

Multilevel Monte Carlo method
with applications to stochastic
partial differential equations

A. Barth and A. Lang

Revised: June 2012

Research Report No. 2011-68
November 2011

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

MULTILEVEL MONTE CARLO METHOD WITH APPLICATIONS TO STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS

ANDREA BARTH AND ANNIKA LANG

ABSTRACT. In this work, the approximation of Hilbert-space-valued random variables is combined with the approximation of the expectation by a multilevel Monte Carlo method. The number of samples on the different levels of the multilevel approximation are chosen such that the errors are balanced. The overall work then decreases in the optimal case to $O(h^{-2})$ if h is the error of the approximation. The multilevel Monte Carlo method is applied to functions of solutions of parabolic and hyperbolic stochastic partial differential equations as needed e.g., for option pricing. Simulations complete the paper.

1. INTRODUCTION

In science, various problems with uncertainties arise. Some of these problems are for example given by stochastic differential equations, by stochastic partial differential equations, or by a random partial differential equations. The uncertainty may be in the initial condition of the system, in the shape of the domain, in the diffusion coefficients, or in some external noise that enters the system. One is for example interested in the expected value of the system or of the variance. A possible way to estimate these values besides deterministic methods like stochastic Galerkin methods, is to use Monte Carlo methods. In the latter, the unknown state like the solution of a stochastic differential equation is approximated and then simulated many times. The average over all simulations is then an estimator for the expected solution. This leads in its classical version to a computationally expensive method. To reduce the computational work, Heinrich introduced in [13] the multilevel Monte Carlo method to approximate functionals of Banach-space-valued random variables. In [11], Giles developed the multilevel Monte Carlo method for stochastic differential equations. This method combines the error in the estimation of an expectation in an optimal way with the errors that arise due to the approximation of the solution of a stochastic differential equation. In the last years, various authors have used this method for different problems. It was considered for stochastic differential equations driven by a Brownian motion in [14, 15] and in the references therein and driven by a Lévy process e.g. in [9, 18]. Applications to random PDEs can be found in [5, 7, 19] among others, while stochastic partial differential equations of Itô type were considered in [4, 12].

In this paper, we detach the multilevel Monte Carlo estimator for the expectation of a random variable from the differential equation. Therefore, we consider a Hilbert-space-valued random variable Y and a sequence of approximations $(Y_\ell, \ell \in \mathbb{N}_0)$. Our interest is to balance the errors that occur from the approximation of the random variable and from the sampling

Key words and phrases. Multilevel Monte Carlo, stochastic partial differential equations, stochastic Finite Element methods, stochastic parabolic equation, stochastic hyperbolic equation, multilevel approximations.

This research was supported in part by the European Research Council under grant ERC AdG 247277. The authors further thank the anonymous referee for very helpful comments.

such that the error remains the same but the computational work is smaller than with standard (singlelevel) Monte Carlo methods. We therefore have to choose the number of samples on each level of estimation in relation to the speed of convergence of the approximation of the random variable. It is shown that if the variance of $(Y_\ell - Y_{\ell-1}, \ell \in \mathbb{N})$ converges fast enough, the overall work is of order $O(h_\ell^{-2})$, where $O(h_\ell)$ is the rate of weak convergence of the approximation $(Y_\ell, \ell \in \mathbb{N}_0)$ to Y .

The introduced multilevel Monte Carlo method is applied to stochastic partial differential equations as introduced in [21]. We give examples of parabolic and hyperbolic equations and estimate Lipschitz functions of the solution. The framework includes payoff functions of European options written, for instance, on forward contracts that can be calculated with the introduced method more efficiently than with standard Monte Carlo methods that are mainly used today. Furthermore, we show how the knowledge of weak and strong convergence results influences the number of samples that have to be chosen according to the theory and therefore the overall work.

This work is organized as follows: In Section 2, the multilevel Monte Carlo method is introduced and discussed for Hilbert-space-valued random variables. This algorithm is applied to parabolic and hyperbolic stochastic partial differential equations with additive and multiplicative martingale noises in Section 3. Finally, in Section 4, simulation results are shown for the heat equation and the scalar advection equation, both driven by additive Wiener noise.

2. MULTILEVEL MONTE CARLO FOR RANDOM VARIABLES

In this section, we derive a convergence and a work versus accuracy result for the multilevel Monte Carlo estimator of a Hilbert-space-valued random variable. This is used to calculate errors and computational work for the approximation of stochastic partial differential equations in Section 3. A multilevel Monte Carlo method for (more general) Banach-space-valued random variables has been introduced in [13], where the author derives bounds on the error for given work. Here, we do the contrary and bound the overall work for a given accuracy.

First, we state a lemma on the convergence in the number of samples of a Monte Carlo estimator. Therefore, let Y be a random variable with values in a Hilbert space B and $(\hat{Y}^i, i \in \mathbb{N})$ be a sequence of independent, identically distributed copies of Y . Then, the strong law of large numbers states that the *Monte Carlo estimator* $E_N[Y]$ defined by

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

converges P -almost surely to $\mathbb{E}[Y]$ for $N \rightarrow +\infty$. In the following lemma we see that it also converges in mean square to $\mathbb{E}[Y]$ if Y is square integrable, i.e., $Y \in L^2(\Omega; B)$ with

$$L^2(\Omega; B) := \{v : \Omega \rightarrow B, v \text{ strongly measurable, } \|v\|_{L^2(\Omega; B)} < +\infty\},$$

where

$$\|v\|_{L^2(\Omega; B)} = \mathbb{E}[\|v\|_B^2]^{1/2}.$$

In contrast to the almost sure convergence of $E_N[Y]$ derived from the strong law of large numbers, in mean square, a convergence rate can be deduced from the following lemma in terms of $N \in \mathbb{N}$.

Lemma 2.1. *For any $N \in \mathbb{N}$ and for $Y \in L^2(\Omega; B)$, it holds that*

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

The lemma is proven in [4, Lemma 4.1]. It shows that the sequence of so-called *Monte Carlo estimators* ($E_N[Y], N \in \mathbb{N}$) converges with rate $O(N^{-1/2})$ in mean square to the expectation of Y .

Next, let us assume that $(Y_\ell, \ell \in \mathbb{N}_0)$ is a sequence of approximations of Y , e.g., $Y_\ell \in V_\ell$, where $(V_\ell, \ell \in \mathbb{N}_0)$ is a sequence of finite dimensional subspaces of B . For given $L \in \mathbb{N}_0$, it holds that

$$Y_L = Y_0 + \sum_{\ell=1}^L (Y_\ell - Y_{\ell-1})$$

and due to the linearity of the expectation that

$$\mathbb{E}[Y_L] = \mathbb{E}[Y_0] + \sum_{\ell=1}^L \mathbb{E}[Y_\ell - Y_{\ell-1}].$$

A possible way to approximate $\mathbb{E}[Y_L]$ is to approximate $\mathbb{E}[Y_\ell - Y_{\ell-1}]$ with the corresponding Monte Carlo estimator $E_{N_\ell}[Y_\ell - Y_{\ell-1}]$ with a number of independent samples N_ℓ depending on the level ℓ . We set

$$E^L[Y_L] := E_{N_0}[Y_0] + \sum_{\ell=1}^L E_{N_\ell}[Y_\ell - Y_{\ell-1}]$$

and call $E^L[Y_L]$ the *multilevel Monte Carlo estimator* of $\mathbb{E}[Y_L]$. The following lemma gives convergence results for the estimator depending on the order of weak convergence of $(Y_\ell, \ell \in \mathbb{N}_0)$ to Y and the convergence of the variance of $(Y_\ell - Y_{\ell-1}, \ell \in \mathbb{N})$. If neither estimates on the weak convergence nor on the convergence of the variance are available, one can use — the in general slower — strong convergence rates.

Lemma 2.2. *Let $Y \in L^2(\Omega; B)$ and let $(Y_\ell, \ell \in \mathbb{N}_0)$ be a sequence in $L^2(\Omega; B)$, then, for $L \in \mathbb{N}_0$, it holds that*

$$\begin{aligned} & \|\mathbb{E}[Y] - E^L[Y_L]\|_{L^2(\Omega; B)} \\ & \leq \|\mathbb{E}[Y - Y_L]\|_B + \frac{1}{\sqrt{N_0}} \text{Var}[Y_0]^{1/2} + \sum_{\ell=1}^L \frac{1}{\sqrt{N_\ell}} \text{Var}[Y_\ell - Y_{\ell-1}]^{1/2} \\ & \leq \|Y - Y_L\|_{L^2(\Omega; B)} + \sum_{\ell=0}^L \frac{1}{\sqrt{N_\ell}} (\|Y - Y_\ell\|_{L^2(\Omega; B)} + \|Y - Y_{\ell-1}\|_{L^2(\Omega; B)}), \end{aligned}$$

where $Y_{-1} = 0$.

