Thesis for the Degree of Doctor of Philosophy

Approximating Stochastic Partial Differential Equations with Finite Elements: Computation and Analysis

Andreas Petersson

Division of Applied Mathematics and Statistics
Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
Gothenburg, Sweden, 2019

Approximating Stochastic Partial Differential Equations with Finite Elements: Computation and Analysis

Andreas Petersson

ISBN: 978-91-7905-206-5

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4673
ISSN: 0346-718X

Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg, Sweden
Phone: +46 (0)31 772 10 00

Author e-mail: `andreas.petersson@chalmers.se`

Cover: *Upper left:* The stochastic heat equation (Example 3.4). *Upper right:* The stochastic wave equation (Example 3.3). *Lower left:* A function connected to the quadratic behaviour of the stochastic heat equation at a fixed time (see summary of Paper 5). *Lower right:* The covariance function of the stochastic wave equation at a fixed time (see summary of Paper 3).

Typeset with LaTeX.
Printed in Gothenburg by Chalmers Reproservice, Sweden 2019.

Approximating Stochastic Partial Differential Equations with Finite Elements: Computation and Analysis

Andreas Petersson

Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg

Abstract

Stochastic partial differential equations (SPDE) must be approximated in space and time to allow for the simulation of their solutions. In this thesis fully discrete approximations of such equations are considered, with an emphasis on finite element methods combined with rational semigroup approximations.

A quantity of interest for SPDE simulations often takes the form of an expected value of a functional applied to the solution. This is the major theme of this thesis, which divides into two minor themes. The first is how to analyze the error resulting from the fully discrete approximation of an SPDE with respect to a given functional, which is referred to as the weak error of the approximation. The second is how to efficiently compute the quantity of interest as well as the weak error itself. The Monte Carlo (MC) and multilevel Monte Carlo (MLMC) methods are common approaches for this.

The thesis consists of five papers. In the first paper the additional error caused by MC and MLMC methods in simulations of the weak error is analyzed. Upper and lower bounds are derived for the different methods and simulations illustrate the results. The second paper sets up a framework for the analysis of the asymptotic mean square stability, the stability as measured in a quadratic functional, of a general stochastic recursion scheme, which is applied to several discretizations of an SPDE. In the third paper, a novel technique for efficiently generating samples of SPDE approximations is introduced, based on the computation of discrete covariance operators. The computational complexities of the resulting MC and MLMC methods are analyzed. The fourth paper considers the analysis of the weak error for the approximation of the semilinear stochastic wave equation. In the fifth paper, a Lyapunov equation is derived, which allows for the deterministic approximation of the expected value of a quadratic functional applied to the solution of an SPDE. The paper also includes an error analysis of an approximation of this equation and an analysis of the weak error, with respect to the quadratic functional, of an approximation of the considered SPDE.

Keywords: Stochastic partial differential equations, numerical approximation, finite element method, Monte Carlo, multilevel Monte Carlo, Lévy process, weak convergence, asymptotic mean square stability, multiplicative noise, covariance operator, stochastic heat equation, stochastic wave equation, Lyapunov equation, white noise, generalized Wiener process

List of included papers

Paper 1 A. Lang and A. Petersson. Monte Carlo versus multilevel Monte Carlo in weak error simulations of SPDE approximations. *Math. Comput. Simulation*, 143:99 – 113, 2018

Paper 2 A. Lang, A. Petersson, and A. Thalhammer. Mean-square stability analysis of approximations of stochastic differential equations in infinite dimensions. *BIT Numer. Math.*, 57(4):963–990, 2017

Paper 3 A. Petersson. Rapid covariance-based sampling of linear SPDE approximations in the multilevel Monte Carlo method. To appear in *Monte Carlo and Quasi-Monte Carlo Methods, MCQMC, Rennes, France, July 2018*. Preprint at arXiv:1806.11523v3, 2019

Paper 4 M. Kovács, A. Lang, and A. Petersson. Weak convergence of fully discrete finite element approximations of semilinear hyperbolic SPDE with additive noise. Submitted. Preprint at arXiv:1909.04571, 2019

Paper 5 A. Andersson, A. Lang, A. Petersson, and L. Schroer. Finite element approximation of Lyapunov equations for the computation of quadratic functionals of SPDE. Preprint at arXiv:1910.05261, 2019

My contributions to the included papers

Paper 1 I wrote the code for the SPDE approximation, performed the cluster computations and contributed to the development of the theory. A significant part of the manuscript was written by me.

Paper 2 I significantly contributed to the theoretical work, the writing and the simulations.

Paper 3 Annika Lang gave valuable input throughout the writing of the paper.

Paper 4 I wrote the manuscript, performed the simulations and produced the major part of the theory.

Paper 5 I significantly contributed to the theoretical work, the writing and the simulations.

Acknowledgements

First of all I would like to thank my supervisor Annika Lang for introducing me to this area of research way back when I started writing my master's thesis. To be honest, I ended up working with this field mainly because I thought that a great teacher makes for a great supervisor. After all these years of research that turned out to be both frustrating and fascinating, I have to commend 24-year-old-me on my great judgment. Thank you, Annika, for all the support and patience, both professional and emotional, that went well beyond what can be expected from a supervisor.

Next, I want to thank my co-authors Andreas Thalhammer, Mihály Kovács and Adam Andersson and my co-supervisor Stig Larsson. The best research in this thesis was accomplished in close collaboration with others, and I hope to have the opportunity to work with so many kind and talented people again. Thanks also to David Cohen for making my licentiate seminar a pleasant experience and being a great guide to Beijing and to Raphael Kruse for giving me the benefit of having a great doctoral thesis on SPDE to read for inspiration from day one.

The administrative staff (Fia, Lotta, Marie and everyone else) deserve a great deal gratitude for all the work you have done for me, particularly for all the help I got when I ended up being responsible for teaching a course on my own. Thank you for still responding to me even when I asked the same question for the fourth time. Along with all of my fellow PhD students and other colleagues at Mathematical Sciences, you help create a friendly working environment that facilitates the achievement of great things.

Thanks also to Vetenskapsrådet, and by extension all Swedish taxpayers, for the financial support, and to all of the students who actually attended the lectures and exercise sessions given by me. Food comes first, then research.

Finally, I am incredibly grateful to all of my friends and family for supporting me through each step of this process. In particular I want to thank my parents for taking care of my cats during stressful times and my cats for being cats. But most of all I am grateful for the help and support I have received from my partner Andrea. You know I could not have done this without you.

"Man darf nie an die ganze Straße auf einmal denken, verstehst du? Man muss immer nur an den nächsten Schritt denken, an den nächsten Atemzug, an den nächsten Besenstrich. Und immer wieder nur an den nächsten. [...] Dann macht es Freude; das ist wichtig, dann macht man seine Sache gut. Und so soll es sein."

— *Beppo* in Michael Ende's *Momo*, 1973

Contents

1 Introduction

In the appended five papers, we study some theoretical and computational aspects of approximations to stochastic partial differential equations (SPDE) of the form

$$\begin{aligned} \mathrm{d}X(t) &= (AX(t) + F(t, X(t)))\ \mathrm{d}t + G(t, X(t))\,\mathrm{d}L(t), \\ X(0) &= X_0, \end{aligned} \tag{1}$$

where $t \in (0, T]$, $T < \infty$. Here the process X takes values in a Hilbert space H, the stochastic Lévy process L takes values in another Hilbert space U, F maps elements from $[0, T] \times H$ to H and G maps elements of $[0, T] \times H$ into a space of linear operators from U to H. Here the noise is said to be *multiplicative*, since the operator G depends on X. If it does not, it is said to be *additive*. The operator A is the generator of a strongly continuous semigroup of bounded linear operators on H. An example of such an operator is the Laplacian Δ on the space $H = L^2(\mathcal{D}) = L^2(\mathcal{D}, \mathbb{R})$ of real-valued square integrable functions on some domain $\mathcal{D} \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$, so that (1) becomes a stochastic heat equation, which one can interpret as describing heat flow perturbed by some noise, perhaps due to measurement errors. In papers 3–5 the noise is restricted to the case that $L = W$ is a Wiener process, and in the case that $H = L^2(\mathcal{D})$ we can informally think of $W(t)$, for a fixed $t \in [0, T]$, as a Gaussian random field on $\mathcal{D}$, with W having independent increments in time. SPDE have many uses in fields such as biology, engineering and finance, see, e.g., [21, 22] for an overview of such applications.

As the solution X to (1) is a stochastic process, a natural quantity of interest is the expected value $\mathbb{E}[X(t)]$ of it at some time $t \in [0, T]$, or the expected value $\mathbb{E}[\phi(X(t))]$ of some functional $\phi\colon H \to \mathbb{R}$ of the solution. Since analytical solutions to (1) are hardly ever available, an approximation $\hat{X}(t)$ is used instead and the quantity $|\,\mathbb{E}[\phi(\hat{X}(t))] - \mathbb{E}[\phi(X(t))]|$ is referred to as a weak error. The topic of weak error analysis has been met with increasing interest in the SPDE community during recent years, not least because it is related to the convergence in distribution of $\hat{X}(t)$ to $X(t)$. One of the main topics of this thesis is the analysis of weak errors of finite element approximations. Such approximations are applicable in the example described above – the main idea is to solve (1) in a finite-dimensional subspace $V_h \subset H$ consisting of, e.g., piecewise linear functions on some mesh with mesh size h on $\mathcal{D}$.

In order to approximate quantities like $\mathbb{E}[\phi(X(t))]$, it is not sufficient to discretize X, one also has to approximate the expectation operator $\mathbb{E}[\cdot]$. For this problem, Monte Carlo methods are often employed, which are based on generating a large number of realizations of $\hat{X}$. That is to say, an approximation of the solution to (1) has to be computed many times, which is computationally expensive. This may explain why simulations that illustrate theoretical results on weak convergence are rarely available, as one typically employs Monte Carlo methods to approximate the weak error $|\,\mathbb{E}[\phi(\hat{X}(t))] - \mathbb{E}[\phi(X(t))]|$ itself. Paper 1 is made up of an analysis of the additional error caused by approximating the weak error using various Monte Carlo approaches.

One of the methods considered in Paper 1 is the multilevel Monte Carlo method, which is based on approximating $\mathbb{E}[\phi(X(t))]$ by applying the Monte Carlo method to a sequence

$\hat{X}_0, \hat{X}_1, \ldots, \hat{X}_\ell, \ldots$ of approximations of X indexed by a level ℓ. Typically the accuracy of $\hat{X}_\ell$ increases as $\ell \to \infty$ but so does the computational cost. The main idea of the multilevel Monte Carlo method is to compute a different number of realizations for each level, from a few when ℓ is big to many when ℓ is small. By partitioning the number of realizations over the levels in a (close to) optimal way, the multilevel Monte Carlo method can be made more efficient than standard Monte Carlo methods while retaining the same accuracy. To ensure that it is more efficient, the approximation should be sufficiently stable at all levels. More precisely, this refers to the asymptotic mean square stability of $\hat{X}$, which, for linear approximations, is the property that $\mathbb{E}[\|\hat{X}(t)\|_H^2] \to 0$ as $t \to \infty$. Paper 2 sets up a framework for analyzing the asymptotic mean square stability of numerical approximations to (1), in the setting that F and G are linear operators, as well as more general finite-dimensional recursion schemes.

The framework for analyzing mean square stability is based on analyzing $\mathbb{E}[\hat{X} \otimes \hat{X}]$, the expected value of the tensor product of $\hat{X}$ with itself. While writing Paper 2, we noted that in the context of finite elements, this quantity, or more specifically the *covariance* $\mathsf{Cov}(\hat{X}) = \mathbb{E}[\hat{X} \otimes \hat{X}] - \mathbb{E}[\hat{X}] \otimes \mathbb{E}[\hat{X}]$ of $\hat{X}$ can be computed relatively cheaply. When the noise of (1) is additive and the Lévy process L is a Wiener process in H, the approximation $\hat{X}$ is uniquely determined by its covariance and expected value. Therefore, computing the covariance provides a cheap way to generate samples of $\hat{X}$. In Paper 3 we explain how this is done, and show that the idea can be extended to the multilevel Monte Carlo setting for further computational gains.

In Paper 4, we take a step back and analyze the weak error of fully discrete finite element approximations of (1) in the specific setting of the *stochastic wave equation*, which can, for example, be used for the modeling of DNA strings suspended in a fluid (see [9]). The results are again restricted to the case that G does not depend on X but F is allowed to depend on X in a non-linear way.

In the final paper, we return to the question of mean square stability of $\hat{X}$ by analyzing methods of computing $\mathbb{E}[\|\hat{X}(t)\|_H^2]$. In particular, we analyze finite element discretizations of *Lyapunov equations* related to (1). This allows us to compute $\mathbb{E}[\|\hat{X}(t)\|_H^2]$ without resorting to Monte Carlo methods and to derive weak convergence rates for $\hat{X}$ in a novel way, in the special case that the functional ϕ is quadratic. We compare this method to Monte Carlo methods and a covariance-based method similar to that of Paper 3, and it turns out that this novel method is superior if we are interested in computing $\mathbb{E}[\|\hat{X}(t)\|_H^2]$ for many initial values X_0 in parallel.

The following sections provide the theoretical background for the papers along with summaries of them. In Section 2 we set up our notation and review basic results from the fields of functional analysis and probability theory, with an emphasis on random fields and stochastic integration. Section 3 contains a short introduction to SPDE and the approximations we consider along with a short introduction to Monte Carlo and multilevel Monte Carlo methods applied to SPDE. In Section 4 we summarize the included papers.

The notation used in this introductory section does not always coincide with that of the included papers.

2 Preliminaries

Here we introduce the concepts and notation needed for the construction of solutions and approximations of SPDE. For proofs and more details on the standard claims made in this part of the thesis, the reader is referred to [8, 15, 23].

2.1 Functional analysis

Let $(U, \langle\cdot,\cdot\rangle_U)$ and $(H, \langle\cdot,\cdot\rangle_H)$ be real separable Hilbert spaces and let $(B, \|\cdot\|_B)$ and $(E, \|\cdot\|_B)$ be real Banach spaces. We write $\mathcal{L}(B, E)$ for the Banach space of bounded linear operators from B to E, or $\mathcal{L}(B)$ if $E = B$. Given an orthonormal basis $(e_i)_{i\in\mathbb{N}}$ of H, we denote by $\mathcal{L}_2(U, H) \subseteq \mathcal{L}(U, H)$ the Hilbert space of *Hilbert–Schmidt operators* with inner product

$$\langle F, G\rangle_{\mathcal{L}_2(U,H)} = \sum_{i=1}^{\infty} \langle Fe_i, Ge_i\rangle_H ,$$

and whenever $U = H$ we write $\mathcal{L}_2(H)$ for $\mathcal{L}_2(H, H)$. This inner product is independent of the chosen orthonormal basis $(e_i)_{i\in\mathbb{N}}$ of U. The embedding $\mathcal{L}_2(U, H) \subset \mathcal{L}(U, H)$ is continuous with embedding constant 1, i.e., for $F \in \mathcal{L}_2(U, H)$, $\|F\|_{\mathcal{L}(U,H)} \leq \|F\|_{\mathcal{L}_2(U,H)}$. We shall also have use of operators of *trace class*, i.e., those operators $Q \in \mathcal{L}(H)$ for which

$$\|Q\|_{\mathcal{L}_1(H)} = \inf\left\{ \sum_{i=1}^{\infty} \|a_i\|_H \|b_i\|_H \colon Qv = \sum_{i=1}^{\infty} \langle v, b_i\rangle_H a_i \right\} < \infty.$$

The set of trace class operators $\mathcal{L}_1(H)$ is continuously embedded into $\mathcal{L}_2(H)$. Moreover, if $Q \in \mathcal{L}_1(H)$, the *trace*

$$\operatorname{Tr}(Q) = \sum_{i=1}^{\infty} \langle Qe_i, e_i\rangle_H$$

is well-defined, fulfils the inequality $|\operatorname{Tr}(Q)| \leq \|Q\|_{\mathcal{L}_1(H)}$ and is independent of the specific choice of orthonormal basis $(e_i)_{i\in\mathbb{N}}$ of H. If, in addition to being of trace class, Q is self-adjoint and positive semidefinite, then $Q^{1/2}$ is well-defined as the unique self-adjoint and positive semidefinite operator for which $Q^{1/2}Q^{1/2} = Q$, and in this case $\operatorname{Tr}(Q) = \|Q\|_{\mathcal{L}_1(H)} = \|Q^{1/2}\|^2_{\mathcal{L}_2(H)}$.

