H)} \\
&\leq \Big\|a_{\ell:L}\left(X_\ell\left(t^\ell_{k(\ell)}\right) - X\left(t^\ell_{k(\ell)}\right) + X\left(t^\ell_{k(\ell)}\right) - X\left(t^L_{k(L)}\right)\right) \\
&\quad + b_{\ell:L}\left(X_\ell\left(t^\ell_{k(\ell)+1}\right) - X\left(t^\ell_{k(\ell)+1}\right) + X\left(t^\ell_{k(\ell)+1}\right) - X\left(t^L_{k(L)}\right)\right)\Big\|_{L^2(\Omega;H)} \\
&\leq a_{\ell:L}\left\|\left(X_\ell\left(t^\ell_{k(\ell)}\right) - X\left(t^\ell_{k(\ell)}\right)\right\|_{L^2(\Omega;H)} + a_{\ell:L}\left\|X\left(t^\ell_{k(\ell)}\right) - X\left(t^L_{k(L)}\right)\right)\right\|_{L^2(\Omega;H)} \\
&\quad + b_{\ell:L}\left\|\left(X_\ell\left(t^\ell_{k(\ell)+1}\right) - X\left(t^\ell_{k(\ell)+1}\right)\right\|_{L^2(\Omega;H)} \\
&\quad + b_{\ell:L}\left\|X\left(t^\ell_{k(\ell)+1}\right) - X\left(t^L_{k(L)}\right)\right)\right\|_{L^2(\Omega;H)} \\
&\leq (a_{\ell:L} + b_{\ell:L})C(T)\left(h_\ell + \sqrt{\delta t^\ell}\right)\left(1 + \|X_0\|_{L^2(\Omega;V)}\right) \\
&\quad + a_{\ell:L}C(T)\sqrt{t^L_{k(L)} - t^\ell_{k(\ell)}}\left(1 + \|X_0\|_{L^2(\Omega;V)}\right) \\
&\quad + b_{\ell:L}C(T)\sqrt{t^\ell_{k(\ell)+1} - t^L_{k(L)}}\left(1 + \|X_0\|_{L^2(\Omega;V)}\right) \\
&\leq C(T)\left(h_\ell + \sqrt{\delta t^\ell}\right)\left(1 + \|X_0\|_{L^2(\Omega;V)}\right).
\end{aligned}$$

Similarly, we get for II_b with $\ell = 1, \dots, L$

$$II_b \le C(T)\big(h_{\ell-1} + \sqrt{\delta t^{\ell-1}}\big)\big(1 + \|X_0\|_{L^2(\Omega;V)}\big).$$

We use the fact that $h_\ell = 2^{-1}h_{\ell-1}$ and $\delta t^\ell = 2^{-2}\delta t^{\ell-1}$ to get overall

$$\begin{aligned}
&\big\|X_\ell\big(t^L_{k(L)}\big) - X_{\ell-1}\big(t^L_{k(L)}\big)\big\|_{L^2(\Omega;H)}\\
&\quad\le C(T)\big(h_\ell + h_{\ell-1} + \sqrt{\delta t^\ell} + \sqrt{\delta t^{\ell-1}}\big)\big(1 + \|X_0\|_{L^2(\Omega;V)}\big)\\
&\quad\le C(T)3\big(h_\ell + \sqrt{\delta t^\ell}\big)\big(1 + \|X_0\|_{L^2(\Omega;V)}\big).
\end{aligned}$$

For $\ell = 0$, we have with the properties of the mild solution that

$$II_b = \big\|X\big(t^L_{k(L)}\big)\big\|_{L^2(\Omega;H)} \le C(T)\big(1 + \|X_0\|_{L^2(\Omega;H)}\big).$$

With these results we get for II, after employing that $\delta t^\ell \simeq h_\ell^2$ for $\ell = 0, \dots, L$,

$$\begin{aligned}
&\sum_{\ell=0}^{L} \frac{1}{\sqrt{N_\ell}}\big(\big\|X_\ell\big(t^L_{k(L)}\big) - X_{\ell-1}\big(t^L_{k(L)}\big)\big\|_{L^2(\Omega;H)}\big)\\
&\quad\le C(T)\left(\frac{1}{\sqrt{N_0}} + \sum_{\ell=0}^{L}\frac{1}{\sqrt{N_\ell}}h_\ell\right)\big(1 + \|X_0\|_{L^2(\Omega;V)}\big).
\end{aligned}$$