Proof. First, we observe that

$$\begin{aligned} \|\mathbb{E}[Y] - E^L[Y_L]\|_{L^2(\Omega; B)} & \leq \|\mathbb{E}[Y] - \mathbb{E}[Y_L]\|_{L^2(\Omega; B)} + \|\mathbb{E}[Y_L] - E^L[Y_L]\|_{L^2(\Omega; B)} \\ & = \|\mathbb{E}[Y - Y_L]\|_B + \|\mathbb{E}[Y_L] - E^L[Y_L]\|_{L^2(\Omega; B)}. \end{aligned}$$

We call the first component on the right hand side the weak convergence of $(Y_\ell, \ell \in \mathbb{N}_0)$ to Y . The second term satisfies by the linearity of the expectation that

$$\begin{aligned} & \|\mathbb{E}[Y_L] - E^L[Y_L]\|_{L^2(\Omega; B)} \\ &= \left\| \mathbb{E}[Y_0] - E_{N_0}[Y_0] + \sum_{\ell=1}^L (\mathbb{E}[Y_\ell - Y_{\ell-1}] - E_{N_\ell}[Y_\ell - Y_{\ell-1}]) \right\|_{L^2(\Omega; B)} \\ &\leq \|\mathbb{E}[Y_0] - E_{N_0}[Y_0]\|_{L^2(\Omega; B)} + \sum_{\ell=1}^L \|\mathbb{E}[Y_\ell - Y_{\ell-1}] - E_{N_\ell}[Y_\ell - Y_{\ell-1}]\|_{L^2(\Omega; B)}. \end{aligned}$$

Now, Lemma 2.1 implies the first assertion. The second inequality follows from the properties of the integral, i.e., for an integrable, B -valued random variable Y , it holds that

$$\|\mathbb{E}[Y]\|_B \leq \mathbb{E}[\|Y\|_B],$$

on the one hand side and from the fact that

$$\text{Var}[Y_\ell - Y_{\ell-1}]^{1/2} \leq \|Y_\ell - Y_{\ell-1}\|_{L^2(\Omega; B)} \leq \|Y - Y_\ell\|_{L^2(\Omega; B)} + \|Y - Y_{\ell-1}\|_{L^2(\Omega; B)}$$

on the other hand side. \square

This lemma enables us to choose for given weak convergence of $(Y_\ell, \ell \in \mathbb{N}_0)$ and for given convergence of the variance of $(Y_\ell - Y_{\ell-1}, \ell \in \mathbb{N})$ the number of samples N_ℓ on each level $\ell \in \mathbb{N}_0$ such that all terms in the error estimate are equilibrated. Furthermore, we provide bounds on work versus accuracy. As for these bounds constants are essential, we explicitly specify them in the proof. A similar result for real-valued random variables can be found in [12].

Theorem 2.3. *Let $(Y_\ell, \ell \in \mathbb{N}_0)$ converge weakly to Y of order $\alpha > 0$, i.e., there exists a constant C_1 such that*

$$\|\mathbb{E}[Y - Y_\ell]\|_B \leq C_1 2^{-\alpha\ell},$$

for $\ell \in \mathbb{N}_0$. Furthermore, assume that the variance of $(Y_\ell - Y_{\ell-1}, \ell \in \mathbb{N})$ converges with order $\beta > 0$, $\beta \leq \alpha$, i.e., there exists a constant C_2 such that

$$\text{Var}[Y_\ell - Y_{\ell-1}] \leq (C_2)^2 2^{-2\beta\ell},$$

and that $\text{Var}[Y_0] = (C_3)^2$. For chosen level $L \in \mathbb{N}_0$, set $N_\ell = 2^{2(\alpha L - \beta\ell)} \ell^{2(1+\epsilon)}$, $\ell = 1, \dots, L$, $\epsilon > 0$, and $N_0 = 2^{2\alpha L}$, then, the error is bounded by

$$\|\mathbb{E}[Y_L] - E^L[Y_L]\|_{L^2(\Omega; B)} \leq (C_1 + C_3 + C_2 \zeta(1 + \epsilon)) 2^{-\alpha L} =: h_L,$$

where ζ denotes the Riemann zeta function, i.e., $\|\mathbb{E}[Y_L] - E^L[Y_L]\|_{L^2(\Omega; B)}$ has the same order of convergence as $\|\mathbb{E}[Y - Y_\ell]\|_B$. Assume further that the work \mathcal{W}_ℓ^B of one calculation of $Y_\ell - Y_{\ell-1}$, $\ell \geq 1$, is bounded by $C_4 2^{\gamma\ell}$ for a constant C_4 and $\gamma > 0$, that the work to calculate Y_0 is bounded by a constant C_5 , and that the addition of the Monte Carlo estimators costs $C_6 \sum_{\ell=1}^L 2^{\delta\ell}$ for some $\delta \geq 0$ and some constant C_6 . Then, the overall work \mathcal{W}_L is bounded by

$$\mathcal{W}_L = \begin{cases} O(h_L^{-\max\{2, \delta/\alpha\}}) & \text{if } \gamma < 2\beta, \\ O(\max\{h_L^{-(2+(\gamma-2\beta)/\alpha)} |\log(h_L)|^{3+2\epsilon}, h_L^{-\delta/\alpha}\}) & \text{if } \gamma \geq 2\beta. \end{cases}$$

Proof. First, we calculate the error. It holds with the made assumptions that

$$\frac{1}{\sqrt{N_0}} \text{Var}[Y_0]^{1/2} = C_3 2^{-\alpha L}$$

and for $\ell = 1, \dots, L$ that

$$\frac{1}{\sqrt{N_\ell}} \text{Var}[Y_\ell - Y_{\ell-1}]^{1/2} \leq C_2 2^{\beta\ell - \alpha L} \ell^{-(1+\epsilon)} 2^{-\beta\ell} = C_2 2^{-\alpha L} \ell^{-(1+\epsilon)}.$$

So overall we get that

$$\sum_{\ell=1}^L \frac{1}{\sqrt{N_\ell}} \text{Var}[Y_\ell - Y_{\ell-1}]^{1/2} \leq C_2 2^{-\alpha L} \sum_{\ell=1}^L \ell^{-(1+\epsilon)} \leq C_2 2^{-\alpha L} \zeta(1+\epsilon),$$

where ζ denotes the Riemann zeta function. To calculate the error, we assemble all estimates to

$$\|\mathbb{E}[Y_L] - E^L[Y_L]\|_{L^2(\Omega; B)} \leq (C_1 + C_3 + C_2 \zeta(1+\epsilon)) 2^{-\alpha L}.$$

Next, we calculate the necessary work to achieve this error. The overall work consists of the work \mathcal{W}_ℓ^B to calculate $Y_\ell - Y_{\ell-1}$ times the number of samples N_ℓ on all level $\ell = 1, \dots, L$, the work \mathcal{W}_0^B on level 0, and the addition of the Monte Carlo estimators in the end. Therefore, we have

$$\begin{aligned} \mathcal{W}_L &\leq C_5 N_0 + C_4 \sum_{\ell=1}^L N_\ell 2^{\gamma\ell} + C_6 \sum_{\ell=1}^L 2^{\delta\ell} \\ &\leq C_5 2^{2\alpha L} + C_4 \sum_{\ell=1}^L 2^{2(\alpha L - \beta\ell)} \ell^{2(1+\epsilon)} 2^{\gamma\ell} + C_6 \left(\frac{2^{\delta(L+1)} - 1}{2^\delta - 1} - 1 \right) \\ &\leq 2^{2\alpha L} (C_5 + C_4 \sum_{\ell=1}^L 2^{(\gamma-2\beta)\ell} \ell^{2(1+\epsilon)}) + C_6 \frac{2^\delta}{2^\delta - 1} 2^{\delta L}. \end{aligned}$$

If $\gamma < 2\beta$, the sum is absolutely convergent and

$$\mathcal{W}_L \leq (C_5 + C_4 C) 2^{2\alpha L} + C_6 \frac{2^\delta}{2^\delta - 1} 2^{\delta L} = O(h_L^{-\max\{2, \delta/\alpha\}}).$$