We also need the notion of derivatives between Hilbert spaces, which we define in the same way as in [3]. By $\mathcal{C}(U, H)$ we denote the space of continuous mappings from U to H and by $\mathcal{G}^1(U, H) \subset \mathcal{C}(U, H)$ the space of *Gâteaux differentiable mappings* with strongly continuous derivatives, i.e., the space of all continuous mappings $\varphi\colon U \to H$ such that

$$\varphi'(u)v = \lim_{\epsilon\to 0} \frac{1}{\epsilon} \left(\varphi(u + \epsilon v) - \varphi(u)\right)$$

exists as a limit in H for all $u, v \in U$, that $\varphi'(u) \in \mathcal{L}(U, H)$ for all $u \in U$ and that the mapping $U \ni u \mapsto \varphi'(u)v$ is continuous for all $v \in H$. If, in addition, $\varphi' \in \mathcal{C}(U, \mathcal{L}(U, H))$,

then $\varphi \in \mathcal{C}^1(U,H)$, the space of *Fréchet differentiable mappings*. By $\mathcal{G}^2(U,H) \subset \mathcal{G}^1(U,H)$ we denote the space of all mappings $\varphi \in \mathcal{G}^1(U,H)$ such that

$$\varphi''(u)(v,w) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left(\varphi'(u+\epsilon w)v - \varphi'(u)v\right)$$

exists as a limit in H for all $u,v,w \in U$, that $\varphi''(u)\colon U \times U \to H \in \mathcal{L}^{[2]}(U,H)$, the space of all bounded bilinear mappings, for all $u \in U$, that $\varphi''(u)$ is symmetric for all $u \in U$, and that the mapping $U \ni u \mapsto \varphi''(u)(v,w)$ is continuous for all $v,w \in U$. If in addition $\varphi' \in \mathcal{C}(U,\mathcal{L}(U,H))$ and $\varphi'' \in \mathcal{C}(U,\mathcal{L}^{[2]}(U,H))$, then $\varphi \in \mathcal{C}^2(U,H)$, the space of twice Fréchet differentiable mappings. For $n = 1,2$, we denote by $\mathcal{G}^n_{\mathrm{b}}(U,H)$ and $\mathcal{G}^n_{\mathrm{p}}(U,H)$ the sets of all $\varphi \in \mathcal{G}^n(U,H)$ such that all derivatives of φ (but not necessarily φ itself) are bounded and polynomially bounded, respectively, with $\mathcal{C}^n_{\mathrm{b}}(U,H)$ and $\mathcal{C}^n_{\mathrm{p}}(U,H)$ defined analogously. We use the shorthand notations $\mathcal{G}^n(H) = \mathcal{G}^n(H,H)$, $\mathcal{G}^n_{\mathrm{b}}(H) = \mathcal{G}^n_{\mathrm{b}}(H,H)$ and $\mathcal{G}^n_{\mathrm{p}}(H) = \mathcal{G}^n_{\mathrm{p}}(H,H)$, and employ analogous notation for the spaces of Fréchet differentiable mappings. For $\varphi \in \mathcal{G}^1_{\mathrm{p}}(U,H)$ and $u,v \in U$ the *mean value theorem* in H is given by

$$\varphi(u) - \varphi(v) = \int_0^1 \varphi'(v + s(u-v))(u-v)\,\mathrm{d}s.$$

Next, let us assume that $\Lambda\colon \mathrm{dom}(\Lambda) \subseteq H \to H$ is a densely defined, linear, self-adjoint and positive definite operator with compact inverse Λ^{-1}. By the spectral theorem applied to Λ^{-1} we get an orthonormal eigenbasis $(e_i)_{i\in\mathbb{N}}$ of H and a positive sequence $(\lambda_i)_{i\in\mathbb{N}}$ of eigenvalues of Λ that is increasing with $\lim_i \lambda_i = \infty$. For $r \geq 0$ we define fractional powers of Λ by

$$\Lambda^{\frac{r}{2}} v = \sum_{i=1}^{\infty} \lambda_i^{\frac{r}{2}} \left\langle v, e_i \right\rangle_H e_i$$

for

$$v \in \dot{H}^r = \mathrm{dom}(\Lambda^{\frac{r}{2}}) = \left\{ v \in H\colon \|v\|_r^2 = \sum_{i=1}^{\infty} \lambda_i^r \left\langle v, e_i \right\rangle_H^2 < \infty \right\}.$$

We also consider negative powers of Λ. For these we first define, for $r < 0$,

$$\dot{H}^r = \left\{ v = \sum_{j=1}^{\infty} v_j e_j\colon (v_j)_{j\in\mathbb{N}} \subset \mathbb{R} \text{ such that } \sum_{j=1}^{\infty} \lambda_j^r v_j^2 < \infty \right\},$$

and for $v \in \dot{H}^r$, we set

$$\Lambda^{\frac{r}{2}} v = \sum_{j=1}^{\infty} \lambda_j^{\frac{r}{2}} v_j e_j.$$

In this way, $\dot{H}^r$, $r \in \mathbb{R}$, becomes a separable Hilbert space when equipped with the inner product

$$\left\langle \cdot, \cdot \right\rangle_r = \left\langle \Lambda^{\frac{r}{2}} \cdot, \Lambda^{\frac{r}{2}} \cdot \right\rangle_H.$$

Moreover, it holds for $s \leq r$ that $\dot{H}^r \subseteq \dot{H}^s$, where the embedding is dense and continuous, that $\dot{H}^0 = H$ and that for $r \geq 0$, we may identify $\dot{H}^{-r}$ with $(\dot{H}^r)'$, the dual space of $\dot{H}^r$ with respect to the inner product $\langle \cdot, \cdot \rangle_H$, see [15, Appendix B]. We next consider a concrete example of such an operator Λ.

Example 2.1. Let $H = L^2(\mathcal{D})$ be the space of square integrable functions on a bounded convex domain $\mathcal{D} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, with polygonal boundary. Let for a function v on $\mathcal{D}$ the operator Λ be given by

$$\Lambda v = -\nabla \cdot (a(\cdot)\nabla v) + c(\cdot)v$$

with zero Dirichlet boundary conditions, where $a, c \colon \mathcal{D} \to \mathbb{R}$ are sufficiently smooth functions with $c(x) \geq 0$ and $a(x) \geq a_0 > 0$ for all $x \in \mathcal{D}$. Then Λ is a densely defined, linear, self-adjoint and positive definite operator with compact inverse and its fractional powers give rise to a family of Hilbert spaces $(\dot{H}^r)_{r\in\mathbb{R}}$. One can show that $\dot{H}^1 = H_0^1(\mathcal{D})$ and $\dot{H}^2 = H^2(\mathcal{D}) \cap H_0^1(\mathcal{D})$, where $H^k(\mathcal{D})$, $k \geq 0$, denotes the Sobolev space of order k and $H_0^1(\mathcal{D})$ is the subspace of $H^1(\mathcal{D})$ containing the functions that are zero at the boundary of $\mathcal{D}$, with norm equivalence. If $a \equiv 1$ and $c \equiv 0$, i.e., if $\Lambda = -\Delta$, then it also holds that $(\dot{H}^s)_{s\in[0,2]}$ are related to $(H^s)_{s\in[0,2]}$ by

$$\dot{H}^s = \begin{cases} H^s \text{ if } s \in [0, 1/2), \\ \{u \in H^s \colon u = 0 \text{ a.e. on } \partial\mathcal{D}\} \text{ if } s \in (1/2, 3/2) \cup (3/2, 2] \end{cases}$$

with norm equivalence (see, e.g., [28, Theorem 4.5]). Here H^s denotes the Sobolev–Slobodeckij space of order s,.

From the results of [15, Appendix B] we also see that the operator $A = -\Lambda$ is the generator of a *semigroup*, the notion of which we define below.

Definition 2.2. Let H be a Hilbert space. A family $(E(t))_{t\in[0,\infty)}$ with $E(t) \in \mathcal{L}(H)$ for all $t \geq 0$ is called a *semigroup of operators on* H if

(*i*) $E(0) = I$, where I is the identity operator and

(*ii*) $E(t + s) = E(t)E(s)$ for all $s, t \geq 0$.

If in addition to this

(*iii*) $\lim_{t\searrow 0} E(t)v = v$ for all $v \in H$,

it is said to be *strongly continuous* or a C_0*-semigroup*. If it also satisfies

(*iv*) $\|E(t)\|_{\mathcal{L}(H)} \leq 1$ for all $t \geq 0$,

then it is called a C_0*-semigroup of contractions*. The linear operator A defined by

$$Av = \lim_{t\searrow 0} \frac{E(t)v - v}{t},$$

with $\mathrm{dom}(A)$ being the space of all $v \in H$ such that the limit exists, is called the *infinitesimal generator* of the semigroup.

The following two semigroups are considered in the papers of this thesis, corresponding to parabolic and hyperbolic equations, respectively.

Example 2.3. With Λ being a densely defined, linear, self-adjoint and positive definite operator with compact inverse Λ^{-1} on H as described above, $A = -\Lambda$ is the generator of a strongly continuous semigroup E with the representation $E(t)v = \sum_{i=1}^{\infty} e^{-\lambda_i t} \langle v, e_i \rangle_H e_i$ for $t \geq 0$, and $v \in H$. In this case, it is an analytic semigroup, that is, it can be extended to be a complex analytic function $E(z)$ for z in a sector containing the real axis.

Example 2.4. With Λ as in the previous example, let $\mathcal{H}^r = \dot{H}^r \oplus \dot{H}^{r-1}$, $r \in \mathbb{R}$, be the Hilbert space of all vectors $v = [v_1, v_2]$ with $v_1 \in \dot{H}^r$ and $v_2 \in \dot{H}^{r-1}$ with inner product $\langle v, w \rangle_{\mathcal{H}} = \langle v_1, w_1 \rangle_{\dot{H}^r} + \langle v_2, w_2 \rangle_{\dot{H}^{r-1}}$ for $v = [v_1, v_2]^\top, w = [w_1, w_2]^\top \in \mathcal{H}^r$. We use the notation $\mathcal{H}$ for $\mathcal{H}^0$. Then

$$A = \begin{bmatrix} 0 & I \\ -\Lambda & 0 \end{bmatrix}$$

is the generator of a strongly continuous semigroup E on $\mathcal{H}$. In this case, it is in fact a group, i.e., $E(t)$ is well-defined also for $t \leq 0$ with $E(t) = E(-t)^{-1}$.

We close this section with a brief review of tensor products of Hilbert spaces. For Hilbert spaces H and U the *algebraic tensor product* $H \otimes_0 U$ is the vector space of finite sums $\sum_{i=1}^{n} v_i \otimes u_i$, where $v_i \in H$ and $u_i \in U$ for $i = 1, \ldots, n$, along with the equivalence relations

$$(v_1 + v_2) \otimes u_1 = v_1 \otimes u_1 + v_2 \otimes u_1,$$
$$v_1 \otimes (u_1 + u_2) = v_1 \otimes u_1 + v_1 \otimes u_2$$

and

$$(\lambda v_1) \otimes u_1 = v_1 \otimes (\lambda u_1) = \lambda (v_1 \otimes u_1),$$

for $\lambda \in \mathbb{R}$. The *Hilbert tensor product* $H \otimes U$, or just $H^{(2)}$ when $U = H$, is defined as the completion of the algebraic tensor product with respect to the norm induced by the inner product

$$\langle v_1 \otimes u_1, v_2 \otimes u_2 \rangle_{H \otimes U} = \langle v_1, v_2 \rangle_H \langle u_1, u_2 \rangle_U .$$

If $(e_{1,i})_{i \in \mathbb{N}}$ and $(e_{2,i})_{i \in \mathbb{N}}$ are orthonormal bases of H and U, respectively, then $(e_{1,i} \otimes e_{2,j})_{i,j \in \mathbb{N}}$ is an orthonormal basis of $H \otimes U$. The same statement holds when we drop the orthonormality requirement if the spaces involved are finite-dimensional.

The Hilbert tensor product can also be constructed by identifying $H \otimes U = \mathcal{L}_2(U, H)$ via an isometric isomorphism, where the element $v_1 \otimes u_1$, $v_1 \in H$, $u_1 \in U$, is interpreted as the mapping

$$u_2 \mapsto \langle u_1, u_2 \rangle_U v_1.$$

We use the same notation for an element $F \in \mathcal{L}_2(U, H)$ and its representation $H \otimes U$.

Example 2.5. In the setting that $H = L^2(\mathcal{D}, \mathbb{R})$, the tensor product has a concrete interpretation. It holds that $H \otimes H = \mathcal{L}_2(H) = L^2(\mathcal{D} \times \mathcal{D}, \mathbb{R})$. With $u, v \in H$, $u \otimes v$ is identified with the function $uv \colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ given by $uv(x, y) = u(x)v(y)$ for almost every

$x, y \in \mathcal{D}$. For every operator $F \in \mathcal{L}_2(H)$ there is a function $f\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ such that F can be represented as an integral operator via

$$(Fv)(x) = \int_{\mathcal{D}} f(x,y)v(y)\,\mathrm{d}y$$

for all $v \in H$ and almost every $x \in \mathcal{D}$. Since

$$F = \sum_{i,j=1}^{\infty} \langle F, e_i \otimes e_j \rangle_{H\otimes H}\, e_i \otimes e_j = \sum_{i,j=1}^{\infty} \langle F, e_i \otimes e_j \rangle_{\mathcal{L}_2(H)}\, e_i \otimes e_j = \sum_{i,j=1}^{\infty} \langle Fe_j, e_i \rangle_H\, e_i \otimes e_j,$$

where $(e_i)_{i\in\mathbb{N}}$ is an orthonormal basis of H, it holds that

$$f(x,y) = \sum_{i,j=1}^{\infty} \langle Fe_j, e_i \rangle_H\, e_i(x) e_j(y)$$

for almost every $x, y \in \mathcal{D}$.

2.2 Probability theory

To be able to speak of stochastic processes in Hilbert spaces, we must first introduce the concept of a Hilbert space-valued random variable. Let $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t\in[0,T]}, P)$ be a complete filtered probability space satisfying the usual conditions, which is to say that $\mathcal{F}_0$ contains all P-null sets and $\mathcal{F}_t = \cap_{s>t}\mathcal{F}_s$ for all $t \in [0,T]$ with $T < \infty$. For a Hilbert space H, an *H-valued random variable*, or just a *random variable* if $H = \mathbb{R}$, is an $(\mathcal{A}, \mathcal{B}(H))$-measurable function $X\colon \Omega \to H$. Here $\mathcal{B}(H)$ refers to the Borel σ-algebra on H. The *expected value* of an H-valued random variable is defined as the Bochner integral

$$\mathbb{E}[X] = \int_{\Omega} X(\omega)\,\mathrm{d}P(\omega)$$

whenever $\|X\|_{L^1(\Omega,H)} < \infty$, where $\|X\|^p_{L^p(\Omega,H)} = \mathbb{E}[\|X\|^p_H]$ for $p \geq 1$. The *covariance* of $X \in L^2(\Omega, H)$ is defined by

$$\mathsf{Cov}(X) = \mathbb{E}\left[(X - \mathbb{E}[X]) \otimes (X - \mathbb{E}[X])\right]$$

and with the identification $H \otimes H = \mathcal{L}_2(H)$ its counterpart, the unique self-adjoint positive semidefinite operator $Q \in \mathcal{L}_2(H)$ for which, with $v \in H$,

$$Qv = \mathbb{E}\left[\langle X - \mathbb{E}[X], v\rangle_H\,(X - \mathbb{E}[X])\right]$$

is called the *covariance operator* of X although when there is no risk of confusion we also refer to this operator as the *covariance* of X. From this definition one can see that

$$\mathrm{Tr}(Q) = \|X - \mathbb{E}[X]\|^2_{L^2(\Omega,H)} < \infty.$$

The identity

$$\mathbb{E}[\langle X - \mathbb{E}[X], v\rangle_H \langle X - \mathbb{E}[X], w\rangle_H] = \langle Qv, w\rangle_H$$

for any $v, w \in H$ is a straightforward consequence of the definition of Q. More generally, the *cross-covariance* of X and $Y \in L^2(\Omega, U)$, where U is another Hilbert space, is defined by

$$\mathsf{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X]) \otimes (Y - \mathbb{E}[Y])] = \mathbb{E}[X \otimes Y] - \mathbb{E}[X] \otimes \mathbb{E}[Y] \in H \otimes U.$$

An H-valued random variable X is said to be *Gaussian* if $X \in H$ P-a.s. and $\langle X, v\rangle_H$ is a real-valued Gaussian random variable for all $v \in H$. In this case $X \in L^p(\Omega, H)$ for all $p \geq 1$ so $m = \mathbb{E}[X]$ and $Q = \mathsf{Cov}(X)$ are well-defined and we write $X \sim \mathcal{N}(m, Q)$. It can be shown that for each $m \in H$ and self-adjoint positive semidefinite $Q \in \mathcal{L}(H)$ with $\mathrm{Tr}(Q) < \infty$, there exists an H-valued random variable X such that $X \sim \mathcal{N}(m, Q)$.