The claim follows by

$$\begin{aligned}
&\big\|\mathbb{E}\big[X\big(t^L_{k(L)}\big)\big] - E^L\big[X_L\big(t^L_{k(L)}\big)\big]\big\|_{L^2(\Omega;H)}\\
&\quad\le C(T)h_L\big(1 + \|X_0\|_{L^2(\Omega;V)}\big)\\
&\qquad + C(T)\left(\frac{1}{\sqrt{N_0}} + \sum_{\ell=0}^{L}\frac{1}{\sqrt{N_\ell}}h_\ell\right)\big(1 + \|X_0\|_{L^2(\Omega;V)}\big)\\
&\quad\le C(T)\left(h_L + \frac{1}{\sqrt{N_0}} + \sum_{\ell=0}^{L}\frac{1}{\sqrt{N_\ell}}h_\ell\right)\big(1 + \|X_0\|_{L^2(\Omega;V)}\big).
\end{aligned}$$

□

We may now relate the work in space $\mathcal{W}_\ell^H$, the number of time steps $\mathcal{W}_\ell^{\mathbb{T}}$, and the number of Monte Carlo samples N_ℓ at level $\ell = 0, \dots, L$ such that the errors in Theorem 4.3 are equilibrated.

Theorem 4.4 *Assume that for $\ell \in \mathbb{N}_0$, the computation of the linear system associated to the discretized version of the operator A can be solved numerically in $\mathcal{W}_\ell^H \simeq h_\ell^{-d} \simeq 2^{d\ell}$ work and memory for some $d \in \mathbb{N}$. Then the error contributions in Theorem 4.3 are equilibrated when the number of time steps is set to $\mathcal{W}_\ell^{\mathbb{T}} \simeq h_\ell^{-2} \simeq 2^{2\ell}$ and the number of Monte Carlo samples is*

$$N_0 \simeq 2^{2L} \quad \textit{and} \quad N_\ell \simeq 2^{2(L-\ell)}\ell^{2(1+\epsilon)} \tag{4.6}$$

for $\ell = 1, \ldots, L$ *and any* $\epsilon > 0$, $L \in \mathbb{N}_0$. *Therefore, the multilevel Monte Carlo method converges with rate*

$$\sup_{t \in \Theta^L} \left\| \mathbb{E}\left[X(t)\right] - E^L\left[X_L(t)\right] \right\|_{L^2(\Omega;H)} \leq C(T,\epsilon)\left(1 + \|X_0\|_{L^2(\Omega;V)}\right) h_L. \tag{4.7}$$

The total number of operations $\mathcal{W}_L$ *for the computation of the multilevel Monte Carlo estimate* $(E^L[X_L(t)], t \in \Theta^L)$ *is bounded by*

$$\mathcal{W}_L \leq C 2^{(d+2+\epsilon)L} \simeq h_L^{-(d+2+\epsilon)}.$$

In particular, in the multilevel Monte Carlo discretization, the error (4.7) *is related to the overall work* $\mathcal{W}_L$ *by*

$$\sup_{t \in \Theta^L} \left\| \mathbb{E}\left[X(t)\right] - E^L\left[X_L(t)\right] \right\|_{L^2(\Omega;H)} \leq C(T,\epsilon)\left(1 + \|X_0\|_{L^2(\Omega;V)}\right) \mathcal{W}_L^{-1/(d+2+\epsilon)}.$$

Proof For the error to be equilibrated, we choose the number N_ℓ of Monte Carlo samples on discretization level ℓ such that

$$\sum_{\ell=0}^{L} N_\ell^{-1/2} h_\ell \leq N_0^{-1/2}(1 + h_0) + \sum_{\ell=1}^{L} N_\ell^{-1/2} h_\ell \leq C h_L.$$

Setting N_ℓ according to Eq. (4.6) and $h_\ell \simeq 2^{-\ell}$, we derive

$$N_0^{-1/2}(1 + h_0) \simeq h_L \quad \text{and} \quad N_\ell^{-1/2} \simeq 2^{(\ell-L)} \ell^{-(1+\epsilon)}$$

and therefore,

$$\sum_{\ell=1}^{L} N_\ell^{-1/2} h_\ell \simeq \sum_{\ell=1}^{L} 2^{-\ell} 2^{(\ell-L)} \ell^{-(1+\epsilon)} \simeq h_L \sum_{\ell=1}^{L} \ell^{-(1+\epsilon)} \leq \zeta(1+\epsilon) h_L,$$

where ζ denotes the Riemann zeta function. When we insert this estimate into the error bound in Theorem 4.3, we get immediately Eq. (4.7).