For $\gamma \geq 2\beta$, it holds that

$$\begin{aligned} \mathcal{W}_L &\leq 2^{2\alpha L} (C_5 + C_4 2^{(\gamma-2\beta)L} L^{3+2\epsilon}) + C_6 \frac{2^\delta}{2^\delta - 1} 2^{\delta L} \\ &= O(\max\{h_L^{-(2+(\gamma-2\beta)/\alpha)} |\log(h_L)|^{3+2\epsilon}, h_L^{-\delta/\alpha}\}). \quad \square \end{aligned}$$

The error estimates also stay true for $\alpha < \beta$, if one sets $N_\ell = \max\{2^{2(\alpha L - \beta\ell)} \ell^{2(1+\epsilon)}, 1\}$. The overall work \mathcal{W}_L is then dominated by the number of samples $N_0 = 2^{2\alpha L}$ on the coarsest level and the work for one solve on the finest level $\mathcal{W}_L^B = O(2^{\gamma L})$. The work versus accuracy analysis leads therefore to

$$\mathcal{W}_L = O(h_L^{-2 \cdot \max(1, \gamma/\alpha)}),$$

for $\gamma < 2\beta$. Nevertheless, if one has the approximation of stochastic partial differential equations in mind, one will most likely bound $\|\mathbb{E}[Y - Y_\ell]\|_B$ with the order of weak convergence and $\text{Var}[Y_\ell - Y_{\ell-1}]$ with the order of strong convergence and it holds that the order of weak convergence is at least as good as the order of strong convergence, i.e., $\alpha \geq \beta$.

We remark that the computation of the sum of the Monte Carlo estimators does not increase the computational complexity if $Y_\ell \in V_\ell$, for all $\ell \in \mathbb{N}_0$, and $(V_\ell, \ell \in \mathbb{N}_0)$ is a sequence of nested finite dimensional subspaces of B .

3. APPLICATION TO STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS

In this section, we apply the multilevel Monte Carlo results from the previous section to stochastic partial differential equations. We aim to approximate expressions of the form $\mathbb{E}[\varphi(X(t))]$, where $X = (X(t), t \in [0, T])$ is the solution of a stochastic partial differential equation and φ is a measurable mapping from a separable Hilbert space H into a separable Hilbert space B . We consider three examples and give convergence and work versus accuracy estimates. The first example deals with the approximation of functions of solutions of parabolic stochastic partial differential equations, the second with the weak semidiscrete approximation of the stochastic heat and wave equation, and in the third, a first order hyperbolic equation is considered.

3.1. Parabolic problem. In this example, we use the framework from [4]. Therefore, consider stochastic processes on a filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ satisfying the “usual conditions” with values in a separable Hilbert space $(U, (\cdot, \cdot)_U)$. 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

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

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

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

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

$$\int_U \|\psi\|_U^2 \mu(d\psi) < +\infty.$$

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, we have the representation $Qe_n = \gamma_n e_n$, where $\gamma_n \geq 0$ is the eigenvalue corresponding to e_n for $n \in \mathbb{N}$. Then, the square root of Q is defined as

$$Q^{1/2}\psi = \sum_n (\psi, e_n)_U \gamma_n^{1/2} e_n,$$

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 $\psi, \phi \in \mathcal{H}$. Let $L_{HS}(\mathcal{H}, H)$ refer to the space of all Hilbert–Schmidt operators from \mathcal{H} to a separable Hilbert space H , and by $\|\cdot\|_{L_{HS}(\mathcal{H}, H)}$ we denote the corresponding norm. By Proposition 8.16 in [21], we have that

$$(3.1) \quad \mathbb{E}[\|\int_0^t \Psi(s) dM(s)\|_H^2] \leq \mathbb{E}[\int_0^t \|\Psi(s)\|_{L_{HS}(\mathcal{H}, H)}^2 ds],$$

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

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

On the separable Hilbert space H , we consider the initial value problem

$$(3.2) \quad dX(t) = (AX(t) + F(X(t))) dt + G(X(t)) dM(t),$$

for $t \in [0, T]$, $T < +\infty$, subject to the initial condition $X(0) = X_0 \in L^2(\Omega; H)$, which is \mathcal{F}_0 -measurable. The operator A with domain $\mathcal{D}(A) \subset H$ is assumed to be the generator of an analytic semigroup S on H . 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 [10]). We assume that A is boundedly invertible on $\mathcal{D}(A)$, and that $(-A)^{-1} : H \rightarrow \mathcal{D}(A)$ is a bounded linear operator. In this section, it is sufficient that A_α exists for $\alpha = 1/2$ and we set $A_{\beta/2} = (A_{1/2})^\beta$, for $\beta \in \mathbb{N}$. We further set $V = \mathcal{D}(A_{1/2})$ and denote by V^* the dual of V . By the Riesz representation theorem, we identify H with its dual and work with the Gelfand triple $V \subset H \cong H^* \subset V^*$ with continuous and dense inclusions. The generator of the semigroup $A : \mathcal{D}(A) \subset H \rightarrow H$ can then be extended to a bounded linear operator $A : V \rightarrow V^*$ via the continuous bilinear form $B_A : V \times V \rightarrow \mathbb{R}$ defined by

$$B_A(\phi, \psi) = \langle A\phi, \psi \rangle_{V^*, V},$$

for $\phi, \psi \in V$. Here, $\langle \cdot, \cdot \rangle_{V^*, V}$ denotes the dual pairing of V and V^* . We set

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

for $\phi \in V$, and define the norm on $L^2(\Omega; V)$ accordingly. Furthermore, by Theorem 6.13 in [20], there exists a constant $C > 0$ such that for all $t \in [0, T]$ and $\phi \in V$

$$(3.3) \quad \|(S(t) - I)\phi\|_H \leq C t^{1/2} \|A_{1/2}\phi\|_H = C t^{1/2} \|\phi\|_V.$$

The operator F maps from H into H and G is a mapping from H into the linear operators from \mathcal{H} into H . We assume that the stochastic process M is in $\mathcal{M}_b^2(U)$. Examples of such processes are given in [4].

Next, we make assumptions such that Equation (3.2) has a mild solution. Therefore, we impose linear growth and Lipschitz conditions on the operators $F : H \rightarrow H$ and $G : H \rightarrow L(U, H)$.

Assumption 3.1. Let $Z = H, V$. Assume that there exist constants $C_1, C_2 > 0$ such that for all $\psi_1, \psi_2 \in Z$ it holds that

$$\begin{aligned} \|F(\psi_1)\|_Z &\leq C_1(1 + \|\psi_1\|_Z), \\ \|F(\psi_1) - F(\psi_2)\|_H &\leq C_1\|\psi_1 - \psi_2\|_H, \end{aligned}$$

and

$$\begin{aligned} \|G(\psi_1)\|_{L_{HS}(\mathcal{H}; Z)} &\leq C_2(1 + \|\psi_1\|_Z), \\ \|G(\psi_1) - G(\psi_2)\|_{L_{HS}(\mathcal{H}; H)} &\leq C_2\|\psi_1 - \psi_2\|_H. \end{aligned}$$

Assumption 3.1 implies that Equation (3.2) has a unique mild solution in H by results in Chapter 9 in [21] and that the predictable process $X : [0, T] \times \Omega \rightarrow H$ is given by

$$(3.4) \quad X(t) = S(t)X_0 + \int_0^t S(t-s)F(X(s)) ds + \int_0^t S(t-s)G(X(s)) dM(s).$$

For further discussions on stochastic differential equations in infinite dimensions, the reader is referred to [8] and [21] and the references therein.

A certain regularity on the initial condition causes the regularity of the mild solution $X = (X(t), t \in [0, T])$ which is specified in the following lemma. This lemma is proven in [4].

Lemma 3.2. *If $\|X_0\|_{L^2(\Omega;V)} < +\infty$, then, the solution X defined in Equation (3.4) is element of $L^2(\Omega;V)$ for all times in $[0, T]$. In particular, for all $t \in [0, T]$ it holds that*

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

This finishes the introduction of the used parabolic framework. Next, we continue with the discretization of Equation (3.4). We use the same approximation scheme as in [4]. For completeness of exposition we include it here.

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

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

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

$$(A_\ell \phi_\ell, \psi_\ell)_H = B_A(\phi_\ell, \psi_\ell),$$

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

$$dX_\ell(t) = (A_\ell X_\ell(t) + P_\ell F(X_\ell(t))) dt + P_\ell G(X_\ell(t)) dM(t),$$

for $t \in (0, T]$, with initial condition $X_\ell(0) = P_\ell X_0$. The semidiscrete mild solution reads

$$X_\ell(t) = S_\ell(t)X_\ell(0) + \int_0^t S_\ell(t-s)P_\ell F(X_\ell(s)) ds + \int_0^t S_\ell(t-s)P_\ell G(X_\ell(s)) dM(s).$$

We shall remark here that we do not approximate the noise. If $U = H$ and V_ℓ contains a finite subset of the eigenbasis of M , the noise is automatically finite dimensional (see e.g. [16]). 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 [2] which properties especially of the eigenvalues of M imply that the overall order of convergence is preserved.