An *H-valued stochastic process* $(X(t))_{t\in[0,T]}$ is a family of H-valued random variables. It is said to be *adapted to the filtration* $(\mathcal{F}_t)_{t\in[0,T]}$ if for each $t \in [0, T]$, $X(t)$ is $\mathcal{F}_t$-measurable. Two H-valued stochastic processes X, Y are said to be *modifications* of one another if for all $t \in [0, T]$, $X(t) = Y(t)$ P-almost surely.

A class of stochastic processes that is important in this thesis is that of so called Lévy processes. We restrict ourselves to the case of square integrable Lévy processes. For a trace class, self-adjoint, positive semidefinite operator $Q \in \mathcal{L}(H)$, an H-valued stochastic process $(L(t))_{t\in[0,T])}$ is said to be a *mean zero square integrable Q-Lévy process with respect to the filtration* $(\mathcal{F}_t)_{t\in[0,T]}$ if

- $L(0) = 0$ P-almost surely,
- L is continuous in probability, i.e., for any $\epsilon > 0$ and $t \in [0, T]$,

$$\lim_{\substack{s\to t \\ s\geq 0}} P(\|L(t) - L(s)\|_H > \epsilon) = 0,$$

- L has stationary increments,
- L is adapted to $(\mathcal{F}_t)_{t\in[0,T]}$,
- $L(t) - L(s)$ is independent of $\mathcal{F}_s$ for all $0 \leq s < t \leq T$,
- L is square integrable, i.e., $\|L(t)\|_{L^2(\Omega,H)} < \infty$ for all $t \geq 0$ and
- $L(t) - L(s)$, $0 \leq s < t \leq T$, has zero mean and covariance $(t - s)Q$.

With these assumptions it holds that L is an H-valued square integrable martingale.

An expansion of a mean zero square integrable Q-Lévy process $(L(t))_{t\in[0,T])}$ on the orthonormal eigenbasis $(q_i)_{i\in\mathbb{N}}$ of Q is called the *Karhunen–Loève expansion*

$$L = \sum_{i=1}^{\infty} \sqrt{\mu_i} L_i q_i, \tag{2}$$

where $(\mu_i)_{i\in\mathbb{N}}$ is the decreasing sequence of positive eigenvalues of Q in H, and we note that such eigenpairs exist since the assumption of a finite trace ensures that Q is compact. For the same reason, the sum (2) converges in $L^2(\Omega, H)$. Furthermore, $(L_i)_{i\in\mathbb{N}}$ is a sequence of uncorrelated identically distributed real-valued Q-Lévy processes with $Q = 1$.

A stochastic process $(W_t)_{t\in[0,T]}$ that is a square integrable Lévy process with the additional property that for all $0 \leq s < t \leq T$ the increment $W(t) - W(s) \sim \mathcal{N}(0, (t-s)Q)$, is said to be a *$Q$-Wiener process* (or a *standard Brownian motion* in the case of $H = \mathbb{R}$ and $Q = 1$) with respect to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$. This is the most important Lévy process for our purposes. If W is a Q-Wiener process, then for any $i \in \mathbb{N}$, the process L_i in (2) is a real-valued standard Brownian motion. For $(W_t)_{t\in[0,T]}$ we can also consider the case that Q is not of trace class. Then $(W_t)_{t\in[0,T]}$ is called a *generalized* Wiener process. The random variable $W(t)$, $t \in [0, T]$, is no longer strictly speaking H-valued, in the sense that $\|W(t)\|_{L^2(\Omega,H)} = \infty$. Instead, in analogy with the Karhunen–Loève expansion of a Q-Wiener process, one formally sets, for an orthonormal basis $(e_i)_{i\in\mathbb{N}}$ of H and a sequence $(\beta_i)_{i\in\mathbb{N}}$ of independent standard Brownian motions,

$$W = \sum_{i=1}^{\infty} \beta_i Q^{1/2} e_i. \tag{3}$$

We note that $H_0 = Q^{1/2}(H)$ is a Hilbert space equipped with the inner product

$$\langle \cdot, \cdot \rangle_{H_0} = \left\langle Q^{-\frac{1}{2}} \cdot, Q^{-\frac{1}{2}} \cdot \right\rangle_H,$$

with $Q^{-1/2}$ denoting the pseudo-inverse of $Q^{1/2}$. It turns out that (3) is well-defined as a standard Q-Wiener process in another Hilbert space $H_1 \supset H \supset H_0$ for which there exists a Hilbert–Schmidt embedding $J\colon H_0 \to H_1$. One can show that for each Hilbert space H, such a pair (H_1, J) always exist. The covariance operator of (3) in H_1 is then given by $Q_1 = JJ^*$, where J^* is the adjoint.

Example 2.6. Returning to the family of spaces $(\dot{H}^r)_{r\in\mathbb{R}}$ of domains of fractional powers of the operator Λ with eigenbasis $(e_i)_{i\in\mathbb{N}}$, a *white noise process* W (i.e., the case that $Q = I$) can be understood as an element of $L^2(\Omega, \dot{H}^{-r})$ for sufficiently large $r > 0$ since for $t \in [0, T]$, by the dominated convergence theorem,

$$\|W(t)\|^2_{L^2(\Omega,\dot{H}^{-r})} = t \sum_{i=1}^{\infty} \|e_i\|^2_{-r} = t \sum_{i=1}^{\infty} \lambda_i^{-r} < \infty$$

whenever there is a constant $C > 0$ and $\epsilon > 0$ such that $\lambda_i^{-r} \leq C i^{-1-\epsilon}$. In the setting of Example 2.1 with the operator $\Lambda = -\Delta$ being the Laplacian with zero Dirichlet boundary conditions, there exist constants C_1 and C_2 such that $C_1 i^{2/d} \leq \lambda_i \leq C_2 i^{2/d}$, see [19]. This means that $W(t) \in L^2(\Omega, \dot{H}^{-r})$ for all $r > d/2$ and $t \in [0, T]$.

2.3 Random fields

Let us now consider the specific setting that $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain and $H = L^2(\mathcal{D})$. We introduce random fields on $\mathcal{D}$ and give a condition under which such fields are elements of $L^2(\Omega, H)$.

Definition 2.7. A *random field* $X :: \mathcal{D} \times \Omega \to \mathbb{R}$ is a collection of random variables $(X(x))_{x\in\mathcal{D}}$ such that the mapping $x \times \omega \mapsto X(x)[\omega]$ is measurable with respect to the product σ-algebra $\mathcal{B}(\mathcal{D}) \otimes \mathcal{A}$.

Definition 2.8. A *second order random field* X is a random field with $X(x) \in L^2(\Omega, \mathbb{R})$ for all $x \in \mathcal{D}$. Its *covariance function* C is given by $C(x, y) = \mathsf{Cov}(X(x), X(y))$ for $x, y \in \mathcal{D}$ and its *mean function* m by $m(x) = \mathbb{E}[X(x)]$.

If we now assume that the pair of functions C and m fulfils

$$\int_{\mathcal{D}} C(x, x) + m(x)^2 \,\mathrm{d}x < \infty,$$

then, as a consequence of the joint measurability and Fubini's theorem, the mapping given by $\omega \mapsto X(\cdot)[\omega]$, which we denote by X, is in $L^2(\Omega, H)$ and the mean function m is equal to the expected value of X. For the same reason, the covariance operator of X is for $f \in H$ and $x \in \mathcal{D}$ given by

$$\begin{aligned}(Qf)(x) &= \mathbb{E}\left[\langle X - \mathbb{E}[X], f\rangle_H \left(X(x) - \mathbb{E}[X(x)]\right)\right] \\ &= \mathbb{E}\left[\int_{\mathcal{D}} \left(X(y) - m(y)\right)\left(X(x) - m(x)\right) f(y) \,\mathrm{d}y\right] = \int_{\mathcal{D}} C(x, y) f(y) \,\mathrm{d}y,\end{aligned}$$

i.e., Q can be represented in terms of C in the identification $\mathcal{L}_2(H) = L^2(\mathcal{D} \times \mathcal{D}, \mathbb{R})$, recalling that $\mathcal{L}_1(H) \subset \mathcal{L}_2(H)$. This fact can be exploited to numerically approximate the eigenpairs of Q from knowing only the covariance function, which can then in turn be used to generate samples of, e.g., a Q-Wiener process process by using the Karhunen–Loève expansion (2). We say that C is a *kernel* of the covariance operator, or just a *covariance kernel* for short. With the parameters $\sigma^2, \kappa \in \mathbb{R}$, an example of a common covariance kernel is the *exponential kernel*

$$C(x, y) = \frac{\sigma^2}{(2\pi)^{d/2}\kappa(d-1)!!} \exp(-\kappa|x - y|)$$

which one obtains as a special case of the *Matérn covariance kernel*

$$C(x, y) = \frac{2^{1-\nu}\sigma^2}{(4\pi)^{d/2}\Gamma(\nu + d/2)\kappa^{2\nu}} (\kappa|x - y|)^\nu K_\nu(\kappa|x - y|)$$

by setting $\nu = 1/2$. Here K_ν denotes the modified Bessel function of the second kind and Γ denotes the gamma function. In Figure 1 we see a realization, evaluated at four times, of an approximation of a Q-Wiener process taking values in $H = L^2(\mathcal{D})$, where $\mathcal{D} = (0, 1)^2$. Here Q is the covariance operator corresponding to a Matérn kernel with parameters $\nu = 3$, $\kappa = 25$ and σ^2 chosen to make $C(x, x) = 5$. The approximation was generated with FEniCS (see [2]) using the approach of [20].

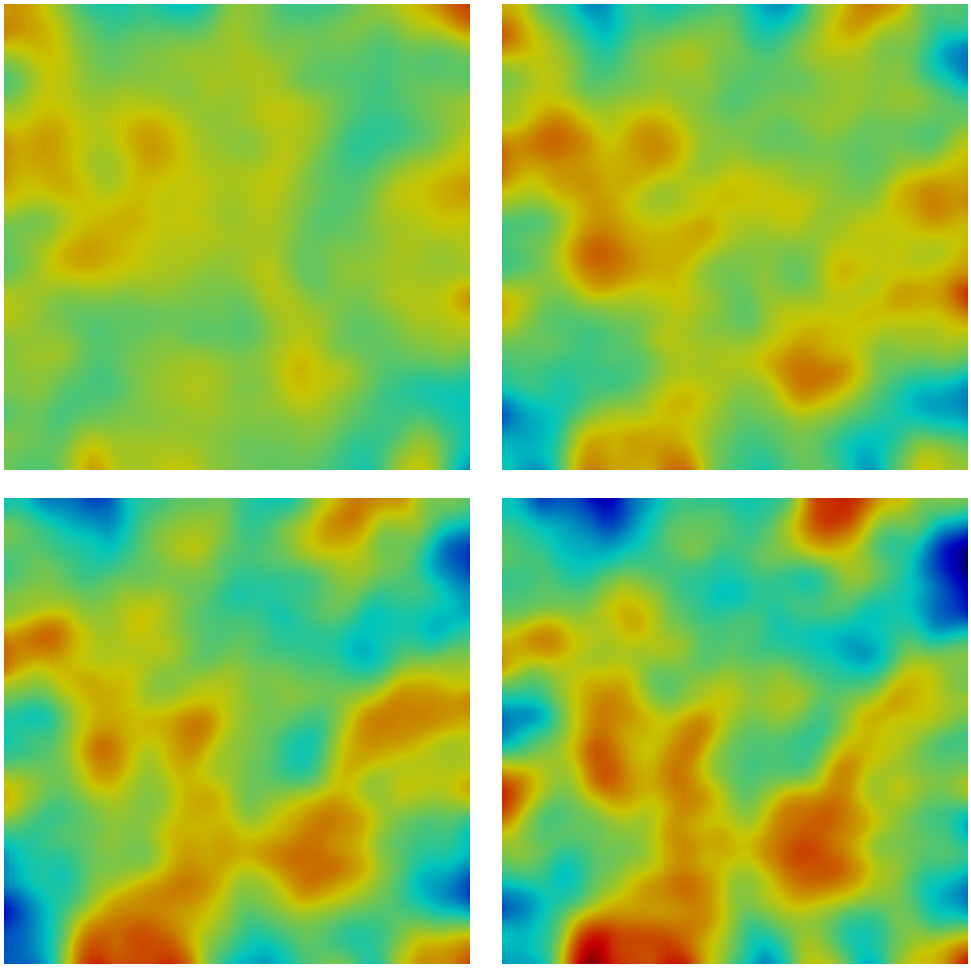


Figure 1: Realization of a Q-Wiener process in $H = L^2((0,1)^2, \mathbb{R})$, sampled at times $t = 0.25, 0.50, 0.75$ and 1.00.

2.4 Stochastic integration

To make sense of solutions to (1) we need a theory for the integration of stochastic processes with respect to square integrable martingales. That is to say, in our setting, for two Hilbert spaces H, U, $t \in [0,T]$ and an $\mathcal{L}(U,H)$-valued stochastic process Ψ on $[0,T]$, we want to make sense of the H-valued *stochastic Itô integral*

$$I_t^L(\Psi) = \int_0^t \Psi(s)\,\mathrm{d}L(s)$$

with respect to a mean zero square integrable U-valued Q-Lévy process L. We briefly reiterate the results of [23, Chapter 8] for our simpler setting of integration with respect to such processes. The integral is first defined in terms of so called *simple* integrands, which are those $\mathcal{L}(U,H)$-valued stochastic processes Ψ for which there, with $m \in \mathbb{N}$, exist

a sequence of times $0 = t_0 < t_1 < \ldots < t_i < t_{i+1} < \ldots < t_m = T$, a sequence $(\Psi_i)_{i=1}^{m-1}$ of $\mathcal{L}(U, H)$-valued operators and a sequence $(A_i)_{i=1}^{m-1}$ of events in $\mathcal{F}_{t_i}$ such that

$$\Psi(s) = \sum_{i=0}^{m-1} 1_{A_i} 1_{(t_i, t_{i+1}]}(s) \Psi_i$$

for $s \in [0, T]$, where 1_{A_i} and $1_{(t_i, t_{i+1}]}$ denote *indicator functions*, i.e., for a set A

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{otherwise.} \end{cases}$$

For these processes one sets

$$I_t^L(\psi) = \sum_{i=0}^{m-1} 1_{A_i} \Psi_i (L(t_{i+1} \wedge t) - L(t_i \wedge t))$$

and we obtain the so called *Itô isometry*

$$\mathbb{E}\left[\|I_t^L(\Psi)\|_H^2\right] = \mathbb{E}\left[\int_0^t \|\Psi(s)\|_{\mathcal{L}_2^0}^2 \, \mathrm{d}s\right], \tag{4}$$

where $U_0 = Q^{1/2}(U)$ and $\mathcal{L}_2^0 = \mathcal{L}_2(U_0, H)$. The space $\mathcal{N}_T^L(H)$ of admissible integrands is now defined as the completion of the space of simple processes with respect to the norm

$$\|\Psi\|_T = \left(\mathbb{E}\left[\int_0^T \|\Psi(s)\|_{\mathcal{L}_2^0}^2 \, \mathrm{d}s\right]\right)^{\frac{1}{2}},$$

and $I_t^L \colon \mathcal{N}_T^L(H) \to L^2(\Omega, H)$ is well-defined as a continuous extension. Since $\|\Psi\|_t \leq \|\Psi\|_T$, the Itô isometry (4) holds true for any $t \in [0, T]$ and any admissible integrand.