The work $\mathcal{W}_L$ to obtain the multilevel Monte Carlo approximation is bounded by

$$\begin{aligned} \mathcal{W}_L &= \sum_{\ell=0}^{L} \mathcal{W}_\ell^H \cdot \mathcal{W}_{2\ell}^{\mathbb{T}} \cdot N_\ell \simeq 2^{d \cdot 0} 2^0 2^{2L} + \sum_{\ell=1}^{L} 2^{d\ell} 2^{2\ell} 2^{2(L-\ell)} \ell^{2(1+\epsilon)} \\ &\leq C 2^{(d+2)L} L^{2(1+\epsilon)} \leq C 2^{(d+2+\epsilon)L} \\ &\simeq h_L^{-(d+2+\epsilon)}, \end{aligned}$$

where $C > 0$ is a constant varying from line to line that is independent of L. Inserting this estimate into the error bound (4.7) and noting that $h_L \simeq 2^{-L}$, we receive the last assertion. □

5 Conclusions

In this paper we gave an a-priori error and complexity analysis of a Galerkin Euler–Maruyama discretization combined with multilevel Monte Carlo sampling for the numerical estimation of expectations of solutions of a class of parabolic partial differential equations driven by square integrable martingales. We proved that the multilevel Monte Carlo Euler–Maruyama approach lowers the computational complexity to calculate the expectation of the solution of a parabolic stochastic partial differential equation compared to a standard Monte Carlo method. The combination of different sample sizes on various subsequent discretization levels lowers the overall complexity to the complexity of one solve of the deterministic partial differential equation, namely, from $\mathrm{O}(2^{(d+4)L})$ to $\mathrm{O}(2^{(d+2+\epsilon)L})$ for a given refinement level L. Expressed in degrees of freedom of the numerical method, we obtain a bound for the computational complexity of the singlelevel Monte Carlo method of $\mathrm{O}(\mathcal{W}_L^H \cdot (\mathcal{W}_L^H)^{4/d})$ and for the multilevel Monte Carlo method of $\mathrm{O}(\mathcal{W}_L^H \cdot (\mathcal{W}_L^H)^{2/d})$. In other words, for $d = 1$ the work of multilevel Monte Carlo is dominated by the number of samples on the coarsest level. For higher space dimensions it is dominated by the effort to solve the corresponding partial differential equation. The low order of the convergence rate stems from the low regularity requirements that we assumed for the equation.

The results in the present paper could be extended in two directions. First, in [21] it is shown that under stronger assumptions on A, the requirement that the mild solution belongs to $L^2(\Omega; V)$ can be relaxed, while still maintaining the same order of convergence of the approximation. More precisely, if A is densely defined, self-adjoint, positive definite, and not necessarily bounded but with compact inverse, the Euler–Maruyama scheme still converges in mean square of order $1/2$ in time and 1 in space for less regular choices of F and G. A second extension of the present analysis of the multilevel Monte Carlo method is to couple it with a Milstein type method as introduced in [4]. There, the authors give an approximation scheme for a more general but linear advection-diffusion problem

$$dX(t) = (A + B)X(t)\,dt + G\big(X(t)\big)\,dM(t)$$

on a bounded domain $D \subset \mathbb{R}^d$, $d \in \mathbb{N}$, with $H = L^2(D)$, where A is the diffusion and B the advection operator. Similar results can be achieved when a nonlinearity $F : H \to H$ is added and sufficient assumptions on the regularity of the coefficients and the solution are made, such that the stochastic partial differential equation reads

$$dX(t) = (A + B)X(t) + F\big(X(t)\big)\,dt + G\big(X(t)\big)\,dW(t),$$

where we set $M = W$ although the following also holds for any continuous martingale in $\mathcal{M}_b^2(U)$. The increment in the Milstein scheme is then given by