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

$$\Theta^n = \{t_k^n = \frac{T}{2^n} k = \delta t^n k, k = 0, \dots, 2^n\}.$$

Note that the time discretization does not have to be equidistant but that we assume it here for the sake of simplicity of the following analysis. We approximate the semigroup $S_\ell(t_k^n)$ for $t_k^n \in \Theta^n$ by the rational approximation $r(\delta t^n A_\ell)^k$ and assume the following:

Assumption 3.3. For a given finite dimensional space $V_\ell \in \mathcal{V}$ and time discretization $\Theta^n \in \mathbb{T}$, there exists a constant $C > 0$ such that the rational approximation of the semigroup satisfies the error bound

$$\|(S(t_k^n) - r(\delta t^n A_\ell)^k P_\ell)v\|_H \leq C(h_\ell + \sqrt{\delta t^n})\|v\|_V,$$

for $v \in V$ and $k = 0, \dots, n$.

The fully discrete Euler approximation is given for $t_k^n = \delta t^n k \in \Theta^n$ and for $\ell \in \mathbb{N}_0$ by

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

which may be rewritten as

$$\begin{aligned} (3.5) \quad X_{\ell,n}(t_k^n) &= r(\delta t^n A_\ell)^k P_\ell X_0 + \sum_{j=1}^k \int_{t_{j-1}^n}^{t_j^n} r(\delta t^n A_\ell)^{k-j+1} P_\ell F(X_{\ell,n}(t_{j-1}^n)) ds \\ &\quad + \sum_{j=1}^k \int_{t_{j-1}^n}^{t_j^n} r(\delta t^n A_\ell)^{k-j+1} P_\ell G(X_{\ell,n}(t_{j-1}^n)) dM(s). \end{aligned}$$

This approximation converges strongly to the mild solution X , which was shown in Theorem 4.3 in [4]. We use this result next to show $L^2(\Omega; B)$ convergence of $(\varphi(X_{\ell,n}), \ell, n \in \mathbb{N}_0)$ to $\varphi(X)$ in the following proposition, where φ is a Lipschitz function with values in a separable Hilbert space B , i.e., we assume that there exists a constant C such that $\varphi : H \rightarrow B$ satisfies for all $\psi_1, \psi_2 \in H$

$$(3.6) \quad \|\varphi(\psi_1) - \varphi(\psi_2)\|_B \leq C \|\psi_1 - \psi_2\|_H.$$

Proposition 3.4. *If X is the mild solution given in Equation (3.4), $(X_{\ell,n}, \ell, n \in \mathbb{N}_0)$ is the sequence of discrete mild solutions introduced in Equation (3.5), and φ satisfies Equation (3.6), then, for all $\ell, n \in \mathbb{N}_0$, the error is bounded by*

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

Proof. Let $n, \ell \in \mathbb{N}_0$, $t \in \Theta^n$ and let φ satisfy Equation (3.6), then

$$\|\varphi(X(t)) - \varphi(X_{\ell,n}(t))\|_{L^2(\Omega; B)} \leq C \|X(t) - X_{\ell,n}(t)\|_{L^2(\Omega; H)}$$

and the assertion follows with Theorem 4.3 in [4]. \square

Next, we consider the singlelevel Monte Carlo estimator $E_N[\varphi(X_{\ell,n})]$ of the approximate solution $X_{\ell,n}$ and compare it to $\mathbb{E}[\varphi(X_{\ell,n})]$, which is needed in the subsequent proofs.

Remark 3.5. For $n, \ell \in \mathbb{N}_0$ and for $t \in \Theta^n$, Lemma 2.1 implies that

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

Furthermore, we use the properties of φ to derive that

$$\|\varphi(X_{\ell,n}(t))\|_{L^2(\Omega; B)} \leq C_0(1 + \|X_{\ell,n}(t_k^n)\|_{L^2(\Omega; H)}) \leq C(T)(1 + \|X_0\|_{L^2(\Omega; H)}),$$

where C_0 denotes the linear growth constant of φ , which is induced by the global Lipschitz constant, and the last step was proven in [4]. This estimate implies that

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

The previous two results enable us to give an error bound on the approximation of the expectation by a singlelevel Monte Carlo method.

Corollary 3.6. *If X is the mild solution given in Equation (3.4), $(X_{\ell,n}, \ell, n \in \mathbb{N}_0)$ is the sequence of discrete mild solutions introduced in Equation (3.5), and φ satisfies Equation (3.6), then, for all $\ell, n \in \mathbb{N}_0$, the error is bounded by*

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

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

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

The first term on the right hand side is bounded by Proposition 3.4. The assertion follows with Lemma 2.1 and Remark 3.5 for the second term. \square

One should mention here that if bounds on $\|\mathbb{E}[\varphi(X(t))] - \mathbb{E}[\varphi(X_{\ell,n}(t))]\|_{L^2(\Omega; B)}$ are available, these should have been used in the proof instead of the strong errors.

Corollary 3.6 raises the question of how to choose the space discretization, the time approximation, and the number of samples to minimize the overall error and the overall work at once. If we choose

$$(3.7) \quad \delta t^n \simeq h_\ell^2$$

and set $h_\ell \simeq 2^{-\ell}$, the errors are balanced when $n = 2\ell$. Here, the notation $\delta t^n \simeq h_\ell^2$ denotes the abbreviation of the statement $\delta t^n = O(h_\ell^2)$ and $h_\ell^2 = O(\delta t^n)$. With the shown convergence rate in Corollary 3.6, it can easily be seen that we equilibrate the discretization and the 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 for some parameter $d \in \mathbb{N}$. If $H = L^2(D)$ over a domain D , the parameter $d = \dim(D)$ refers to the dimension of the problem. Then, the overall work \mathcal{W}_ℓ is given by

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

and the error bound in Corollary 3.6 in terms of the overall computational work reads

$$\sup_{t \in \Theta^n} \|\mathbb{E}[\varphi(X(t))] - E_N[\varphi(X_{\ell,n}(t))]\|_{L^2(\Omega; B)} \leq C(T) h_\ell \simeq (\mathcal{W}_\ell)^{-1/(d+4)}.$$

With the knowledge of work versus accuracy for the singlelevel Monte Carlo approximation of the expectation, we continue with the application of the multilevel approach as presented in Theorem 2.3. As we are not aware of any weak convergence rates for our approximation scheme, we use the strong ones presented in Proposition 3.4 and insert them into the second estimate in Lemma 2.2. Here we just cover the case, when we estimate a function of the solution at fixed time $t \in [0, T]$ that is an element of all grids, e.g., at time T . How to interpolate the solutions on coarse levels to the time grid on the finest level is shown in [4] and depends on the chosen approximation scheme.

Corollary 3.7. *For $L \in \mathbb{N}_0$ and $\ell = 0, \dots, L$, set $h_\ell^2 \simeq 2^{-2\ell} \simeq \delta t^{2\ell}$,*

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

for $\ell = 1, \dots, L$ and any $\epsilon > 0$. Then, for fixed $t \in [0, T]$, it holds that

$$(3.8) \quad \|\mathbb{E}[\varphi(X(t))] - E^L[\varphi(X_{L,2L}(t))]\|_{L^2(\Omega;B)} \leq C(\epsilon) h_L (1 + \|X_0\|_{L^2(\Omega;V)}).$$

Furthermore, the computational complexity \mathcal{W}_L for the computation of the multilevel Monte Carlo estimate is bounded by

$$\mathcal{W}_L = O(h_L^{-(2+d)} |\log(h_L)|^{3+2\epsilon}),$$

if it is assumed that the computation in each time step of each sample has linear complexity, i.e., $\mathcal{W}_L^H \simeq h_L^{-d}$ for some $d \in \mathbb{N}$, which includes the subtraction and addition of approximate solutions on different levels ℓ .