One can also construct $\mathcal{N}_T^L(H)$ by

$$\mathcal{N}_T^L(H) = L^2(([0, T] \times \Omega, \mathcal{P}_T, \, \mathrm{d}t \otimes P), \mathcal{L}_2^0),$$

where $\mathcal{P}_T$ denotes the predictable σ-algebra, i.e., the σ-algebra generated by the set

$$\{(s, t] \times A \subseteq [0, T] \times \Omega \,|\, 0 \leq s < t, A \in \mathcal{F}_s\}.$$

A stochastic process which is measurable with respect to this σ-algebra is said to be *predictable*, so we can integrate all predictable $\mathcal{L}_2^0$-valued stochastic processes.

This integral is well-defined also in the case that $L = W$ is a generalized Wiener process, having covariance Q that is not necessarily of trace class. Recall that in this case, W can be understood as a U_1-valued standard Q-Wiener process with covariance JJ^* for a Hilbert–Schmidt embedding $J \colon U_0 \to U_1$, where $U_1 \supset U \supset U_0$. By noting that

$$\|\Psi\|_{\mathcal{L}_2^0}^2 = \|\Psi J^{-1}\|_{\mathcal{L}_2(Q_1^{1/2}(U_1), H)}^2$$

for $\Psi \in \mathcal{L}_2^0$, one obtains a well-defined stochastic integral with respect to generalized Wiener processes by

$$\int_0^t \Psi(s)\,\mathrm{d}W(s) = \int_0^t \Psi(s)J^{-1}\,\mathrm{d}W(s)$$

for predictable $\mathcal{L}_2^0$-valued stochastic processes Ψ, see also [25, Chapter 2.5.2]. It can be shown that this definition is independent of the specific choice of U_1 and J.

3 Stochastic partial differential equations and approximations

In this section we return to (1), the SPDE of the introduction, to discuss what we mean by a solution to it. We introduce the approximations considered in this thesis, along with a brief review of Monte Carlo methods in the context of SPDE approximations.

3.1 Stochastic partial differential equations

Recall that the considered SPDE is given by

$$\begin{aligned} \mathrm{d}X(t) &= (AX(t) + F(t, X(t)))\ \mathrm{d}t + G(t, X(t))\,\mathrm{d}L(t), \\ X(0) &= X_0, \end{aligned} \tag{1}$$

where $t \in [0,T]$, $X_0 \in L^2(\Omega, H)$, $G\colon [0,T] \times H \to \mathcal{L}_2^0$, $F\colon [0,T] \times H \to H$ and L is a U-valued Q-Lévy process or a generalized Wiener process with covariance Q. This is to be understood as the integral equation

$$X(t) = X_0 + \int_0^t AX(s) + F(s, X(s))\,\mathrm{d}s + \int_0^t G(s, X(s))\,\mathrm{d}L(s), \tag{5}$$

where the first integral is of Bochner type and the second is the stochastic integral introduced in Section 2.4. In order to make sense of this process, there are several notions of solutions in the literature. Solutions formulated in terms of the integral equation (5) are referred to as *strong solutions*. We will, however, be concerned with the weaker concept of *mild solutions* which are formulated in terms of the semigroup $(E(t))_{t\in[0,T]}$ generated by A.

Definition 3.1. Let $X_0 \in L^2(\Omega, H)$. A predictable process $X = (X(t))_{t\in[0,T]}$ is called a *mild solution to* (1) if

$$\sup_{t\in[0,T]} \|X(t)\|_{L^2(\Omega,H)} < \infty$$

and for all $t \in [0,T]$

$$X(t) = E(t)X_0 + \int_0^t E(t-s)F(s, X(s))\,\mathrm{d}s + \int_0^t E(t-s)G(s, X(s))\,\mathrm{d}L(s),\ P\text{-a.s.} \tag{6}$$

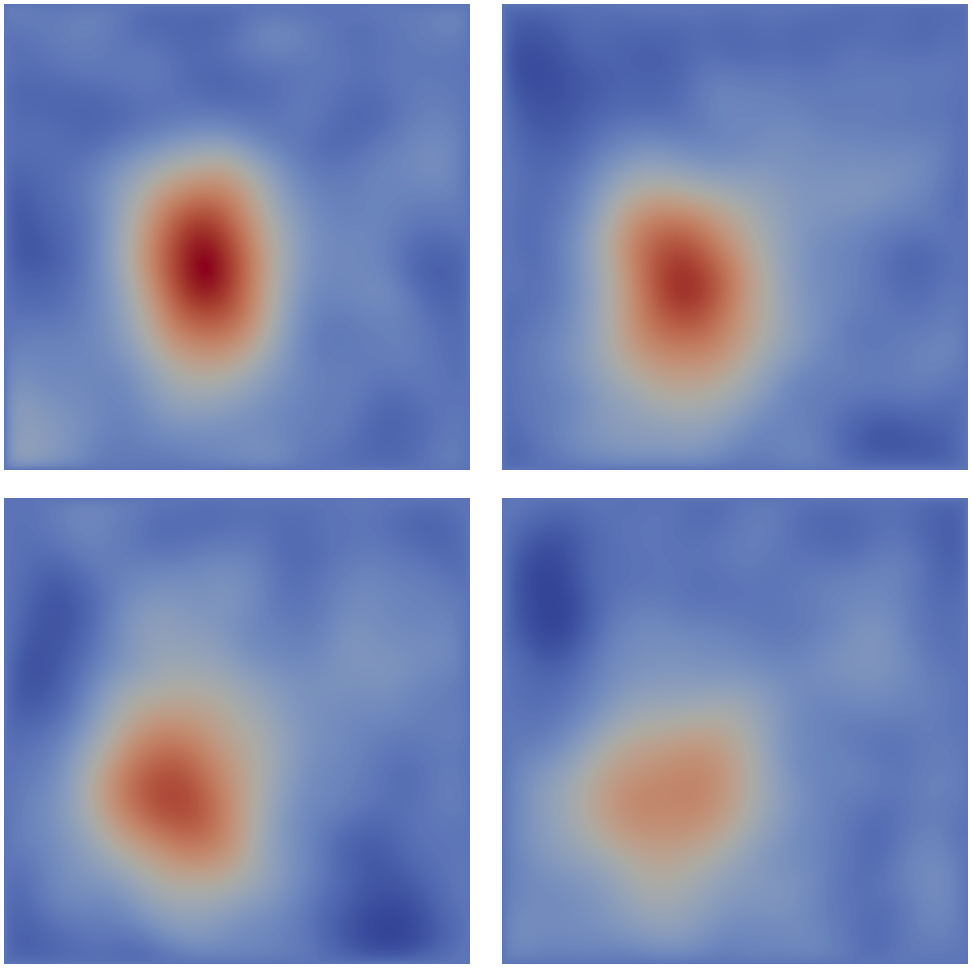


Figure 2: Realization of a stochastic reaction-diffusion equation with additive noise (Example 3.2) on $\mathcal{D} = (0,1)^2$, sampled at times $t = 1.5, 4, 6.5$ and 9.

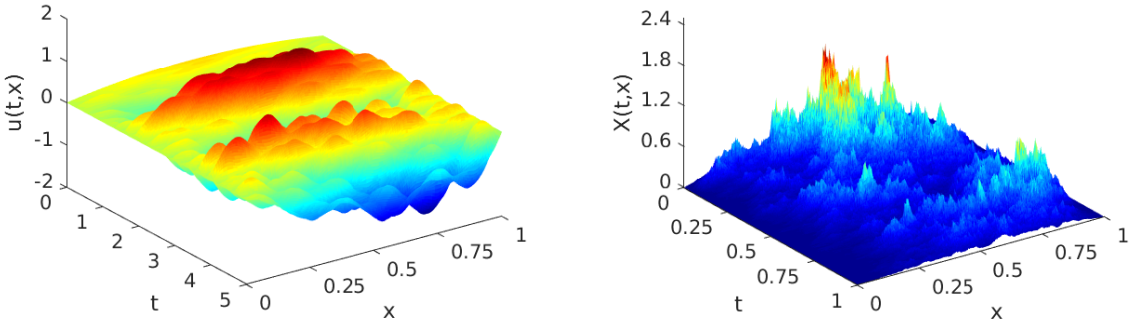
We return to specific assumptions that guarantee the existence of mild solutions below, but let us first consider a few concrete examples of SPDE.

Example 3.2 (Stochastic reaction-diffusion equation with additive noise)**.** Returning to the setting of Example 2.1, with $H = L^2(\mathcal{D})$ and $\Lambda v = -\nabla \cdot (a(\cdot)\nabla v) + c(\cdot)v$ with zero Dirichlet boundary conditions on $\mathcal{D}$, let $L = W$ be a Q-Wiener process in H. Let F be given by $F(t,v)(x) = b(t,x) \cdot \nabla v(x) + d(t,x)$ for $x \in \mathcal{D}$ and a function v on $\mathcal{D}$. When $d\colon \mathcal{D} \times [0,T] \to \mathbb{R}$ is smooth and $b\colon \mathcal{D} \times [0,T] \to \mathbb{R}^d$ is smooth and non-zero, $F(t,\cdot)$ does not, strictly speaking, map into H but is instead a member of $\mathcal{L}(H, \dot{H}^{-1})$ (cf. [15, Example 2.22]), which suffices for the existence of a mild solution. Let $G = g(t)\cdot$, where $g\colon [0,T] \to \mathbb{R}$ is smooth, and suppose also that the initial value X_0 is smooth. Then, with $A = -\Lambda$, (1) is the mathematical model of the equation

$$\mathrm{d}X(t,x) = \Big(\nabla \cdot (a(x)\nabla X(t,x)) + b(t,x) \cdot \nabla X(t,x) + c(t,x)X(t,x) + d(t,x)\Big)\,\mathrm{d}t$$

$$+ g(t)\,\mathrm{d}W(t,x)$$

for all $t \in (0,T]$ and $x \in \mathcal{D}$, with $X(t,x) = 0$ for all $t \in (0,T], x \in \partial\mathcal{D}$ and $X(0,x) = X_0(x)$ for all $x \in \mathcal{D}$, where X_0 is some sufficiently smooth element in H. From [15, Theorem 2.25] it follows that this equation has a well-defined mild solution. In Figure 2, we show a sample path of this equation at four different times, with $\mathcal{D} = (0,1)^2$, $a(x) = c_1$, $b(t,x) = c_2 \exp(-c_3 t)[1,1]^\top$, $c(x) = d(t,x) = 0$, $g(t) = c_4$ and $X_0(x) = c_5 x_1 x_2 (1-x_1)(1-x_2)$ for all $t \geq 0$ and $x = [x_1, x_2]^\top \in \mathcal{D}$, where c_1, c_2, c_3, c_4 and c_5 are some appropriate constants. The covariance operator of the noise is given by the same Matérn covariance kernel as in Figure 1. Similar *stochastic reaction-diffusion equations* have been used to model weather patterns. We will return to this example when illustrating convergence of SPDE approximations below.



(a) Stochastic wave equation with additive noise.

(b) Stochastic heat equation with multiplicative noise.

Figure 3: Realizations of the two SPDE from Examples 3.3 and 3.4 on $\mathcal{D} = [0,1]$ with $T = 5$ (left) and $T = 1$ (right).

Example 3.3 (Stochastic wave equation with additive noise)**.** Let $\dot{H}^0 = L^2(\mathcal{D})$, and $\Lambda = -\Delta$ with zero Dirichlet boundary conditions (i.e., a special case of the operator in Example 2.1), let $L = W$ be a generalized Wiener process in $\dot{H}^0$ with covariance Q. The *stochastic wave equation* is given by

$$\mathrm{d}\dot{u}(t,x) = \Delta u(t,x)\,\mathrm{d}t + \mathrm{d}W(t,x),$$

for all $t \in (0,T]$ and $x \in \mathcal{D}$, with $u(t,x) = 0$ for all $t \in (0,T], x \in \partial\mathcal{D}$ and $u(0,x) = u_0(x)$, $\dot{u}(0,x) = v_0(x)$ for all $x \in \mathcal{D}$. Here $\dot{u}$ denotes the first time derivative of u and u_0, v_0 are sufficiently smooth elements in $\dot{H}^0$. With A as in Example 2.4, this can be put into the framework of (1) by writing it as

$$\mathrm{d}X(t) = AX(t)\,\mathrm{d}t + B\,\mathrm{d}W(t)$$

on $\mathcal{H} = \dot{H}^0 \oplus \dot{H}^{-1}$, with A and $\mathcal{H}$ as in Example 2.4, and with

$$B = \begin{bmatrix} 0 \\ I \end{bmatrix} \text{ and } X(t) = \begin{bmatrix} X_1(t) \\ X_2(t) \end{bmatrix} = \begin{bmatrix} u(t) \\ \dot{u}(t) \end{bmatrix}$$

for $t \in [0, T]$. See Figure 3(a) for a realization of this equation in $\mathcal{D} = (0, 1)$ with $T = 5$, $u_0(x) = x(1 - x)$, $v_0(x) = 0$ for $x \in \mathcal{D}$. The covariance operator Q is given by a kernel q with $q(x, y) = 10 \cos(20(x - y)) + 5 \min(x, y)$ for $x, y \in \mathcal{D}$. In Paper 4 we consider a more general form of this equation and show existence and uniqueness for its mild solution. Similar equations have been used to model the vertical movement of a DNA molecule suspended in liquid, see [9].

Example 3.4 (Stochastic heat equation with multiplicative noise)**.** Let H be as in Example 3.2, let $A = -\Lambda = \Delta$ with zero Dirichlet boundary conditions, let $L = W$ be a generalized Wiener process in $\dot{H}^0$ with covariance Q, let $F = 0$ and let G be a linear operator given by $(G(u)v)(x) = u(x)v(x)$ for almost every $x \in \mathcal{D}$. The SPDE (1) is then a model for the *stochastic heat equation with multiplicative noise* given by

$$\mathrm{d}X(t, x) = \Delta X(t, x)\,\mathrm{d}t + X(t, x)\,\mathrm{d}W(t, x),$$

for all $t \in (0, T]$ and $x \in \mathcal{D}$, with $X(t, x) = 0$ for all $t \in (0, T], x \in \partial\mathcal{D}$ and $X(0, x) = X_0(x)$ for all $x \in \mathcal{D}$. Strictly speaking, the operator G on H does not map into $\mathcal{L}_2^0$ if Q is not of trace class, but for $d = 1$, one can still deduce the existence of the mild solution (6) when $Q = I$, i.e., when W is a white noise process. In Figure 3(b) we show a realization of the equation for the case that $Q = I$. The initial value is given by a hat function and the operators Δ and Q have been rescaled by appropriate constants. In Paper 5, we study an operator connected to the quadratic behaviour of this equation.

3.2 Spatial discretization

Since an analytic solution to (1) is rarely available, one has to *discretize* the equation in space and in time in order to simulate it on a computer. We speak of a *fully discrete approximation* of the mild solution X if it is discretized in both space and time. To arrive at such an approximation is the goal of this section and the next. In this first part, we consider spatial discretizations. The main idea is to seek solutions to (1) in some finite-dimensional subspace of H, where the operators involved are replaced with finite-dimensional counterparts.

For this, let $(V_h)_{h\in(0,1]}$ be a family of subspaces of H such that $\dim(V_h) = N_h < \infty$, $h \in (0, 1]$ and denote by $P_h \colon H \to V_h$ the orthogonal projector onto V_h. From [15, Section 3.2] we cite two examples of subspace families in the context of $H = L^2(\mathcal{D})$ for some bounded convex domain $\mathcal{D}$ with polygonal boundary.

Example 3.5 (Standard finite element method)**.** Let $(T_h)_{h\in(0,1]}$ be a regular quasi-uniform family of triangulations of $\mathcal{D}$ with h being the maximal mesh size. We let V_h be the space of all functions that are continuous and piecewise linear on T_h and zero at the boundary of $\mathcal{D}$.

Example 3.6 (Spectral Galerkin method)**.** We restrict ourselves to the case that $\mathcal{D} = [0,1]$ and $A = -\Lambda = \Delta$ with zero Dirichlet boundary conditions. In this setting the orthonormal eigenbasis $(e_i)_{i\in\mathbb{N}}$ and the sequence of eigenvalues $(\lambda_i)_{i\in\mathbb{N}}$ to Λ are explicitly known to be $e_i = \sqrt{2}\sin(i\pi\cdot)$ and $\lambda_i = i^2\pi^2$ for all $i \in \mathbb{N}$. If we now, for $N \in \mathbb{N}$, set $h = \lambda_{N+1}^{-1/2}$ and $V_h = \operatorname{span}(e_1, e_2, \dots, e_N)$ we get a sequence of finite-dimensional spaces.