$$\begin{aligned}
X_{\ell,n}\big(t_j^n\big) &= r\big(\delta t^n A_\ell\big) X_{\ell,n}\big(t_{j-1}^n\big) + \int_{t_{j-1}^n}^{t_j^n} r\big(\delta t^n A_\ell\big) P_\ell B X_{\ell,n}\big(t_{j-1}^n\big)\,ds \\
&\quad + \int_{t_{j-1}^n}^{t_j^n} r\big(\delta t^n A_\ell\big) P_\ell F\big(X_{\ell,n}\big(t_{j-1}^n\big)\big)\,ds \\
&\quad + \int_{t_{j-1}^n}^{t_j^n} r\big(\delta t^n A_\ell\big) P_\ell G\big(X_{\ell,n}\big(t_{j-1}^n\big)\big)\,dW(s) \\
&\quad + \int_{t_{j-1}^n}^{t_j^n} \bigg(r\big(\delta t^n A_\ell\big) P_\ell G\bigg(\int_{t_{j-1}^n}^{s} G\big(X_{\ell,n}\big(t_{j-1}^n\big)\big)\,dW(r)\bigg)\bigg)\,dW(s),
\end{aligned}$$

where $r(\delta t^n A_\ell)$ denotes the rational approximation of the semigroup generated by A that depends on the time approximation (with parameter δt^n) and space approximation level (with parameter ℓ). A combination of the proof in [4] and [16] leads to the following bound, dependent on the regularity parameter $\alpha \in \mathbb{N}$

$$\sup_{t\in\Theta^n}\left\|X(t)-X_{\ell,n}(t)\right\|_{L^2(\Omega;H)} \leq C_1\left(h_\ell^\alpha+\left(\delta t^n\right)^{\alpha/2}\right)\sup_{t\in\mathbb{T}}\left\|X(t)\right\|_{L^2(\Omega;H^\alpha)} + C_2\delta t^n\sup_{t\in\mathbb{T}}\left\|X(t)\right\|_{L^2(\Omega;H^1)}, \tag{5.1}$$

where C_1 and C_2 are constants. For $\alpha = 1$, which meets the prerequisites on the regularity of this paper, we would not get better convergence than for the Euler–Maruyama scheme introduced in this paper. However, for more regular data, i.e., $\alpha = 2$ and $X_0 \in L^2(\Omega; H^2(D) \cap H_0^1(D))$, we obtain in particular

$$\sup_{t\in\Theta^n}\left\|X(t)-X_{\ell,n}(t)\right\|_{L^2(\Omega;H)} = \mathrm{O}\left(h_\ell^2+\delta t^n\right).$$

We have still the (severe) consistency constraint $\delta t^n \simeq h_\ell^2$. To compare this error bound to the result for the Euler–Maruyama scheme proven in Theorem 4.2, we employ Lemma 4.1, which results in a convergence rate of $\mathrm{O}(h_\ell^2 + \delta t^n + N^{-1/2})$ for the singlelevel Monte Carlo method. To prove the result for the multilevel Monte Carlo approach, we need to change the linear interpolation of the solution given in Eq. (4.4) to

$$X_\ell(t) = r\left((t-\hat{t})A_\ell\right)X_\ell(\hat{t}) + \int_{\hat{t}}^t r\left((t-\hat{t})A_\ell\right)P_\ell B X_\ell(\hat{t})\,ds + \int_{\hat{t}}^t r\left((t-\hat{t})A_\ell\right)P_\ell F\left(X_\ell(\hat{t})\right)ds + \int_{\hat{t}}^t r\left((t-\hat{t})A_\ell\right)P_\ell G\left(X_\ell(\hat{t})\right)dW(s), \tag{4.4$'$}$$