Proof. First, we note that by Proposition 3.4 it holds that

$$\|\varphi(X(t)) - \varphi(X_{\ell,2\ell}(t))\|_{L^2(\Omega;B)} \leq C(t)(h_\ell + \sqrt{\delta t^{2\ell}})(1 + \|X_0\|_{L^2(\Omega;V)})$$

and $h_\ell + \sqrt{\delta t^{2\ell}} \simeq 2^{-\ell}$. Then, Lemma 2.2 implies with Lemma 3.2 that

$$\begin{aligned} & \|\mathbb{E}[\varphi(X(t))] - E^L[\varphi(X_{L,2L}(t))]\|_{L^2(\Omega;B)} \\ & \leq C(t)(1 + \|X_0\|_{L^2(\Omega;V)}) \left(h_L + \sqrt{\delta t^{2L}} + \sum_{\ell=0}^L \frac{1}{\sqrt{N_\ell}} (h_\ell + \sqrt{\delta t^{2\ell}} + h_{\ell-1} + \sqrt{\delta t^{2(\ell-1)}}) \right) \end{aligned}$$

with $h_{-1} + \sqrt{\delta t^{-2}} := 1$. In the framework of Theorem 2.3, we have that $\alpha = \beta = 1$. Thus, the errors are equilibrated if $N_0 = 2^{2L}$ and $N_\ell = 2^{2(L-\ell)} \ell^{2(1+\epsilon)}$, for $\epsilon > 0$, which implies Equation (3.8).

To calculate the overall work, we first note, that $\gamma = d+2$ in the framework of Theorem 2.3. Therefore, we have $\gamma \geq 2\beta$ and

$$(\gamma - 2\beta)/\alpha = d.$$

Thus, Theorem 2.3 yields that

$$\mathcal{W}_L = O(h_L^{-(2+d)} |\log(h_L)|^{3+2\epsilon}). \quad \square$$

This corollary shows that the work to estimate $\mathbb{E}[\varphi(X(t))]$ reduces from $O(h_L^{-(4+d)})$ with the singlelevel Monte Carlo method to $O(h_L^{-(2+d)} |\log(h_L)|^{3+2\epsilon})$, when the multilevel Monte Carlo method is applied.

3.2. Semidiscrete approximation of the heat and wave equation. In [17], the authors present weak convergence results of a semidiscrete approximation for a heat and a wave equation and show that the rate of weak convergence is essentially twice the rate of strong convergence. In this section, we use these results to calculate the overall work of a multilevel Monte Carlo estimator and show that the knowledge of a faster weak than strong convergence decreases the number of necessary samples N_ℓ on each level $\ell = 0, \dots, L$ to preserve the order of convergence and therefore the overall work.

Let $D \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be a bounded domain and consider on $H = L^2(D)$ the heat equation

$$(3.9) \quad dX(t) = \Delta X(t) dt + dW(t)$$

with initial condition $X(0) = X_0$ driven by a Q -Wiener process $W = (W(t), t \in [0, T])$ with $Q \in L_1^+(H)$. Here Δ denotes the Laplace operator with Dirichlet boundary conditions.

Let us discretize D by a sequence of triangulations $(\mathcal{T}_\ell, \ell \in \mathbb{N}_0)$ defined over finite numbers of points. Then, let $(\mathcal{S}_\ell, \ell \in \mathbb{N}_0)$ denote a corresponding family of Finite Element spaces

consisting of piecewise, continuous polynomials of degree $\leq r - 1$ such that $\mathcal{S}_\ell \rightarrow H$ for $\ell \rightarrow +\infty$ and furthermore, $\mathcal{S}_\ell \subset H_0^1(D)$ for $\ell \in \mathbb{N}_0$. We denote by Δ_ℓ the discrete Laplacian and by $P_\ell : H \rightarrow \mathcal{S}_\ell$ the orthogonal projection. Then, the spatially discrete approximation is

$$(3.10) \quad dX_\ell(t) = \Delta_\ell X_\ell(t) dt + P_\ell dW(t)$$

with initial condition $X_\ell(0) = P_\ell X_0$, for $t \in [0, T]$. In [17], it is especially shown that the sequence $(X_\ell, \ell \in \mathbb{N}_0)$ converges weakly of essentially order 2 to X , while it converges strongly with order one, i.e., for a smooth Lipschitz functional $\varphi : H \rightarrow \mathbb{R}$

$$|\mathbb{E}[\varphi(X(t)) - \varphi(X_\ell(t))]| = O(h_\ell^2 \log(h_\ell))$$

and

$$\|X(t) - X_\ell(t)\|_{L^2(\Omega; H)} = O(h_\ell),$$

for $t \in [0, T]$, where the meshwidths $(h_\ell, \ell \in \mathbb{N}_0)$ are given by

$$h_\ell := \max_{K \in \mathcal{T}_\ell} \{\text{diam}(K)\},$$

for $\ell \in \mathbb{N}_0$. Furthermore, we assume that the sequence of triangulations is derived by regular subdivision, which implies that $h_\ell = h_{\ell-1}/2$, for all $\ell \in \mathbb{N}$.

With the same approach it is shown in [17] that under sufficient smoothness conditions, the approximation $X_\ell := (X_\ell^1, X_\ell^2)$ of the stochastic wave equation

$$(3.11) \quad \begin{aligned} dX^1(t) &= X^2(t) dt \\ dX^2(t) &= \Delta X^1(t) dt + dW(t) \end{aligned}$$

with initial conditions $X^1(0) = X_{0,1}$ and $X^2(0) = X_{0,2}$ has similar properties. While in this case the strong order of convergence is $r/(r+1)$, the weak order of convergence is essentially twice as fast, i.e.,

$$|\mathbb{E}[\varphi(X(t)) - \varphi(X_\ell(t))]| = O(h_\ell^{2\frac{r}{r+1}} \log(h_\ell))$$

and

$$\|X(t) - X_\ell(t)\|_{L^2(\Omega; H)} = O(h_\ell^{\frac{r}{r+1}}),$$

for $t \in [0, T]$.

In the following corollary, we show that the overall work of a multilevel Monte Carlo method decreases if weak convergence results are available.

Corollary 3.8. *For the Finite Element approximation (3.10) of the heat equation (3.9), the multilevel Monte Carlo estimator satisfies for $L \in \mathbb{N}_0$ and any $\eta > 0$*

$$\|\mathbb{E}[\varphi(X(t))] - E^L[\varphi(X_L(t))]\|_{L^2(\Omega; H)} = O(h_L^{2-\eta})$$

for an overall work of

$$\mathcal{W}_L = \begin{cases} O(h_L^{-2(2-\eta)}) & \text{if } d = 1, \\ O(h_L^{-2(2-\eta)} |\log(h_L^{2-\eta})|^{3+2\epsilon}) & \text{if } d = 2, \\ O(h_L^{-2(1+d/(2-\eta))} |\log(h_L^{2-\eta})|^{3+2\epsilon}) & \text{if } d > 2. \end{cases}$$

The corresponding multilevel Monte Carlo approximation of the wave equation (3.11) satisfies for $r \geq 2$

$$\|\mathbb{E}[\varphi(X(t))] - E^L[\varphi(X_L(t))]\|_{L^2(\Omega; H)} = O(h_L^{2r/(r+1)-\eta})$$

for an overall work of

$$\mathcal{W}_L = \begin{cases} \mathcal{O}(h_L^{-2(2r/(r+1)-\eta)}) & \text{if } d = 1 \\ \mathcal{O}(h_L^{-2(2r/(r+1)-\eta+d)} |\log(h_L^{2r/(r+1)-\eta})|^{3+2\epsilon}) & \text{if } d > 1. \end{cases}$$

Proof. Let us first consider the heat equation. Then, Lemma 2.2 and the properties of φ lead to

$$\begin{aligned} & \|\mathbb{E}[\varphi(X(t))] - E^L[\varphi(X_L(t))]\|_{L^2(\Omega;H)} \\ & \leq \|\mathbb{E}[\varphi(X(t)) - \varphi(X_L(t))]\| + \frac{1}{\sqrt{N_0}} \|X_0(t)\|_{L^2(\Omega;H)} \\ & \quad + \sum_{\ell=1}^L \frac{1}{\sqrt{N_\ell}} (\|X(t) - X_\ell(t)\|_{L^2(\Omega;H)} + \|X(t) - X_{\ell-1}(t)\|_{L^2(\Omega;H)}) \\ & \leq C \left(h_L^2 \log(h_L) + \frac{1}{\sqrt{N_0}} + \sum_{\ell=1}^L \frac{1}{\sqrt{N_\ell}} h_\ell \right) \\ & \leq C \left(h_L^{2-\eta} + \frac{1}{\sqrt{N_0}} + \sum_{\ell=1}^L \frac{1}{\sqrt{N_\ell}} h_\ell \right) \end{aligned}$$

for fixed $L \in \mathbb{N}$ and $\eta > 0$. In the framework of Theorem 2.3, we have that $\alpha = 2 - \eta$ and $\beta = 1$. Therefore, the errors are equilibrated if we set $N_0 = 2^{2(2-\eta)L}$ and $N_\ell = 2^{2((2-\eta)L-\ell)}$, for $\ell = 1, \dots, L$. Then, the overall error is by Theorem 2.3 bounded by

$$\|\mathbb{E}[\varphi(X(t))] - E^L[\varphi(X_L(t))]\|_{L^2(\Omega;H)} = \mathcal{O}(h_L^{2-\eta}),$$

and the overall work is by Theorem 2.3 with $\gamma = d$ bounded by

$$\mathcal{W}_L = \begin{cases} \mathcal{O}(h_L^{-2(2-\eta)}) & \text{if } d = 1, \\ \mathcal{O}(h_L^{-2(2-\eta)} |\log(h_L^{2-\eta})|^{3+2\epsilon}) & \text{if } d = 2, \\ \mathcal{O}(h_L^{-2(1+d/(2-\eta))} |\log(h_L^{2-\eta})|^{3+2\epsilon}) & \text{if } d > 2. \end{cases}$$