Given the family $(V_h)_{h\in(0,1]}$, assuming also that $V_h \in \dot{H}^1$ for all $h \in (0,1]$ (as is the case in the previous two examples), one way of arriving at a discrete version of $A = -\Lambda$ is to set $A_h = -\Lambda_h$, defined by the relationship

$$\langle \Lambda_h v_h, w_h\rangle_H = \langle v_h, w_h\rangle_1 = \left\langle \Lambda^{\frac{1}{2}} v_h, \Lambda^{\frac{1}{2}} w_h \right\rangle_H$$

for all $v_h, w_h \in V_h$. It is straightforward to see that this operator is self-adjoint and positive definite (hence invertible) on V_h which, in the same way as before, entails the existence of an orthonormal eigenbasis $(e_{h,i})_{i=1,\dots,N_h}$ of V_h and an increasing positive sequence $(\lambda_{h,i})_{i=1,\dots,N_h}$ of eigenvalues of Λ_h. For the same reason, $A_h = -\Lambda_h$ generates a C_0-semigroup E_h of contractions on V_h. When A is replaced by A_h, F by P_hF, G by P_hG and X_0 by P_hX_0 in (1), the resulting equation is called a *semidiscrete approximation* of the SPDE.

3.3 Spatio-temporal discretization

To arrive at a fully discrete approximation of (1), we consider *rational approximations* of the semigroup $(E_h(t))_{t\in[0,T]}$ on a finite-dimensional space $V_h \subset H$, where we follow the approach of [26]. Let us therefore consider a uniform time grid given by $t_j = j\Delta t$ for $j = 0, \dots, N_{\Delta t}$, where $N_{\Delta t} \in \mathbb{N}$ and $\Delta t = TN_{\Delta t}^{-1}$.

A *rational approximation of order* ρ of the exponential function is a rational function $R\colon \mathbb{C} \to \mathbb{C}$ such that there exist constants $C, \delta > 0$ satisfying for all $z \in \mathbb{C}$ with $|z| < \delta$,

$$|R(z) - \exp(z)| \le C|z|^{\rho+1}.$$

By applying R to $\Delta t A_h$ we obtain a rational approximation of the semigroup generated by A_h. This is supposed to be understood in the spectral sense, i.e., we assume that $-A_h$ has an orthonormal eigenbasis $(e_{h,k})_{k=1,\dots,N_h}$ of V_h, as in the previous section, along with a sequence $(\lambda_{h,k})_{k=1,\dots,N_h}$ of eigenvalues, and set

$$R(\Delta t A_h)v_h = \sum_{k=1}^{N_h} R(-\Delta t\lambda_{h,k}) \langle v_h, e_{h,k}\rangle_H e_{h,k}.$$

Since R is rational there exist polynomials r_{n} and r_{d} such that $R = r_{\mathrm{d}}^{-1} r_{\mathrm{n}}$.

The fully discrete approximation $X_{h,\Delta t} = (X_{h,\Delta t}^{t_j})_{j=0,\dots,N_{\Delta t}}$ of the mild solution to (1) is now given by the recursion scheme

$$\begin{aligned} X_{h,\Delta t}^{t_{j+1}} &= D_{\Delta t,h}^{\det,j}\big(X_{h,\Delta t}^{t_j}\big) + D_{\Delta t,h}^{\mathrm{stoch},j}\big(X_{h,\Delta t}^{t_j}\big), \\ X_{h,\Delta t}^{0} &= P_h X_0 \end{aligned} \tag{7}$$

for $j = 0, \dots, N_{\Delta t} - 1$. Here we set, for $v_h \in V_h$, the *deterministic approximation operator* on V_h to be either

$$D^{\mathrm{det},j}_{\Delta t,h}(v_h) = R(\Delta t A_h)v_h + r_{\mathrm{d}}^{-1}(\Delta t A_h)\Delta t P_h F(t_j, v_h) \tag{8}$$

or

$$D^{\mathrm{det},j}_{\Delta t,h}(v_h) = R(\Delta t A_h)v_h + R(\Delta t A_h)\Delta t P_h F(t_j, v_h). \tag{9}$$

Similarly, we let for $j = 0, \dots, N_{\Delta t} - 1$, the *stochastic approximation operator* be given by

$$D^{\mathrm{stoch},j}_{\Delta t,h}(v_h) = r_{\mathrm{d}}^{-1}(\Delta t A_h)P_h G(t_j, v_h)\Delta L^j, \tag{10}$$

or

$$D^{\mathrm{stoch},j}_{\Delta t,h}(v_h) = R(\Delta t A_h)P_h G(t_j, v_h)\Delta L^j, \tag{11}$$

where ΔL^j denotes the Lévy increment $L(t_{j+1}) - L(t_j)$.

Example 3.7. An important example of a rational approximation of $(E_h(t))_{t\in[0,T]}$ is the *backward Euler scheme*, where $R(\Delta t A_h)$ is defined through $r_{\mathrm{d}}(x) = 1 - x$ and $r_{\mathrm{n}}(x) = 1$ for all $x \in \mathbb{R}$, $x \neq 1$. Then, (8) coincides with (9) and (10) coincides with (11). One can rewrite Scheme (7) as

$$X^{t_{j+1}}_{h,\Delta t} - X^{t_j}_{h,\Delta t} = \left(A_h X^{t_{j+1}}_{h,\Delta t} + P_h F(t_j, X^{t_j}_{h,\Delta t})\right)\Delta t + P_h G(t_j, X^{t_j}_{h,\Delta t})\Delta L^j, \tag{12}$$

for $j = 0, \dots, N_{\Delta t} - 1$. For (8) and (10), another example is the *forward Euler scheme* defined through $r_{\mathrm{d}}(x) = 1$ and $r_{\mathrm{n}}(x) = 1 + x$ for all $x \in \mathbb{R}$ which can similarly be rewritten as

$$X^{t_{j+1}}_{h,\Delta t} - X^{t_j}_{h,\Delta t} = \left(A_h X^{t_j}_{h,\Delta t} + P_h F(t_j, X^{t_j}_{h,\Delta t})\right)\Delta t + P_h G(t_j, X^{t_j}_{h,\Delta t})\Delta L^j,$$

for $j = 0, \dots, N_{\Delta t} - 1$. Both of these schemes are considered in the stability analysis of Paper 2. The *Crank–Nicolson scheme* is employed in the simulation section of Paper 4 and is obtained from (9) and (11) with $R(z) = (1 + z/2)/(1 - z/2)$, which can be written as

$$\left(1 - \frac{1}{2}A_h\right) X^{t_{j+1}}_{h,\Delta t} = \left(1 + \frac{1}{2}A_h\right)\left(X^{t_j}_{h,\Delta t} + P_h F(t_j, X^{t_j}_{h,\Delta t})\Delta t + P_h G(t_j, X^{t_j}_{h,\Delta t})\Delta L^j\right),$$

3.4 Strong and weak convergence

We now briefly review two different notions of convergence of $X^T_{h,\Delta t}$ to $X(T)$, the mild solution of (1) at time T. The approximation is said to *converge strongly* to X if

$$\|X^T_{h,\Delta t} - X(T)\|_{L^2(\Omega,H)} \to 0$$

as $h, \Delta t \to 0$. However, one might not always be interested in approximating X in a mean square sense but only in the expected value of a functional of the solution. We say that $\hat{X}$ *converges weakly* to X if

$$\left|\mathbb{E}\left[\phi\big(X^T_{h,\Delta t}\big) - \phi\big(X(T)\big)\right]\right| \to 0$$

as $h, \Delta t \to 0$ for all sufficiently smooth functionals $\phi\colon H \to \mathbb{R}$. In the classical case, the *test function* ϕ is only assumed to be Lipschitz continuous, this is equivalent to $X_{h,\Delta t}^T \to X(T)$ in distribution, in a certain sense. Clearly, strong convergence then implies weak convergence, but for a particular test function one typically expects the weak convergence rate to be higher than the strong convergence rate. To show this one needs to replace the Lipschitz continuity with some kind of differentiability condition. In the field of convergence analysis of SPDE approximations, a common rule of thumb is that the weak convergence rate is twice that of the strong convergence. We illustrate this in the simple setting below, which coincides with parts of Paper 3. We first make some assumptions on the parameters of the SPDE.

Assumption 3.8. We assume that $H = L^2(\mathcal{D})$ and $A = -\Lambda$ are given by the expressions in Example 2.1, that F and G only depend on $t \in [0,T]$ and that the initial value $X_0 \in \dot{H}^1$ is deterministic. The noise $L = W = (W(t))_{t\in[0,T]}$ is an $(\mathcal{F}_t)_{t\in[0,T]}$-adapted generalized Wiener process in H with covariance operator Q that is self-adjoint and positive semidefinite but not necessarily of trace class. Moreover, for some $\delta \in [1/2, 1]$, there exists a constant $C > 0$ such that $G\colon [0,T] \to \mathcal{L}_2^0$ and $F\colon [0,T] \to H$ satisfy

$$\|G(t_1) - G(t_2)\|_{\mathcal{L}_2^0} \leq C|t_1 - t_2|^\delta, \text{ for all } t_1, t_2 \in [0,T]$$

and

$$\|F(t_1) - F(t_2)\|_H \leq C|t_1 - t_2|^\delta, \text{ for all } t_1, t_2 \in [0,T].$$

This assumption guarantees the existence of the mild solution (6) by [15, Theorem 2.25] and the uniform bound

$$\sup_{t\in[0,T]} \|X(t)\|_{L^p(\Omega,H)} < \infty \tag{13}$$

holds true for any $p \geq 1$. The following discretization ensures the strong convergence of $\hat{X}$ to X.

Assumption 3.9. The equation is discretized in space with V_h given by either Example 3.5 or Example 3.6. The fully discrete approximation is given by the backward Euler scheme (12) in Example 3.7.

The next result on strong convergence is a direct consequence of [15, Theorem 3.14].

Theorem 3.10. *Under Assumptions 3.8 and 3.9, for any $p \geq 1$, there exists a constant C such that, for any fully discrete approximation $X_{h,\Delta t}$,*

$$\|X_{h,\Delta t}^T - X(T)\|_{L^p(\Omega,H)} \leq C\left(h + \Delta t^{\frac{1}{2}}\right).$$

We say that the strong convergence is of rate 1 in space and 1/2 in time. As mentioned above, the weak convergence rate is in many instances twice that of the strong rate (see e.g., [3, 11, 14] for fully discrete approximations with additive noise and [5, 7, 10] for semi-discrete approximations with multiplicative noise). This is true also in this setting, when $\delta = 1$. The following theorem is proven in [15, Theorem 5.12].

Theorem 3.11. *Under Assumptions 3.8 and 3.9, for all $\phi \in C_{\mathrm{p}}^2(H,\mathbb{R})$, there exists a constant $C > 0$ such that*

$$\left|\mathbb{E}\left[\phi(X(T)) - \phi(X_{h,\Delta t}^T)\right]\right| \leq C\left(1 + |\log(h)|\right)\left(h^2 + \Delta t^\delta\right), \text{ for all } h, \Delta t \in (0,1].$$

3.5 Monte Carlo methods

Since the expected value of functionals of the mild solution of (1) cannot in general be evaluated explicitly, we have to introduce an approximation of the expectation operator $\mathbb{E}[\cdot]$. We will consider two approximations that are both based on simulating a large number of approximate solutions to (1). We formulate the theory in the general setting of real-valued random variables.

The *Monte Carlo estimator* E_N of a real-valued random variable $Y \in L^2(\Omega,\mathbb{R})$ is given by

$$E_N[Y] = \frac{1}{N}\sum_{i=1}^N Y^{(i)},$$

where $(Y^{(i)})_{i=1,\dots,N}$ is a sequence of independent, identically distributed random variables that have the same law as Y. The convergence of $E_N[Y]$ to $\mathbb{E}[Y]$ as $N \to \infty$ is ensured by a mean square version of the *law of large numbers*

$$\begin{aligned} \|\,\mathbb{E}[Y] - E_N[Y]\|_{L^2(\Omega,\mathbb{R})}^2 &= \left\| \frac{1}{N}\sum_{i=1}^N \left(\mathbb{E}[Y] - Y^{(i)}\right)\right\|_{L^2(\Omega,\mathbb{R})}^2 \\ &= \frac{1}{N^2}\sum_{i=1}^N \left\|\mathbb{E}[Y] - Y^{(i)}\right\|_{L^2(\Omega,\mathbb{R})}^2 \\ &= \frac{1}{N}\,\mathsf{Var}(Y) \leq \frac{1}{N}\|Y\|_{L^2(\Omega,\mathbb{R})}^2. \end{aligned} \tag{14}$$

Instead of a single random variable Y we can consider a sequence $(Y_\ell)_{\ell\in\mathbb{N}_0}$ of random variables, where $Y_\ell \in L^2(\Omega,\mathbb{R})$ and where the index $\ell \in \mathbb{N}_0$ is referred to as a *level*. The *multilevel Monte Carlo estimator* E^L of $Y_L \in (Y_\ell)_{\ell\in\mathbb{N}_0}$ is, for $L \in \mathbb{N}$, defined by

$$E^L[Y_L] = E_{N_0}[Y_0] + \sum_{\ell=1}^L E_{N_\ell}[Y_\ell - Y_{\ell-1}], \tag{15}$$

where $(N_\ell)_{\ell=0,\dots,L}$ consists of level specific numbers of samples in the respective Monte Carlo estimators. A telescoping sum argument shows that as an estimator of $\mathbb{E}[Y_L]$ the multilevel Monte Carlo estimator is unbiased. Under the assumption that $(Y_\ell)_{\ell\in\mathbb{N}_0}$ converges to some random variable Y, a calculation similar to (14) leads to the error estimate

$$\|\,\mathbb{E}[Y] - E^L[Y_L]\|_{L^2(\Omega,\mathbb{R})} \leq |\,\mathbb{E}[Y - Y_L]\,| + \left(\frac{1}{N_0}\|Y_0\|_{L^2(\Omega,\mathbb{R})}^2 + \sum_{l=1}^L \frac{1}{N_\ell}\|Y_\ell - Y_{\ell-1}\|_{L^2(\Omega,\mathbb{R})}^2\right)^{\frac{1}{2}}.$$

In this context, the advantage of using a multilevel Monte Carlo estimator compared to a standard Monte Carlo estimator is due to the flexibility allowed by letting the sample sizes $(N_\ell)_{\ell=0,\dots,L}$ depend on a bound on $\|Y_\ell - Y_{\ell-1}\|_{L^2(\Omega,\mathbb{R})}$, $\ell = 1,\dots,L$. In the case that sampling Y_ℓ for small ℓ is computationally cheaper than sampling Y_L, we can let the sampling effort be concentrated on the coarser levels $\ell \ll L$. We then need to choose the sample sizes in such a way no single term dominates the overall error. The following theorem, which is essentially the same as [16, Theorem 1], see also [24, Proposition 2], shows how this is done and provides bounds on the overall computational work.