where $\hat{t} = \lfloor t\rfloor_{\Theta^\ell}$, i.e., $\hat{t}$ is the largest time discretization point on level ℓ, which is smaller than t. This is the extension to all $t \in \mathbb{T}$ of the fully discrete mild solution. With this approach a convergence of X_ℓ in (4.4′) to the mild solution X in mean square of order $\mathrm{O}(\delta t^\ell + h_\ell^2)$ is achieved which is needed in the proof of Theorem 4.3. Using this interpolation we achieve an error bound of the multilevel Monte Carlo approximation for the Milstein scheme of $C(T)(h_L^2 + N_0^{-1/2} + \sum_{\ell=0}^L N_\ell^{-1/2} h_\ell^2)$ if we set $\delta t^\ell \simeq h_\ell^2$. This reduces the overall complexity of $\mathrm{O}(2^{(d+6)L})$ for the singlelevel Monte Carlo approximation to $\mathrm{O}(2^{4L})$ for $d = 1$ and $\mathrm{O}(2^{(d+2+\epsilon)L})$ for $d > 1$. Here, we choose for the optimal numbers of samples on each level $\ell = 1, \dots, L$ in the multilevel Monte Carlo method $N_\ell = 2^{4(L-\ell)}\ell^{2(1+\epsilon)}$ for $\epsilon > 0$ and $N_0 = 2^{4L}$. As in the case of an Euler–Maruyama approach, the work is dominated by the number of samples on the coarsest level in the multilevel Monte Carlo approach in low dimensions, whereas for $d > 1$ the work is dominated again by the Finite Element method. This leads to asymptotically the same work for the multilevel Monte Carlo approach for an Euler–Maruyama and a Milstein type scheme. The faster convergence of the Milstein method, leading to a higher sampling effort only comes into play in low dimensions where the overall work is dominated by the Monte Carlo method. Expressed in degrees of freedom, the difference between the computational complexity of the singlelevel Monte Carlo and the multilevel Monte Carlo method for a Milstein scheme becomes even more apparent: For $d > 1$ we have for the multilevel Monte Carlo method $\mathrm{O}(\mathcal{W}_L^H \cdot (\mathcal{W}_L^H)^{2/d})$, the same as for the multilevel Monte Carlo coupled with an Euler–Maruyama method, whereas for $d = 1$ we have $\mathrm{O}((\mathcal{W}_L^H)^4)$; for the singlelevel Monte Carlo method we get $\mathrm{O}(\mathcal{W}_L^H \cdot (\mathcal{W}_L^H)^{6/d})$. Here, we assumed that generating the source term in the case of the Milstein method can be done with the same computational effort as for the Euler–Maruyama scheme, i.e., the Karhunen–Loève expansion of the driving noise is truncated according to Lemma 4.1 and Lemma 4.2 in [4]. Note, that in Eq. (5.1) the convergence rate depends on the regularity parameter α. In other words, for a solution with less regularity, more sampling is necessary to balance Monte Carlo and (time and space) discretization errors.

Equation (5.1) indicates that the ratio between spatial and temporal discretization error is the same for the Milstein method as for the Euler–Maruyama method (provided, however, that additional regularity of the data in the case of the Milstein scheme is used). Nevertheless, Eq. (5.1) also reveals that a lower regularity in space does not imply a different ratio of space and time step for the Milstein scheme. The optimal relationship of space and time stepping remains $\delta t^n \simeq h_\ell^2$ to ensure equilibration of the consistency errors. Other choices like $\delta t^n \simeq h_\ell^\beta$ for $\beta \in (0, 2]$ do not improve the relationship of work versus accuracy under the considered regularity assumptions. Moreover, this relationship is optimal for $\beta = 2$.

In this paper we do not approximate the noise. However, in [5] the authors show how to truncate the Karhunen–Loève expansion dependent on the number of discretization points, e.g., the discretization level. In [4], an extended result for the case of a Milstein scheme is presented.

Acknowledgements This research was supported in part by the European Research Council under grant ERC AdG 247277.

References

1. Auscher, P., Tchamitchian, P.: Square roots of elliptic second order divergence operators on strongly Lipschitz domains: L^2 theory. J. Anal. Math. **90**, 1–12 (2003)
2. Barth, A.: A finite element method for martingale-driven stochastic partial differential equations. Commum. Stoch. Anal. **4**(3), 355–375 (2010)
3. Barth, A., Lang, A.: L^p and almost sure convergence of a Milstein scheme for stochastic partial differential equations. SAM Report 2011-15, Seminar for Applied Mathematics, ETH Zurich (2011)
4. Barth, A., Lang, A.: Milstein approximation for advection-diffusion equations driven multiplicative noncontinuous martingale noises. Appl. Math. Optim. (2012). doi:10.1007/s00245-012-9176-y
5. Barth, A., Lang, A.: Simulation of stochastic partial differential equations using finite element methods. Stochastics **84**(2–3), 217–231 (2012)
6. Barth, A., Schwab, Ch., Zollinger, N.: Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients. Numer. Math. **119**(1), 123–161 (2011)
7. Chow, P.L., Jiang, J.L., Menaldi, J.L.: Pathwise convergence of approximate solutions to Zakai's equation in a bounded domain. In: Da Prato, G. et al. (eds.) Stochastic Partial Differential Equations and Applications. Proceedings of the Third Meeting on Stochastic Partial Differential Equations and Applications held at Villa Madruzzo, Trento, Italy, January 1990. Pitman Res. Notes Math. Ser., vol. 268, pp. 111–123. Longman Scientific & Technical, Harlow (1992)