Similarly, we get the result for the wave equation with $\alpha = 2r/(r+1) - \eta$, $\beta = r/(r+1)$, and $\gamma = d$. \square

If we use the convergence rates of the strong error estimates for the heat equation, Theorem 2.3 would imply the same amount of work for $d = 1, 2$ but for $d > 2$

$$\mathcal{W}_L = \mathcal{O}(h_L^{-(2-\eta)d} |\log(h_L^{2-\eta})|^{3+2\epsilon})$$

of work to achieve the same order of convergence, which is worse. For example for $d = 3$, the exponent of the estimate using the weak error bound is $5 + 3\eta/(2 - \eta)$, while the use of the strong error estimates leads to $6 - 3\eta$, which is worse for small $\eta > 0$.

3.3. First order hyperbolic problem. In this example, we consider the framework of Section 4 in [1] and use the notation of Section 3.1. Let $D \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be a bounded domain with smooth boundary ∂D and assume on the Hilbert space $H = L^2(D)$ the first order hyperbolic problem

$$(3.12) \quad dX(t) = BX(t) dt + G(X(t)) dM(t)$$

with initial condition $X(0) = X_0$, where $M \in \mathcal{M}_b^2(U)$, $G : H \rightarrow L(U, H)$ is linear and B is a first order differential operator that is the generator of a C_0 -semigroup of contractions

$S = (S(t), t \in [0, T])$. For a given vector field b , the first order differential operator B is defined as

$$B\phi(x) := \sum_{i=1}^d b_i(x) \partial_i \phi(x),$$

for $x \in D$ and $\phi \in C_c^1(D)$, where ∂_i denotes the derivative in the i -th direction. The inflow boundary is the set

$$\partial D^- := \{x \in \partial D, b(x) \cdot \mathbf{n}(x) > 0\},$$

where $\mathbf{n}(x)$ denotes the exterior normal to ∂D at x . For convenience, we impose Dirichlet boundary conditions $X(t) = 0$ on the inflow boundary ∂D^- , for all $t \in [0, T]$. This particular structure has to be taken into consideration when defining the finite dimensional spaces for the approximation of the stochastic partial differential equation (3.12). Let $(\mathcal{S}_\ell^-, \ell \in \mathbb{N}_0)$ be a family of Finite Element spaces in $H^1(D)$ consisting of piecewise linear, continuous polynomials with respect to a family of triangulations $(\mathcal{T}_\ell, \ell \in \mathbb{N}_0)$, which vanish on the inflow boundary ∂D^- . We define the bilinear form $B_B : H^1(D) \times H \rightarrow \mathbb{R}$ by

$$B_B(\phi, \psi) := (B\phi, \psi)_H,$$

for all $\phi \in H^1(D)$ and $\psi \in H$. We approximate the solution X of Equation (3.12) by the linearized backward Euler scheme as introduced in Section 3 in [1]. The fully discrete problem is to find $X_{\ell,n}$ for $\ell, n \in \mathbb{N}_0$ such that for all time discretization points $t_k^n \in \Theta^n$

$$(3.13) \quad (X_{\ell,n}(t_k^n), \phi)_H = (X_0, \phi)_H + \delta t^n \sum_{i=1}^k B_B(X_{\ell,n}(t_{i-1}^n), \phi)_H + \sum_{i=1}^k \int_{t_{i-1}^n}^{t_i^n} (G^*(X_{\ell,n}(t_{i-1}^n) \phi), dM(s))_{\mathcal{H}},$$

where G^* denotes the adjoint of G . This approximation converges by Theorem 4.3 in [1] with

$$\|X(t) - X_{\ell,n}(t)\|_{L^2(\Omega; H)} = O(h_\ell + \sqrt{\delta t^n}),$$

when we impose sufficient smoothness on the equation, where $(h_\ell, \ell \in \mathbb{N}_0)$ denotes the sequence of meshwidths as introduced in Section 3.2. If we plug this estimate into Theorem 2.3, we get the following corollary. As the convergence is similar to the parabolic case in Section 3.1, the corollary is similar to Corollary 3.7 and the proof is therefore omitted.

Corollary 3.9. *Let Equation (3.13) define the approximation to the hyperbolic problem (3.12). For $L \in \mathbb{N}_0$ and $\ell = 0, \dots, L$, set $h_\ell^2 \simeq 2^{-2\ell} \simeq \delta t^{2\ell}$,*

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

for $\ell = 1, \dots, L$ and any $\epsilon > 0$. Then, for fixed $t \in [0, T]$, it holds that

$$\|\mathbb{E}[\varphi(X(t))] - E^L[\varphi(X_{L,2L}(t))]\|_{L^2(\Omega; B)} \leq C(\epsilon) h_L (1 + \|X_0\|_{L^2(\Omega; V)}).$$

Furthermore, the computational complexity \mathcal{W}_L for the computation of the multilevel Monte Carlo estimate is bounded by

$$\mathcal{W}_L = O(h_L^{-(2+d)} |\log(h_L)|^{3+2\epsilon}),$$

if it is assumed that the computation in each time step of each sample has linear complexity, i.e., $\mathcal{W}_L^H \simeq h_\ell^{-d}$ for some $d = \dim(D)$, which includes the subtraction and addition of approximate solutions on different levels ℓ .

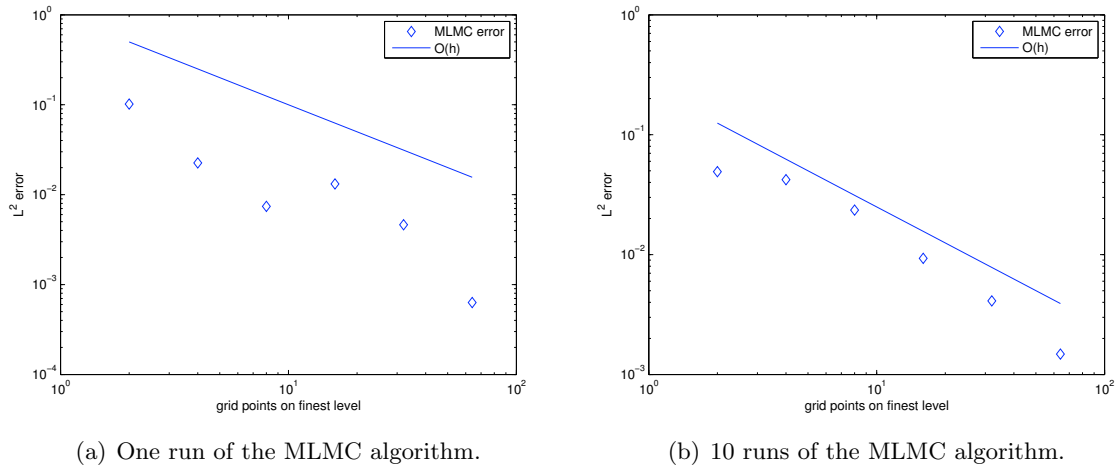


FIGURE 1. Mean square error of the multilevel Monte Carlo estimator of the heat equation driven by additive Wiener noise according to the convergence of an Euler scheme.

4. SIMULATIONS

In this section, some simulation results of the theory of the previous sections are shown. First, we reproduce the error bounds for a parabolic problem, then for a hyperbolic one.