Theorem 3.12. *Let $(a_\ell)_{\ell\in\mathbb{N}_0}$ be a decreasing sequence of positive real numbers that converges to zero and assume that there exist constants C_1, C_2, C_3 and a parameter $\eta \in [0,1]$ such that for all $\ell \in \mathbb{N}$, Y_ℓ fulfils*

$$|\mathbb{E}[Y - Y_\ell]| \leq C_1(1 + |\log(a_\ell)|)a_\ell,$$

Y_ℓ and $Y_{\ell-1}$ fulfil

$$\|Y_\ell - Y_{\ell-1}\|_{L^2(\Omega,\mathbb{R})} \leq C_2 a_\ell^\eta$$

and Y_0 fulfils $\|Y_0\|_{L^2(\Omega,\mathbb{R})} = C_3$. For $L \in \mathbb{N}, \ell = 1,\dots,L$, $\epsilon > 0$, set $N_\ell = \lceil a_L^{-2} a_\ell^{2\eta} \ell^{1+\epsilon} \rceil$, where $\lceil\cdot\rceil$ is the ceiling function, and $N_0 = \lceil a_L^{-2} \rceil$. Then

$$\|\,\mathbb{E}[Y] - E^L[Y_L]\,\|_{L^2(\Omega,\mathbb{R})} \leq \left(C_1^2 + C_3 + C_2 + \zeta(1+\epsilon)\right)^{\frac{1}{2}} (1 + |\log(a_L)|)a_L,$$

where ζ denotes the Riemann zeta function. Furthermore, assume that, for some constants C_4, C_5 and $\kappa, \delta > 0$, the work $\mathcal{W}_\ell^B$ of one calculation of $Y_\ell - Y_{\ell-1}$, $\ell \geq 1$, is bounded by $C_4 a_\ell^{-\kappa}$ and that the work of one calculation of Y_0 is bounded by C_5. Then there exists another constant C_6 such that the overall work $\mathcal{W}_L$ is bounded by

$$\mathcal{W}_L \leq C_6 a_L^{-2} \left(C_5 + C_4 \sum_{\ell=1}^{L} a_\ell^{-(\kappa-2\eta)} \ell^{1+\epsilon} \right).$$

Furthermore, if there exists $a > 1$ and $\zeta > 0$ such that $a_\ell \simeq a^{-\ell}$ then the bound on $\mathcal{W}_L$ simplifies to

$$\mathcal{W}_L = \begin{cases} \mathrm{O}(a_L^2), & \text{if } \kappa < 2\eta \\ \mathrm{O}(a_L^{-(2+\kappa-2\eta)} L^{2+\epsilon}), & \text{if } \kappa \geq 2\eta. \end{cases}$$

Above, the notation $a_\ell \simeq a^{-\ell}$ means that a_ℓ can be bounded from above and below by non-zero constants, not depending on ℓ, times $a^{-\ell}$.

Example 3.13. Let us consider a concrete example of a Monte Carlo simulation under Assumption 3.8 with $\delta = 1$ and Assumption 3.9, where we compare the computational costs of the Monte Carlo and multilevel Monte Carlo estimators. Below we let $C > 0$ be a generic constant that may change from line to line.

Recall that we seek to approximate the $L^2(\mathcal{D})$-valued solution of (1) with the family $\hat{X} = (X_{h,\Delta t}, h \in (0,1], N_{\Delta t} \in \mathbb{N})$, where the V_h-valued sequence $X_{h,\Delta t}$ is given by the

backward Euler scheme (12) and V_h is the space of all piecewise linear functions on T_h, a triangulation with maximal mesh size h. Let us introduce a subsequence of approximations, indexed by levels ℓ, by $\check{X} = (X_\ell = X_{h,\Delta t}, h = 2^{-\ell}, N_{\Delta t} = h^{-2}, \ell \in \mathbb{N}_0)$, and suppose that it is primarily the end time value $X(T)$ that we are interested in. Theorem 3.10 then ensures that for all $p \geq 1$ there exists a constant C such that for all $\ell \in \mathbb{N}_0$

$$\|X(T) - X_\ell^T\|_{L^p(\Omega,H)} \leq C2^{-\ell}.$$

Given a functional $\phi \in C^2_{\mathrm{p}}(H, \mathbb{R})$, we now set $Y = \phi(X(T))$ and let Y_ℓ of $(Y_\ell)_{\ell\in\mathbb{N}_0}$ be given by $Y_\ell = \phi(X_\ell^T)$. By the mean value theorem we have

$$\begin{aligned}
&\|Y - Y_\ell\|^2_{L^2(\Omega,\mathbb{R})} \\
&\quad = \mathbb{E}\left[\left|\phi\big(X(T)\big) - \phi\big(X_\ell^T\big)\right|^2\right] \\
&\quad = \mathbb{E}\left[\left|\int_0^1 \left\langle \phi'\big(X_\ell^T + s\big(X(T) - X_\ell^T\big)\big), X(T) - X_\ell^T\right\rangle_H \,\mathrm{d}s\right|^2\right] \qquad (16)\\
&\quad \leq C\|X(T) - X_\ell^T\|^2_{L^{2(p+1)}(\Omega,H)} \left(1 + \|X(T)\|^{2p}_{L^{2(p+1)}(\Omega,H)} + \|X_\ell^T\|^{2p}_{L^{2(p+1)}(\Omega,H)}\right) \\
&\quad \leq C\|X(T) - X_\ell^T\|^2_{L^{2(p+1)}(\Omega,H)}
\end{aligned}$$

by using Hölder's inequality and the assumption that there exist $p \geq 2$ and $C > 0$ such that $\|\phi'(f)\|_H \leq C(1 + \|f\|_H^p)$ for all $f \in H$ (cf. [15, Chapter 1]). Here the constant of the first inequality depends on p but not on ℓ. In the second inequality, we use the uniform bound (13) on X and the fact that the convergence result of Theorem 3.10 implies a similar bound on X_ℓ to get a constant which depends on X but not on ℓ. By a similar argument one shows that there exists a constant C, not depending on ℓ, such that

$$\|Y_\ell\|_{L^2(\Omega,\mathbb{R})} = \mathbb{E}\left[\left|\phi\left(X_\ell^T\right)\right|^2\right]^{\frac{1}{2}} \leq C$$

for all $\ell \in \mathbb{N}_0$.

We use these results to deduce that the error of the standard Monte Carlo estimation of $\mathbb{E}[\phi(X(T))]$ is, for $\ell \in \mathbb{N}_0$, by the triangle inequality, Theorem 3.11 and (14) bounded by

$$\begin{aligned}
&\left\|\mathbb{E}\left[\phi(X(T))\right] - E_N\left[\phi(X_\ell^T)\right]\right\|_{L^2(\Omega,\mathbb{R})} \\
&\quad \leq \left\|\mathbb{E}\left[\phi(X(T))\right] - \mathbb{E}\left[\phi(X_\ell^T)\right]\right\|_{L^2(\Omega,\mathbb{R})} + \left\|\mathbb{E}\left[\phi(X_\ell^T)\right] - E_N\left[\phi(X_\ell^T)\right]\right\|_{L^2(\Omega,\mathbb{R})} \\
&\quad \leq C(1+\ell)2^{-2\ell} + \frac{1}{\sqrt{N}}\|\phi(X_\ell^T)\|_{L^2(\Omega,\mathbb{R})} \leq C\left((1+\ell)2^{-2\ell} + \frac{1}{\sqrt{N}}\right).
\end{aligned}$$

Therefore, to ensure that the Monte Carlo error does not dominate the error of the approximation of $\mathbb{E}[\phi(X(T))]$ one should set the number of samples $N \simeq (1+\ell)^{-2}2^{4\ell}$. For

the multilevel Monte Carlo scheme, Theorem 3.10 along with (16) ensure that

$$\begin{aligned}\|Y_\ell - Y_{\ell-1}\|_{L^2(\Omega,\mathbb{R})} &\leq \|Y - Y_\ell\|_{L^2(\Omega,\mathbb{R})} + \|Y - Y_{\ell-1}\|_{L^2(\Omega,\mathbb{R})} \\ &\leq C\left(2^{-\ell} + 2^{-\ell-1}\right) \leq C2^{-\ell},\end{aligned}$$

so that the conditions of Theorem 3.12 are fulfilled with $a_\ell = 2^{-2\ell}$, $\ell \in \mathbb{N}_0$, and $\eta = 1/2$. Therefore there exists a constant C such that for all $L \in \mathbb{N}$

$$\| \mathbb{E}[Y] - E^L[Y_L] \|_{L^2(\Omega,\mathbb{R})} \leq C(1+L)2^{-2L}$$

as long as the level dependent sample sizes N_ℓ, $\ell \in \mathbb{N}_0$, are chosen to be $N_\ell = \lceil 2^{4L-2\ell}\ell^{1+\epsilon} \rceil$ for $\ell > 0$ and $N_0 = 2^{4L}$. This means that for a given level L, the majority of samples are computed at a coarse level while retaining the same rate of convergence compared to the standard Monte Carlo method. Assuming that the computational work of solving the backward Euler system (12) at one time step of level ℓ is bounded by $\mathrm{O}(2^{\alpha\ell d})$, where d is the dimension of the underlying spatial domain and $\alpha \in [1,2]$, and that the computational cost of computing $Y_\ell - Y_{\ell-1}$ is roughly equivalent to the cost of Y_ℓ, for $\ell \in \mathbb{N}_0$, the total cost of computing $E^L[Y_L]$ is by Theorem 3.12 bounded by $\mathcal{W}_L = \mathrm{O}(2^{(4+\alpha d)L}L^{2+\epsilon})$. Thus the computational cost of the multilevel Monte Carlo estimation is significantly cheaper than that of the single level Monte Carlo estimator $E_N\left[\phi(X_L^T)\right]$ with $N = (1+L)^{-2}2^{4L}$ samples, which in comparison is bounded by $\mathcal{W}_L = \mathrm{O}(2^{(6+\alpha d)L}L^{-2})$, while retaining the same rate of convergence.

4 Summaries of included papers

4.1 Paper 1: *Monte Carlo versus multilevel Monte Carlo in weak error simulations of SPDE approximations*

Consider the analysis of weak errors for fully discrete approximations of solutions to SPDE, that is to say, in the context of Example 3.13, errors of the type $|\,\mathbb{E}[Y - Y_\ell]|$, where ϕ is some given functional, $Y = \phi(X(T))$ and Y_ℓ of $(Y_\ell)_{\ell\in\mathbb{N}_0}$ is given by $Y_\ell = \phi(X_\ell^T)$ for each level $\ell \in \mathbb{N}_0$. This topic has been investigated in the community of numerical analysis of SPDE for some time. Yet, simulations that illustrate the theoretical results of such investigations are rarely available. Furthermore, while weak convergence results for equations driven by additive noise exist, cf. Section 3.4, theoretical results for the case of *multiplicative noise* are still work in progress for fully discrete finite element discretizations (see, however, Paper 5 for the special case of quadratic test functions). In these cases, simulations of weak convergence rates can inform us about the plausibility of claims on the rate.

One reason for the lack of simulations in the literature is the computational expense of simulating a solution to an SPDE, which must be repeated a large number of times when using a Monte Carlo method to approximate the expectation that is part of the weak error. Due to this computational complexity, it is important to carefully consider which Monte Carlo method one chooses in order to accurately simulate weak error rates. In Paper 1 we

present four methods of simulating such rates and analyze the additional error caused by the Monte Carlo approximation involved in each of them.

The analysis is done for the more general problem of approximating the quantity $|\,\mathbb{E}[Y - Y_\ell]|$, where $(Y_\ell)_{\ell\in\mathbb{N}_0}$ is a sequence of mean square integrable random variables converging to $Y \in L^2(\Omega,\mathbb{R})$. If one were interested in estimating $\mathbb{E}[Y - Y_\ell]$, the method of *common random numbers* would tell us that when Y and Y_ℓ are positively correlated, which is reasonable to assume in the case that the latter random variable is an approximation of the former, an estimator of the form $\mathbb{E}[Y] - E_N[Y_\ell]$ is outperformed by $E_N[Y - Y_\ell]$, since the former has higher variance and both are unbiased. Now, when estimating $|\,\mathbb{E}[Y - Y_\ell]|$, the estimators $|\,\mathbb{E}[Y] - E_N[Y_\ell]|$ and $|E_N[Y - Y_\ell]|$ are in general biased, so a direct comparison cannot be made. Instead we show that the mean squared error of the former estimator is bounded from below by

$$\big\|\,|\,\mathbb{E}[Y - Y_\ell]| - |\,\mathbb{E}[Y] - E_N[Y_\ell]|\big\|_{L^2(\Omega,\mathbb{R})} \geq -|\,\mathbb{E}[Y - Y_\ell]| + \big(|\,\mathbb{E}[Y - Y_\ell]|^2 + N^{-1}\,\mathsf{Var}[Y_\ell]\big)^{1/2}$$

and from above by

$$\big\|\,|\,\mathbb{E}[Y - Y_\ell]| - |\,\mathbb{E}[Y] - E_N[Y_\ell]|\big\|_{L^2(\Omega,\mathbb{R})} \leq N^{-1/2}(\mathsf{Var}[Y_\ell])^{1/2}.$$

For the latter estimator, the corresponding bounds are shown to be

$$\begin{aligned}&\big\|\,|\,\mathbb{E}[Y - Y_\ell]| - |E_N[Y - Y_\ell]|\big\|_{L^2(\Omega,\mathbb{R})}\\ &\quad\geq -|\,\mathbb{E}[Y - Y_\ell]| + \big(|\,\mathbb{E}[Y - Y_\ell]|^2 + N^{-1}\,\mathsf{Var}[Y - Y_\ell]\big)^{1/2}\end{aligned}$$

and

$$\big\|\,|\,\mathbb{E}[Y - Y_\ell]| - |E_N[Y - Y_\ell]|\big\|_{L^2(\Omega,\mathbb{R})} \leq N^{-1/2}(\mathsf{Var}[Y - Y_\ell])^{1/2}.$$

Therefore, under the assumption that the quantity of interest $|\,\mathbb{E}[Y - Y_n]| \ll N^{-1/2}$ is very small, which is usually the case in the context of weak error simulations of SPDE approximations, the former estimator will behave like $N^{-1/2}(\mathsf{Var}[Y_\ell])^{1/2}$ and the latter like $N^{-1/2}(\mathsf{Var}[Y - Y_\ell])^{1/2}$. So if Y and Y_ℓ have a positive correlation, which they in general do in such simulations, the additional error of the latter estimator will be significantly smaller than that of the former.

In addition to this, the additional error caused by estimating $|\,\mathbb{E}[Y - Y_L]|$ with a multilevel Monte Carlo estimator $|\,\mathbb{E}[Y] - E^L[Y_L]|$ is analyzed. We find that the mean squared error of this estimator is bounded from below by

$$\begin{aligned}&\big\|\,|\,\mathbb{E}[Y - Y_L]| - |\,\mathbb{E}[Y] - E^L[Y_L]|\big\|_{L^2(\Omega,\mathbb{R})}\\ &\quad\geq -|\,\mathbb{E}[Y - Y_L]| + \left(|\,\mathbb{E}[Y - Y_L]|^2 + N_0^{-1}\,\mathsf{Var}[Y_0] + \sum_{\ell=1}^{L} N_\ell^{-1}\,\mathsf{Var}[Y_\ell - Y_{\ell-1}]\right)^{1/2}\end{aligned}$$

and from above by

$$\big\|\,|\,\mathbb{E}[Y - Y_L]| - |\,\mathbb{E}[Y] - E^L[Y_L]|\big\|_{L^2(\Omega,\mathbb{R})} \leq \left(N_0^{-1}\,\mathsf{Var}[Y_0] + \sum_{\ell=1}^{L} N_\ell^{-1}\,\mathsf{Var}[Y_\ell - Y_{\ell-1}]\right)^{1/2}.$$

When choosing the sample sizes similarly to the choice made in Theorem 3.12, it turns out that

$$\|\,|\,\mathbb{E}[Y - Y_L]| - |\,\mathbb{E}[Y] - E^L[Y_L]|\|_{L^2(\Omega,\mathbb{R})} \simeq |\,\mathbb{E}[Y - Y_L]|,$$

which is to say that the additional error coming from the Monte Carlo method will asymptotically not affect the observed rate of the weak error simulations. For completeness, the additional error caused by the multilevel Monte Carlo estimator $|E^L[Y - Y_L]|$ is also analyzed, although one should note that this estimator is of no practical interest. This is due to the fact that $E_{N_0}[Y - Y_0]$ has to be computed, i.e., many samples of the exact solution must be generated, which destroys the idea of multilevel Monte Carlo methods.

An attempt is then made to simulate weak error rates using these estimators for the one-dimensional stochastic heat equation driven by multiplicative Wiener noise in H, i.e.,

$$\begin{aligned} \mathrm{d}X(t) &= AX(t)\,\mathrm{d}t + G(X(t))\,\mathrm{d}W(t), \\ X(0) &= X_0, \end{aligned}$$

where $H = L^2(\mathcal{D})$ and $A = \Delta$ with two examples of operators G. The equation is discretized by a finite element approximation in space (Example 3.5) and the backward Euler scheme in time (Example 3.7). Due to the large sample sizes involved, this simulation is computationally highly expensive and was therefore performed on a cluster at Chalmers Centre for Computational Science and Engineering (C3SE). We observe that the estimators $E_N[Y - Y_\ell]$ and $|\,\mathbb{E}[Y] - E^L[Y_L]|$, where Y is replaced by a reference solution, outperform $\mathbb{E}[Y] - E_N[Y_\ell]$ in the sense that the simulated weak error rate more closely resembles the prediction of the rule of thumb. Moreover, we simulate the error rates also for the simpler case of approximating a geometric Brownian motion. The lower computational costs of this allowed for finer simulations, which illustrates the theoretical bounds in an even clearer way.