8. Clark, D.S.: Short proof of a discrete Gronwall inequality. Discrete Appl. Math. **16**, 279–281 (1987)
9. Da Prato, G., Zabczyk, J.: Stochastic Equations in Infinite Dimensions. Encyclopedia of Mathematics and Its Applications, vol. 44. Cambridge University Press, Cambridge (1992)
10. Engel, K.J., Nagel, R.: One-Parameter Semigroups for Linear Evolution Equations. Graduate Texts in Mathematics, vol. 194. Springer, Berlin (2000)
11. Geissert, M., Kovács, M., Larsson, S.: Rate of weak convergence of the finite element method for the stochastic heat equation with additive noise. BIT Numer. Math. **49**(2), 343–356 (2009)
12. Giles, M.B.: Improved multilevel Monte Carlo convergence using the Milstein scheme. In: Keller, A. et al. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2006. Selected Papers Based on the Presentations at the 7th International Conference 'Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing', Ulm, Germany, August 14–18, 2006, pp. 343–358. Springer, Berlin (2008)
13. Giles, M.B.: Multilevel Monte Carlo path simulation. Oper. Res. **56**(3), 607–617 (2008)
14. Hutzenthaler, M., Jentzen, A., Kloeden, P.E.: Divergence of the multilevel Monte Carlo method (2011). arXiv:1105.0226 [math.PR]
15. Jentzen, A., Kloeden, P.E.: The numerical approximation of stochastic partial differential equations. Milan J. Math. **77**(1), 205–244 (2009)
16. Jentzen, A., Röckner, M.: A Milstein scheme for SPDEs (2010). arXiv:1001.2751 [math.NA]
17. Kantorovitz, S.: On Liu's analyticity criterion for semigroups. Semigroup Forum **53**(2), 262–265 (1996)
18. Kato, T.: Perturbation Theory for Linear Operators, Corr. printing of the 2nd edn. Grundlehren der mathematischen Wissenschaften, vol. 132. Springer, Berlin (1980)
19. Kloeden, P., Lord, G., Neuenkirch, A., Shardlow, T.: The exponential integrator scheme for stochastic partial differential equations: Pathwise error bounds. J. Comput. Appl. Math. **235**(5), 1245–1260 (2011)
20. Kovács, M., Larsson, S., Saedpanah, F.: Finite element approximation of the linear stochastic wave equation with additive noise. SIAM J. Numer. Anal. **48**, 408–427 (2010)
21. Kruse, R.: Optimal error estimates of Galerkin finite element methods for stochastic partial differential equations with multiplicative noise (2011). arXiv:1103.4504 [math.NA]
22. Lang, A.: Mean square convergence of a semidiscrete scheme for SPDEs of Zakai type driven by square integrable martingales. Proc. Comput. Sci. **1**(1), 1609–1617 (2010). ICCS 2010
23. Mishra, S., Schwab, Ch.: Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data. Math. Comput. **81**, 1979–2018 (2012)
24. Pazy, A.: Semigroups of Linear Operators and Applications to Partial Differential Equations. Applied Mathematical Sciences, vol. 44. Springer, New York (1983)
25. Peszat, S., Zabczyk, J.: Stochastic Partial Differential Equations with Lévy Noise. Encyclopedia of Mathematics and Its Applications, vol. 113. Cambridge University Press, Cambridge (2007)
26. Thomée, V.: Galerkin Finite Element Methods for Parabolic Problems, 2nd revised and expanded edn. Springer Series in Computational Mathematics, vol. 25. Springer, Berlin (2006)