4.1. Parabolic problem. We simulate similarly to [3] the heat equation driven by additive Wiener noise

$$dX(t) = \Delta X(t) dt + dW(t)$$

on the space interval $(0, 1)$ and the time interval $[0, 1]$ with initial condition $X(0, x) = \sin(\pi x)$ for $x \in (0, 1)$. The covariance kernel C_Q of the Q -Wiener process W is given by

$$C_Q(x, y) = \exp(-2(x - y)^2),$$

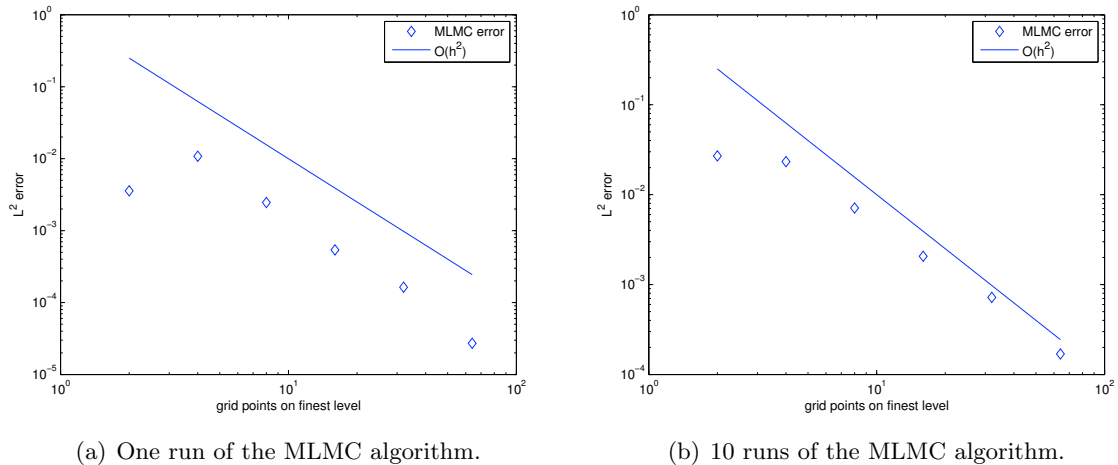
for $x, y \in (0, 1)$, and W is constructed of independent, real-valued Wiener processes $(W_i, i \in \mathbb{N})$. The solution to the corresponding deterministic system with $u(t) = \mathbb{E}[X(t)]$ for $t \in [0, 1]$

$$du(t) = \Delta u(t) dt$$

is in this case $u(t, x) = \exp(-\pi^2 t) \sin(\pi x)$, for $x \in (0, 1)$ and $t \in [0, 1]$.

The space discretization is done with a Finite Element method and the hat function basis, i.e., with the spaces $(\mathcal{S}_h, h > 0)$ of piecewise linear, continuous polynomials, see e.g., Section 3.1 in [4]. We use a Crank–Nicolson method for the time stepping and truncate the Karhunen–Loève expansion of the Wiener process according to Lemma 3.1 in [3] to be able to do simulations. The number of multilevel Monte Carlo samples is calculated according to Section 3.1. In Figure 1(a), the multilevel Monte Carlo estimator $E^L[X_{L,2L}(1)]$ was calculated for $L = 1, \dots, 6$, i.e., we chose φ to be the identity. The plot shows the approximation of

$$\|\mathbb{E}[X(1)] - E^L[X_{L,2L}(1)]\|_H = \left(\int_0^1 (\exp(-\pi^2) \sin(\pi x) - E^L[X_{L,2L}(1, x)])^2 dx \right)^{1/2},$$



(a) One run of the MLMC algorithm.

(b) 10 runs of the MLMC algorithm.

FIGURE 2. Mean square error of the multilevel Monte Carlo estimator of the heat equation driven by additive Wiener noise according to the convergence of a Milstein scheme.

i.e.,

$$e_1(X_{L,2L}) := \left(\frac{1}{m} \sum_{k=1}^m (\exp(-\pi^2) \sin(\pi x_k) - E^L[X_{L,2L}(1, x_k)])^2 \right)^{1/2}.$$

Here, for all levels $L = 1, \dots, 6$, $m = 2^6 + 1$ and x_k , $k = 1, \dots, m$, are the nodal points of the finest discretization, i.e., on level 6. The multilevel Monte Carlo estimator $E^L[X_{L,2L}]$ is calculated at these points by its basis representation, for $L = 1, \dots, 5$, which is equal to the linear interpolation to all grid points x_k , $k = 1, \dots, m$. One observes the convergence of one multilevel Monte Carlo estimator, i.e., the almost sure convergence of the method, which can be shown using the mean square convergence and the Borel–Cantelli lemma. In Figure 1(b), the error is estimated according to Corollary 3.7. For the estimation of the $L^2(\Omega; H)$ norm we chose

$$e_N(X_{L,2L}) := \left(\frac{1}{N} \sum_{i=1}^N e_1(X_{L,2L}^i)^2 \right)^{1/2},$$

where $(X_{L,2L}^i, i = 1, \dots, N)$ is a sequence of independent, identically distributed samples of $X_{L,2L}$ and $N = 10$. The simulation results confirm the theory.

Furthermore, it is known that the used approximation scheme for a heat equation with additive noise converges better than presented in Section 3.1. In this case the Euler–Maruyama scheme is equal to the Milstein scheme and converges strongly with order δt^n in time and h_ℓ^2 in space. In Figure 2, we chose the number of samples according to the convergence of a Milstein scheme as presented in Section 5 in [4]. The convergence of one run of the multilevel Monte Carlo algorithm in Figure 2(a) appears to be stable and the slope is as expected, where the calculated error is e_1 . Figure 2(b) shows the error e_{10} , where the $L^2(\Omega; H)$ error was estimated from 10 samples. The convergence plots verify the theoretical results.

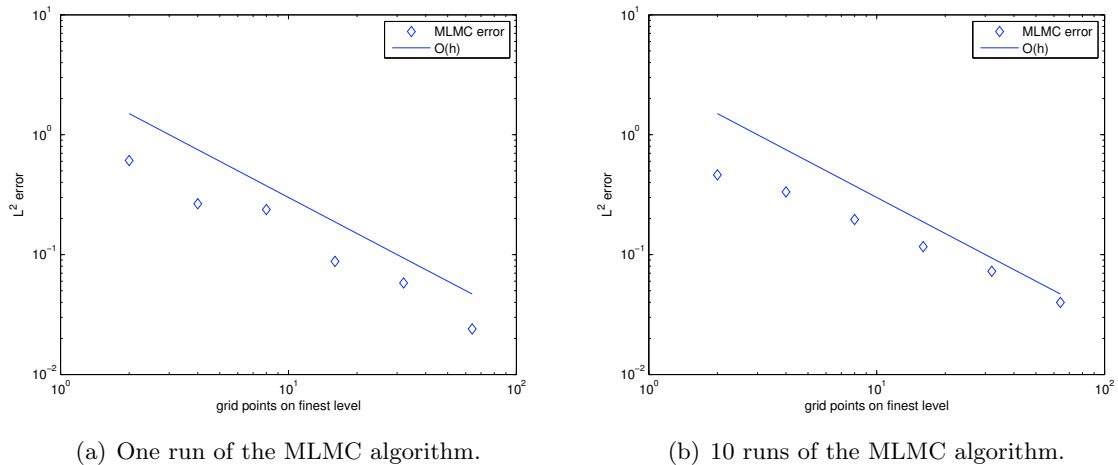


FIGURE 3. Mean square error of the multilevel Monte Carlo estimator of the solution of the advection equation driven by additive Wiener noise.

4.2. First order hyperbolic problem. We simulate the linear advection equation driven by additive Wiener noise

$$dX(t) = \nabla X(t) dt + dW(t)$$

on the space interval $(0, 1)$ and the time interval $[0, 1]$ with initial condition $X(0, x) = \sin(\pi x)$, for $x \in (0, 1)$, and inflow boundary condition $X(t, 1) = -\sin(\pi t)$, for $t \in [0, 1]$. The covariance kernel C_Q of the Q -Wiener process W is given by

$$C_Q(x, y) = \exp(-10(x - y)^2),$$

for $x, y \in (0, 1)$, and W is constructed of independent, real-valued Wiener processes $(W_i, i \in \mathbb{N})$. The solution to the corresponding deterministic system with $u(t) = \mathbb{E}[X(t)]$, for $t \in [0, 1]$,

$$du(t) = \nabla u(t) dt$$

is in this case $u(t, x) = \sin(\pi(x + t))$, for $x \in (0, 1)$ and $t \in [0, 1]$.