4.2 Paper 2: *Mean-square stability analysis of approximations of stochastic differential equations in infinite dimensions*

Let us assume that SPDE (1) is linear and autonomous, i.e., that the mappings F and G do not depend on $t \in [0, T]$, and that F and G fulfil $F \in \mathcal{L}(H)$ and $G \in \mathcal{L}(H, \mathcal{L}(U, H))$. A property of interest of such SPDE is the qualitative behaviour of the second moment of the solution to (1). This is commonly analyzed in terms of the *equilibrium* or *zero solution* $(X_{\mathrm{e}}(t) = 0)_{t\geq 0}$ which is called *mean square stable* if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\mathbb{E}[\|X(t)\|_H^2] < \varepsilon$ for all $t \geq 0$ whenever $\mathbb{E}[\|X_0\|_H^2] < \delta$. It is called *asymptotically mean square stable* if it is mean square stable and there exists $\delta > 0$ such that $\mathbb{E}[\|X_0\|_H^2] < \delta$ implies $\lim_{t\to\infty} \mathbb{E}[\|X(t)\|_H^2] = 0$.

While the main focus of the analysis of approximations of solutions to (1) has been on showing strong and weak convergence, cf. Theorem 3.10 and Theorem 3.11, these properties do not guarantee that the approximation shares the same (asymptotic) mean square stability properties as the analytical solution. The goal of Paper 2 is to generalize the existing theory of asymptotic mean square stability analysis of approximations to the solutions

of finite-dimensional stochastic differential equations to more general approximations, such as the ones introduced in Section 3.3. An important application of mean square stability can be found in multilevel Monte Carlo methods. If the solution is mean square unstable on any of the included levels, this is enough for the estimator to not behave as it should, see, e.g., [1].

The goal of the first part of the paper is the analysis of the asymptotic mean square stability of the *general* linear recursion scheme

$$\begin{aligned} X_h^{j+1} &= D_{\Delta t,h}^{\mathrm{det}} X_h^j + D_{\Delta t,h}^{\mathrm{stoch},j} X_h^j, \\ X_h^0 &= X_h^0, \end{aligned} \tag{17}$$

for $j \in \mathbb{N}_0$, i.e., we do not explicitly assume that the $\mathcal{L}(V_h)$-valued operators $D_{\Delta t,h}^{\mathrm{det}}$ and $D_{\Delta t,h}^{\mathrm{stoch},j}$ are given by (8) and (10) (still, not having an index j in $D_{\Delta t,h}^{\mathrm{det}}$ is motivated by the fact that we apply our results to approximations with autonomous F). The $\mathcal{F}_0$-measurable initial condition X_h^0 is assumed to be square integrable. For this scheme, an *equilibrium* (solution) is given by the zero solution, which is defined as $X_{h,\mathrm{e}}^j = 0$ for all $j \in \mathbb{N}_0$. It is called *mean square stable* if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\mathbb{E}[\|X_h^j\|_H^2] < \varepsilon$ for all $j \in \mathbb{N}_0$ whenever $\mathbb{E}[\|X_h^0\|_H^2] < \delta$ and *asymptotically mean square stable* if it is mean square stable and there exists $\delta > 0$ such that $\mathbb{E}[\|X_h^0\|_H^2] < \delta$ implies $\lim_{j\to\infty} \mathbb{E}[\|X_h^j\|_H^2] = 0$. When there is no risk of confusion, the recursion scheme is itself said to be (asymptotically) mean square stable when this holds.

First, the asymptotic mean square stability is analyzed under the general assumption that the family $(D_{\Delta t,h}^{\mathrm{stoch},j}, j \in \mathbb{N}_0)$ is $\mathcal{F}$-compatible in the sense that the random operator $D_{\Delta t,h}^{\mathrm{stoch},j}$ is $\mathcal{F}_{t_{j+1}}$-measurable and $\mathbb{E}[D_{\Delta t,h}^{\mathrm{stoch},j} | \mathcal{F}_{t_j}] = 0$ for all $j \in \mathbb{N}_0$. It is also assumed that, for all $j \in \mathbb{N}_0$,

$$\|D_{\Delta t,h}^{\mathrm{stoch},j}\|_{L^2(\Omega,\mathcal{L}(V_h))} = \mathbb{E}[\|D_{\Delta t,h}^{\mathrm{stoch},j}\|_{\mathcal{L}(V_h)}^2]^{1/2} < \infty$$

and

$$\mathbb{E}\left[D_{\Delta t,h}^{\mathrm{stoch},j} \otimes D_{\Delta t,h}^{\mathrm{stoch},j} \,\middle|\, \mathcal{F}_{t_j} \right] = \mathbb{E}\left[D_{\Delta t,h}^{\mathrm{stoch},j} \otimes D_{\Delta t,h}^{\mathrm{stoch},j} \right].$$

This is a natural assumption that is true when $D_{\Delta t,h}^{\mathrm{stoch},j}$ is given by (10), since L is a square integrable martingale. Next, under the additional assumption that $(D_{\Delta t,h}^{\mathrm{stoch},j}, j \in \mathbb{N}_0)$ has constant covariance, i.e., that for all $j \in \mathbb{N}_0$,

$$\mathbb{E}\left[D_{\Delta t,h}^{\mathrm{stoch},j} \otimes D_{\Delta t,h}^{\mathrm{stoch},j} \right] = \mathbb{E}\left[D_{\Delta t,h}^{\mathrm{stoch},0} \otimes D_{\Delta t,h}^{\mathrm{stoch},0} \right],$$

it is shown that the zero solution of (17) is asymptotically mean square stable if and only if the *stability operator*

$$\mathcal{S} = D_{\Delta t,h}^{\mathrm{det}} \otimes D_{\Delta t,h}^{\mathrm{det}} + \mathbb{E}[D_{\Delta t,h}^{\mathrm{stoch},0} \otimes D_{\Delta t,h}^{\mathrm{stoch},0}] \in \mathcal{L}(V_h^{(2)})$$

satisfies $\rho(\mathcal{S}) = \max_{i=1,\ldots,N_h^2} |\lambda_i| < 1$, where $\lambda_1, \ldots, \lambda_{N_h^2}$ are the eigenvalues of $\mathcal{S}$.

The second part of the paper treats the asymptotic mean square stability of (17) assuming that it does approximate the mild solution to (1). It is shown that when $D_{\Delta t,h}^{\mathrm{det}}$ and $D_{\Delta t,h}^{\mathrm{stoch},j}$ are given by (8) and (10), the stability operator simplifies to

$$\mathcal{S} = D_{\Delta t,h}^{\mathrm{det}} \otimes D_{\Delta t,h}^{\mathrm{det}} + \Delta t\,(C \otimes C)q \in \mathcal{L}(V_h^{(2)}),$$

where $q = \sum_{k=1}^{\infty} \mu_k q_k \otimes q_k \in U^{(2)}$ and $C \in \mathcal{L}(U, \mathcal{L}(V_h))$ with

$$Cu = r_{\mathrm{d}}^{-1}(\Delta t A_h) P_h G(\cdot) u,$$

recalling that the eigenvectors of Q are given by $(q_i)_{i\in\mathbb{N}}$ and the eigenvalues by $(\mu_i)_{i\in\mathbb{N}}$. A similar result is shown for the higher order Milstein scheme, the convergence of which was analyzed in [6].

In the remainder of the paper we derive sufficient conditions for the asymptotic mean square stability of (17) when $D_{\Delta t,h}^{\mathrm{det}}$ and $D_{\Delta t,h}^{\mathrm{stoch},j}$ are given by (8) and (10), including the backward and forward Euler scheme of Example 3.7. For example, the first of these is seen to be asymptotically mean square stable if

$$\frac{(1 + \Delta t \|F\|_{\mathcal{L}(H)})^2 + \Delta t\, \mathrm{Tr}(Q) \|G\|^2_{\mathcal{L}(H,\mathcal{L}(U,H))}}{(1 + \Delta t \lambda_{h,1})^2} < 1,$$

where $\lambda_{h,1}$ is the smallest eigenvalue of the discrete operator $-A_h$. These conditions are based on the observation that $\rho(\mathcal{S}) \leq \|\mathcal{S}\|_{\mathcal{L}(V_h)}$. Using these results, a condition that ensures the asymptotic mean square stability of both the zero solution to (1) and its approximation with the backward Euler scheme is derived under assumptions that are applicable when, for example, V_h is given by Example 3.5 or 3.6. A similar result is again shown for the Milstein scheme. Simulations using both spectral and finite element Galerkin methods illustrate the theoretical results.

4.3 Paper 3: *Rapid covariance-based sampling of linear SPDE approximations in the multilevel Monte Carlo method*

The goal of this paper is to efficiently compute approximations of the final time solution $X(T)$ of the SPDE

$$\begin{aligned} \mathrm{d}X(t) &= \big(AX(t) + F(t, X(t))\big)\,\mathrm{d}t + G(t)\,\mathrm{d}W(t), \\ X(0) &= X_0, \end{aligned}$$

on H in the special case that $F(t,\cdot)$ is an affine linear mapping for all $t \in [0,T]$, i.e., that there exist an operator $F_t^1 \in \mathcal{L}(H, \dot{H}^{-1})$ and an element $F_t^2 \in \dot{H}^{-1}$ such that $F(t,f) = F_t^1 f + F_t^2$ for all $f \in H$. Here we assume that $A = -\Lambda$, where Λ is a densely defined, positive definite operator on H with a compact inverse, generating a sequence $(\dot{H}^r)_{r\in\mathbb{R}}$ of domains of spectral powers of Λ. We also assume that both $F(\cdot,u)$ and $G(\cdot)$ are Hölder continuous

with exponent 1/2 for all $u \in H$ and the process W is allowed to be a generalized Wiener process in H. This setting allows for the SPDE in Example 3.2.

The equation is discretized on a space $V_h \subset \dot{H}^1$, for example the finite element space of Example 3.5, and the backward Euler scheme (12) is used for the discretization in time. The resulting approximation $X_{h,\Delta t} = (X_{h,\Delta t}^{t_j})_{j=0,\dots,N_{\Delta t}}$ can then, with $R_{h,\Delta t} = (I - \Delta t A_h)$, $F_{h,\Delta t}^{1,j} = \left(I + \Delta t P_h F_{t_j}^1\right)$, and $F_{h,\Delta t}^{2,j} = \Delta t P_h F_{t_j}^2$, be written as

$$R_{h,\Delta t} X_{h,\Delta t}^{t_{j+1}} = F_{h,\Delta t}^{1,j} X_{h,\Delta t}^{t_j} + F_{h,\Delta t}^{2,j} + P_h G(t_j) \Delta W^j$$

for $j = 0, \dots, N_{\Delta t} - 1$. The key observation here is that since F is affine linear and ΔW^j is Gaussian, the distribution of $X_{h,\Delta t}^T$ is uniquely determined by its expected value and covariance. Based on the derivation of the stability operator $\mathcal{S}$ in the previous paper, we derive a recursion scheme for $\mathsf{Cov}(X_{h,\Delta t}^{t_{j+1}})$, denoted by $\Sigma^{t_{j+1}}$, as

$$\left(R_{h,\Delta t}\right)^{\otimes 2} \Sigma^{t_{j+1}} = \left(F_{h,\Delta t}^{1,j}\right)^{\otimes 2} \Sigma^{t_j} + \mathbb{E}\left[\left(P_h G(t_j) \Delta W^j\right)^{\otimes 2}\right]$$

for $j = 0, \dots, N_{\Delta t} - 1$, where $K_h^{\otimes 2}$ denotes the tensor product $K_h \otimes K_h$ of an operator $K_h \in \mathcal{L}(V_h)$ with itself. With this we can therefore obtain a covariance-based Monte Carlo scheme as an alternative to the path-based Monte Carlo scheme of Section 3.5. This is to say, to approximate $\mathbb{E}[\phi(X(T))]$ for a given functional ϕ, we generate samples of $X_{h,\Delta t}^T$ using Σ^T instead of computing the entire path $X_{h,\Delta t}$ several times.

We then show that a similar covariance-based scheme can be derived for the multilevel Monte Carlo method. Recall that we in each summand of (15) need a pair $(X_{h,\Delta t}, X_{h',\Delta t'})$ of approximations of X for a set $(h', \Delta t')$ of coarse and a set $(h, \Delta t)$ of fine mesh sizes and time step sizes. Assuming that the coarse time step size $\Delta t'$ is a multiple of the fine time step size Δt, we can then create an extension $X_{h',\Delta t} = (X_{h',\Delta t}^{t_j})_{j=0}^{N_{\Delta t}}$ of the coarse approximation on the fine time grid, and then rephrase the two recursion schemes as a single scheme

$$\begin{bmatrix} \hat{R}_{h',\Delta t}^j & 0 \\ 0 & R_{h,\Delta t}^j \end{bmatrix} \begin{bmatrix} X_{h',\Delta t}^{t_{j+1}} \\ X_{h,\Delta t}^{t_{j+1}} \end{bmatrix} = \begin{bmatrix} \hat{F}_{h',\Delta t}^{1,j} & 0 \\ 0 & F_{h,\Delta t}^{1,j} \end{bmatrix} \begin{bmatrix} X_{h',\Delta t}^{t_j} \\ X_{h,\Delta t}^{t_j} \end{bmatrix} + \begin{bmatrix} \hat{F}_{h',\Delta t}^{2,j} \\ F_{h,\Delta t}^{2,j} \end{bmatrix} + \begin{bmatrix} P_{h'} \hat{G}(t_j) \\ P_h G(t_j) \end{bmatrix} \Delta W^j$$

for $j = 0, 1, \dots, N_{\Delta t} - 1$ on $V_{h'} \oplus V_h$, where $\hat{R}_{h',\Delta t}^j, \hat{F}_{h',\Delta t}^{1,j}, \hat{F}_{h',\Delta t}^{2,j}$ and $\hat{G}(t_j)$ are suitable extensions of $R_{h,\Delta t}$, $F_{h,\Delta t}^{1,j}$, $F_{h,\Delta t}^{2,j}$ and $P_h G(t_j)$, respectively. We use the extended operators to write a recursion scheme for the cross-covariance Σ_{t_j+1} of $X_{h',\Delta t}^{t_{j+1}}$ and $X_{h,\Delta t}^{t_{j+1}}$ as

$$\left(\hat{R}_{h',\Delta t'}^j \otimes R_{h,\Delta t}\right) \Sigma_{t_{j+1}} = \left(\hat{F}_{h,\Delta t}^{1,j} \otimes F_{h,\Delta t}^{1,j}\right) \Sigma_{t_j} + \mathbb{E}\left[P_{h'} \hat{G}(t_j) \Delta W^j \otimes P_h G(t_j) \Delta W^j\right]$$

for $j = 0, 1, \dots, N_{\Delta t} - 1$. The distribution of the pair $\left(X_{h',\Delta t'}^T, X_{h,\Delta t}^T\right)$ is, in the same way as in the standard Monte Carlo setting, uniquely determined by the covariances of the two discretizations along with the means and the cross-covariance, and therefore we can

design a covariance-based multilevel Monte Carlo scheme based on the computation of these quantities.

Restricting the setting to that of Assumption 3.8, we use a result similar to Theorem 3.12 to deduce that, under reasonable assumptions on the complexity of the finite element algorithms and the generation of the Wiener process, the covariance-based Monte Carlo algorithm outperforms the standard Monte Carlo algorithm and the covariance-based multilevel Monte Carlo algorithm outperforms the standard multilevel Monte Carlo algorithm. We perform numerical simulations in dimension $d = 1$ which are consistent with the claims made on the computational complexity of the considered algorithms.