The space discretization is done with a first order SUPG method as introduced in [6]. We use a Crank–Nicolson method for the time stepping and truncate the Karhunen–Loève expansion of the Wiener process according to Lemma 3.1 in [3] as in the parabolic case. The number of multilevel Monte Carlo samples is calculated according to Section 3.3. Note, that we used for this calculation a convergence result based on a Galerkin method. Since the Galerkin approximation is for a first order hyperbolic equation only asymptotically stable, we use for the simulation a (stabilized) SUPG scheme. However, as the SUPG approximation converges slightly better, the number of multilevel Monte Carlo samples calculated in Section 3.3 is too little. This means that in our simulation the error is dominated by the Monte Carlo error.

As in the parabolic case, Figure 3(a) shows the mean square error of the multilevel Monte Carlo estimator $E^L[X_{L,2L}(1)]$ for $L = 1, \dots, 6$. The plot shows the approximation of

$$\|\mathbb{E}[X(1)] - E^L[X_{L,2L}(1)]\|_H = \left(\int_0^1 (\sin(\pi(x+1)) - E^L[X_{L,2L}(1, x)])^2 dx \right)^{1/2},$$

i.e.,

$$e_1(X_{L,2L}) := \left(\frac{1}{m} \sum_{k=1}^m (\sin(\pi(x_k + 1)) - E^L[X_{L,2L}(1, x_k)])^2 \right)^{1/2}.$$

Here, as in the parabolic simulation, for all levels $L = 1, \dots, 6$, $m = 2^6 + 1$ and x_k , $k = 1, \dots, m$, are the nodal points of the finest discretization, i.e., on level 6. The multilevel Monte Carlo estimator $E^L[X_{L,2L}]$, for $L = 1, \dots, 5$, is again equal to the linear interpolation to all grid points x_k , $k = 1, \dots, m$. We see the convergence of one multilevel Monte Carlo estimator in dependence of the number of grid points on the finest grid. One sample for the estimation of the mean of the multilevel Monte Carlo estimator might, as in the parabolic case, not be sufficient. Therefore, in Figure 3(b) the error $e_N(X_{L,2L})$, for $N = 10$, is plotted. As before, for the estimation of the $L^2(\Omega; H)$ norm, we chose

$$e_N(X_{L,2L}) := \left(\frac{1}{N} \sum_{i=1}^N e_1(X_{L,2L}^i)^2 \right)^{1/2}.$$

The convergence plot verifies the theoretical findings.

REFERENCES

- [1] A. BARTH, *A Finite Element Method for martingale-driven stochastic partial differential equations*, Comm. Stoch. Anal., 4 (2010), pp. 355–375.
- [2] A. BARTH AND A. LANG, *Milstein approximation for advection-diffusion equations driven by multiplicative noncontinuous martingale noises*. SAM-Report 2011-36, 2011.
- [3] ———, *Simulation of stochastic partial differential equations using finite element methods*, Stochastics, 84 (2012), pp. 217–231.
- [4] A. BARTH, A. LANG, AND C. SCHWAB, *Multilevel Monte Carlo Finite Element method for parabolic stochastic partial differential equations*. SAM-Report 2011-30, May 2011.
- [5] A. BARTH, C. SCHWAB, AND N. ZOLLINGER, *Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients*, Numerische Mathematik, 119 (2011), pp. 123–161.
- [6] P. BOCHEV, M. GUNZBURGER, AND J. SHADID, *Stability of the SUPG finite element method for transient advection-diffusion problems.*, Comp. Meth. Appl. Mech. Engrg., 193 (2004), pp. 2301–2323.
- [7] K. CLIFFE, M. GILES, R. SCHEICHL, AND A. TECKENTRUP, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Computing and Visualization in Science, 14 (2011), pp. 3–15.
- [8] G. DA PRATO AND J. ZABCZYK, *Stochastic Equations in Infinite Dimensions*, vol. 44 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 1992.
- [9] S. DEREICH AND F. HEIDENREICH, *A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations*, Stochastic Processes and their Applications, 121 (2011), pp. 1565–1587.
- [10] K.-J. ENGEL AND R. NAGEL, *One-Parameter Semigroups for Linear Evolution Equations*, Graduate Texts in Mathematics. 194. Berlin: Springer, 2000.
- [11] M. B. GILES, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.
- [12] M. B. GILES AND C. REISINGER, *Stochastic finite differences and multilevel Monte Carlo for a class of SPDEs in finance*, SIAM J. Fin. Math., (2012).
- [13] S. HEINRICH, *Multilevel Monte Carlo methods*, in Large-Scale Scientific Computing, Third International Conference, LSSC 2001, Sozopol, Bulgaria, June 6-10, 2001, Revised Papers, S. Margenov, J. Wasniewski, and P. Y. Yalamov, eds., vol. 2179 of Lecture Notes in Computer Science, Springer, 2001, pp. 58–67.
- [14] H. HOEL, E. VON SCHWERIN, A. SZEPESY, AND R. TEMPONE, *Adaptive multilevel Monte Carlo simulation*, in Numerical Analysis of Multiscale Computations, vol. 82 of Lecture Notes in Computational Science and Engineering, 2012, pp. 217–234.
- [15] M. HUTZENTHALER, A. JENTZEN, AND P. E. KLOEDEN, *Divergence of the multilevel Monte Carlo method*. arXiv: math.PR.1105.0226, May 2011.

- [16] P. KLOEDEN, G. LORD, A. NEUENKIRCH, AND T. SHARDLOW, *The exponential integrator scheme for stochastic partial differential equations: Pathwise error bounds*, J. Comput. Appl. Math., 235 (2011), pp. 1245–1260.
- [17] M. KOVÁCS, S. LARSSON, AND F. LINDGREN, *Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise*, BIT Num. Math, 52 (2012), pp. 85–108.
- [18] H. MARXEN, *The multilevel Monte Carlo method used on a Lévy driven SDE*, Monte Carlo Methods and Applications, 16 (2010), pp. 167–190.
- [19] S. MISHRA AND C. SCHWAB, *Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data*, Math. Comp., (2012).
- [20] A. PAZY, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Applied Mathematical Sciences, 44. New York – Berlin – Heidelberg – Tokyo: Springer, 1983.
- [21] S. PESZAT AND J. ZABCZYK, *Stochastic Partial Differential Equations with Lévy Noise*, vol. 113 of Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2007.

(Andrea Barth)

ETH, SEMINAR FÜR ANGEWANDTE MATHEMATIK
RÄMISTRASSE 101
8092 ZÜRICH

E-mail address: `andrea.barth@math.ethz.ch`

(Annika Lang)

ETH, SEMINAR FÜR ANGEWANDTE MATHEMATIK
RÄMISTRASSE 101
8092 ZÜRICH

E-mail address: `annika.lang@math.ethz.ch`

Research Reports

| No. | Authors/Title |
|-------|---|
| 11-68 | <i>A. Barth and A. Lang</i> Multilevel Monte Carlo method with applications to stochastic partial differential equations |
| 11-67 | <i>C. Effenberger and D. Kressner</i> Chebyshev interpolation for nonlinear eigenvalue problems |
| 11-66 | <i>R. Guberovic, Ch. Schwab and R. Stevenson</i> Space-time variational saddle point formulations of Stokes and Navier-Stokes equations |
| 11-65 | <i>J. Li, H. Liu and H. Sun</i> Enhanced approximate cloaking by SH and FSH lining |
| 11-64 | <i>M. Hansen and Ch. Schwab</i> Analytic regularity and best N -term approximation of high dimensional parametric initial value problems |
| 11-63 | <i>R. Hiptmair, G. Phillips and G. Sinha</i> Multiple point evaluation on combined tensor product supports |
| 11-62 | <i>J. Li, M. Li and S. Mao</i> Convergence analysis of an adaptive finite element method for distributed flux reconstruction |
| 11-61 | <i>J. Li, M. Li and S. Mao</i> A priori error estimates of a finite element method for distributed flux reconstruction |
| 11-60 | <i>H. Heumann and R. Hiptmair</i> Refined convergence theory for semi-Lagrangian schemes for pure advection |
| 11-59 | <i>V.A. Hoang and Ch. Schwab</i> N -term Galerkin Wiener chaos approximations of elliptic PDEs with lognormal Gaussian random inputs |
| 11-58 | <i>X. Claeys and R. Hiptmair</i> Electromagnetic scattering at composite objects: A novel multi-trace boundary integral formulation |
| 11-57 | <i>S. Mishra and L.V. Spinolo</i> Accurate numerical schemes for approximating initial-boundary value problems for systems of conservation law |
| 11-56 | <i>P. Grohs</i> Finite elements of arbitrary order and quasiinterpolation for data in Riemannian manifolds |