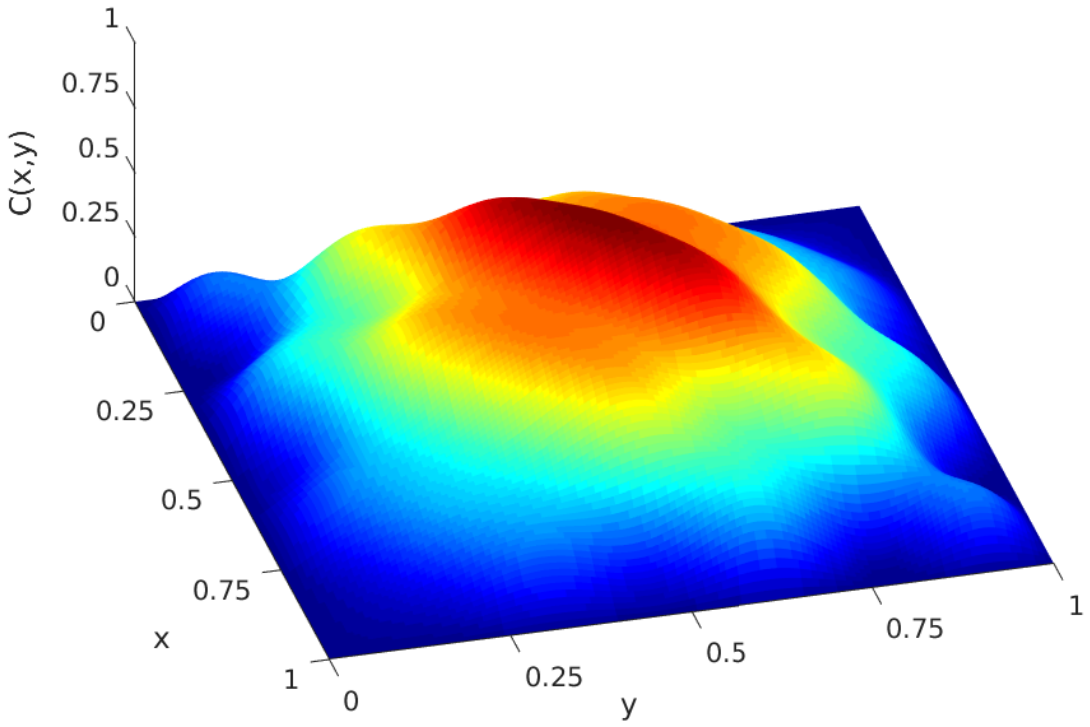


Figure 4: Estimate of the covariance function of $u(T)$, where u is the solution to the stochastic wave equation in Example 2.4.

To summarize, in this paper we deduced a method for the estimation of $\mathsf{Cov}(X(T))$ based on the computation of $\Sigma^T = \mathsf{Cov}(X^T_{h,\Delta t})$ and noted that, under certain assumptions, it can be computationally cheaper to sample from the covariance instead of computing the entire path if the quantity of interest is given by $\mathbb{E}[\phi(X(T))]$. The quantity $\mathsf{Cov}(X(T))$ is interesting in and of itself, since we, by the identification $H \otimes H = \mathcal{L}_2(H) = L^2(\mathcal{D} \times \mathcal{D}, \mathbb{R})$,

can think of it as the covariance function C of $X(T)$, given by

$$C(x,y) = \sum_{i,j=1}^{\infty} \langle \mathsf{Cov}(X(T))e_i, e_j \rangle_H \, e_i(x) e_j(y) \tag{18}$$

for $x, y \in \mathcal{D}$, where $(e_i)_{i\in\mathbb{N}}$ is an orthonormal basis of H. One can show that the error $\| \mathsf{Cov}(X_{h,\Delta t}^T) - \mathsf{Cov}(X(T))\|_{H\otimes H}$ can be bounded by the strong error $\|X_{h,\Delta t}^T - X(T)\|_{L^2(\Omega;H)}$ and we can obtain a visual representation of an approximation of C by an expansion of $\mathsf{Cov}(X_{h,\Delta t}^T)$ on a basis of $V_h^{(2)}$, similarly to (18). As long as the considered equation is linear and we have discretized it via a fully discrete scheme like (7), we can find a visual representation of an approximation of C for any SPDE. As an example, we show in Figure 4 an estimate of the covariance function of the solution to the stochastic wave equation of Example 2.4 and Figure 3(a) evaluated at $T = 5$. The discretization was accomplished by a finite element method in space and a Crank–Nicolson method in time. The study of the resulting weak error is the topic of the next paper of this thesis.

4.4 Paper 4: *Weak convergence of fully discrete finite element approximations of semilinear hyperbolic SPDE with additive noise*

The topic of Paper 4 is the discretization of the stochastic wave equation

$$\mathrm{d}\dot{u}(t) - \Delta u(t)\,\mathrm{d}t = F(t,u(t))\,\mathrm{d}t + G(t)\,\mathrm{d}W(t)$$

with additive Wiener noise in $H = L^2(\mathcal{D})$ for $t \in (0,T]$, $T < \infty$, i.e., a more general form of the equation of Example 3.3. We are mainly interested in the setting when $u \mapsto F(\cdot,u)$ is nonlinear, in which case the equation is said to be semilinear. As in Example 3.3, the equation is rewritten in the abstract form

$$\mathrm{d}X(t) = AX(t)\,\mathrm{d}t + BF(t,P^1X(t))\,\mathrm{d}t + BG(t)\,\mathrm{d}W(t) \tag{19}$$

on $\mathcal{H}$, where P^1 denotes the projection onto the first component of $v = [v_1, v_2]^\top \in \mathcal{H}$, i.e., $P^1 v = v_1$.

With Λ_h as in Section 3.2 and V_h being the finite element space in Example 3.5 (but consider general polynomials of degree $\kappa - 1$, $\kappa \in \{2,3\}$), we let

$$A_h = \begin{bmatrix} 0 & I \\ -\Lambda_h & 0 \end{bmatrix}$$

be a discrete counterpart to A on the product space $\mathcal{V}_h = V_h \oplus V_h$, equipped with the same inner product as $\mathcal{H}$. Then A_h generates a C_0-semigroup E_h of contractions on $\mathcal{V}_h$. With $D_{\Delta t,h}^{\mathrm{det},j}$ given by (9) and $D_{\Delta t,h}^{\mathrm{stoch},j}$ by (11), we consider the fully discrete approximation

$$\begin{aligned} X_{h,\Delta t}^{t_{j+1}} &= D_{\Delta t,h}^{\mathrm{det},j} X_{h,\Delta t}^{t_j} + D_{\Delta t,h}^{\mathrm{stoch},j} \\ &= R(\Delta t A_h)\left(X_{h,\Delta t}^{t_j} + P_h BF(t_j, P^1 X_{h,\Delta t}^{t_j})\Delta t + P_h BG(t_j)\Delta W^j \right) \end{aligned} \tag{20}$$

for $j = 0, \dots, N_{\Delta t} - 1$, where $R(\Delta t A_h)$ denotes a rational approximation of order ρ of the semigroup E_h. The main goal of the paper is to deduce weak convergence rates of (20). This generalizes results of [13] to the semilinear, non-autonomous setting. For this, we use results on the error of the semigroup approximation from [14], which are only available for the first components $P^1 X_{h,\Delta t}^{t_j}$ and $P^1 X(t) = u(t)$ of the approximation and the solution. Therefore, we restrict ourselves to deriving convergence results for the first component only.

Specifically, we assume, with $\Lambda = -\Delta$ and the sequence $(\dot{H}^r)_{r\in\mathbb{R}}$ as in Example 2.1, that there exist parameters $\beta, \eta, \delta \geq 0$ and $\theta \leq \min(\beta, \delta, 1)$ and a constant $C > 0$ such that the data in (19) fulfils the following requirements:

(i) The mapping $G\colon [0,T] \to \mathcal{L}_2(Q^{1/2}(H), \dot{H}^{\beta-1})$ satisfies

$$\|\Lambda^{\frac{\beta-1}{2}} \left(G(t_1) - G(t_2)\right)\|_{\mathcal{L}_2^0} \leq C|t_1 - t_2|^\eta$$

for all $t_1, t_2 \in [0,T]$ and $\|\Lambda^{\frac{\beta-1}{2}} G(t)\|_{\mathcal{L}_2^0} \leq C$ for some $t \in [0,T]$.

(ii) The function $F\colon [0,T] \times \dot{H}^0 \to \dot{H}^0$ satisfies

$$\|\Lambda^{-\frac{1}{2}} \left(F(t,u) - F(t,v)\right)\|_{\dot{H}^0} \leq C\|u - v\|_{\dot{H}^0}$$

for all $t \in [0,T]$ and $u, v \in \dot{H}^0$,

$$\|\Lambda^{\frac{\alpha}{2}} F(t,u)\|_{\dot{H}^0} \leq C\left(1 + \|\Lambda^{\frac{\alpha}{2}} u\|_{\dot{H}^0}\right)$$

for all $t \in [0,T]$, $u \in \dot{H}^\alpha$ and $\alpha \in \{0, \theta\}$ and

$$\|\Lambda^{-\frac{1}{2}} \left(F(t,u) - F(s,u)\right)\|_{\dot{H}^0} \leq C\left(1 + \|u\|_{\dot{H}^0}\right)|t - s|^\eta$$

for all $s, t \in [0,T]$ and $u \in \dot{H}^0$.

(iii) The initial value $X_0 \in \mathcal{H}^\delta$ is deterministic.

With these assumptions in place, we deduce a strong convergence result, i.e., that for any $p \in [1, \infty)$, there exists a constant $C > 0$ such that for all $h, \Delta t \in (0, 1]$

$$\sup_{n\in\{0,1,\dots,N_{\Delta t}\}} \|P^1 X_{h,\Delta t}^{t_n} - P^1 X(t_n)\|_{L^p(\Omega,\dot{H}^0)} \leq C\left(h^{r\frac{\kappa}{\kappa+1}} + \Delta t^{\min(r\frac{\rho}{\rho+1},\eta,1)}\right),$$

where $r = \min(\beta, \delta, 1)$. Assuming also a differentiability condition on F, namely that there exist parameters $\mu \in [0,2]$, $\nu \in [\max(\mu - 1, 0), \min(r, 1)]$ and a constant $C > 0$ such that for every $t \in [0,T]$, $F(t,\cdot) \in \mathcal{G}_{\mathrm{p}}^1(\dot{H}^0, \dot{H}^{-\min(\mu,1)})$ and

$$\|\Lambda^{-\frac{\mu}{2}} F'(t,u)v\|_{\dot{H}^0} \leq C\left(1 + \|\Lambda^{\frac{\nu}{2}} u\|_{\dot{H}^0}\right)\|\Lambda^{-\frac{\nu}{2}} v\|_{\dot{H}^0}$$

for all $u \in \dot{H}^\nu$ and $v \in \dot{H}^{-\nu}$, we are also able to prove a strong convergence result in a negative norm, i.e., that for any $p \in [1, \infty)$, there exists a constant $C > 0$ such that for all $h, \Delta t \in (0, 1]$

$$\sup_{n \in \{0,1,\dots,N_{\Delta t}\}} \|P^1 X_{h,\Delta t}^{t_n} - P^1 X(t_n)\|_{L^p(\Omega, \dot{H}^{-\nu})} \leq C\big(h^{r' \frac{\kappa}{\kappa+1}} + \Delta t^{\min(r' \frac{\rho}{\rho+1}, \eta, 1)}\big),$$

where $r' = \min(\max(2\nu, \beta), \max(2\nu, 1+\theta), \delta)$. The assumptions on the nonlinearity F are in particular applicable when F is the natural operator that, for $u \in H = L^2(\mathcal{D})$, is given by $F(u)(x) = f(u(x))$ for a.e. $x \in \mathcal{D}$. Here $f \colon \mathbb{R} \to \mathbb{R}$ is a differentiable function such that, for a constant $C > 0$, $|f(x)| \leq C(1 + |x|)$, $|f'(x)| \leq C$ and $|f'(x) - f'(y)| \leq C|x - y|$ for all $x, y \in \mathbb{R}$.

Assuming further that the test function $\phi \in \mathcal{G}_{\mathrm{p}}^2(\dot{H}^0, \mathbb{R})$, that $G(t) = I$ for all $t \in [0, T]$, that $\|\Lambda^{\beta - \frac{1}{2}} Q \Lambda^{-\frac{1}{2}}\|_{\mathcal{L}_1(\dot{H}^0)} < \infty$ (this is the same assumption as in [13]) then we are able to deduce a weak convergence result. We show that for $\mu \leq 1$, there exists a constant $C > 0$ such that, for all $h, \Delta t \in (0, 1]$,

$$\big|\mathbb{E}\left[\phi(P^1 X_{h,\Delta t}^T) - \phi(P^1 X(T))\right]\big| \leq C\big(h^{r' \frac{\kappa}{\kappa+1}} + \Delta t^{\min(r' \frac{\rho}{\rho+1}, \eta, 1)}\big).$$

If, on the other hand, $1 < \mu \leq 2$, we show that there exists a constant $C > 0$ such that, for all $h, \Delta t \in (0, 1]$,

$$\big|\mathbb{E}\left[\phi(P^1 X_{h,\Delta t}^T) - \phi(P^1 X(T))\right]\big| \leq C\big(h^{r' \frac{\kappa}{\kappa+1} + 1 - \mu} + h^{1-\mu} \Delta t^{\min(r' \frac{\rho}{\rho+1}, \eta, 1)}\big).$$

These two cases of μ correspond, with $\mathcal{D} \subset \mathbb{R}^d$, to $d = 1$ and $d = 2$. To the best of our knowledge, this is the first weak convergence result for a semilinear stochastic wave equation applicable to the case $d = 2$. To deduce the results, we use similar techniques as the author of [27] but instead of using Kolmogorov's equation and the Itô formula, which is the common way of deducing weak convergence rates, we complete the analysis by means of so called Malliavin calculus. The reason for why we cannot use the aforementioned techniques is that, in contrast to [27], we do not assume that the discretization is given by a so called trigonometric integrator. There are situations when such integrators could be better suited, such as highly oscillatory data, but for complicated domain geometries the algorithms in this paper could be more advantageous from an implementational point of view, since they do not require any knowledge of the eigenfunctions of Δ or its discrete counterpart.

Simulations using both white noise ($Q = I$, see Example 2.6) and trace class noise given by a covariance kernel as in Section 2.3 serve as illustrations of the theoretical results.

4.5 Paper 5: *Finite element approximation of Lyapunov equations for the computation of quadratic functionals of SPDE*

In the analysis of the asymptotic mean square stability of SPDE schemes as in Paper 2, the quadratic functional $\phi(\cdot) = \|\cdot\|_H^2$ naturally plays an important role. In practice it is

also often used to illustrate weak convergence results, which we have done in Papers 1, 3 and 4. Therefore, it is of interest to 1) deduce rates on the weak error

$$\left|\mathbb{E}\left[\left\|X_{h,\Delta t}^T\right\|_H^2 - \left\|X(T)\right\|_H^2\right]\right|$$

for a given approximation $X_{h,\Delta t}^T$ of $X(T)$, and 2) find efficient ways of computing approximations of $\mathbb{E}\left[\left\|X(T)\right\|_H^2\right]$. We deal with these two questions in Paper 5.

Specifically, we consider the autonomous multiplicative noise SPDE

$$\mathrm{d}X(t) = AX(t)\ \mathrm{d}t + G(X(t))\ \mathrm{d}W(t) \tag{21}$$

for $t \in (0,T]$ with initial condition $X(0) = X_0$ in a Hilbert space H, driven by a generalized Wiener process W in another Hilbert space U with covariance operator Q. We again assume that $A = -\Lambda$, where Λ is a densely defined, positive definite operator on H with a compact inverse, generating a sequence $(\dot{H}^r)_{r\in\mathbb{R}}$ of domains of spectral powers of Λ. The multiplicative noise operator fulfils $G \in \mathcal{L}(H, \mathcal{L}_2(U_0, \dot{H}^{\beta-1}))$ for some $\beta \in (0,1]$, where $U_0 = Q^{1/2}(U)$. We are interested in approximating not only $\mathbb{E}\left[\left\|X(T)\right\|_H^2\right]$ but, with $R_1, R_0 \in \mathcal{L}(H)$, the more general quantity

$$\Phi(x) = \mathbb{E}\left[\int_0^T \|R_1 X(t)\|_H^2\ \mathrm{d}t + \|R_0 X(T)\|_H^2 \Big| X(0) = x\right] \tag{22}$$

for all initial values $x \in H$, which is a particularly interesting problem in light of the definition of asymptotic mean square stability, cf. Paper 2.

In order to approximate (22), we derive the existence of a solution $K\colon [0,T] \to \mathcal{L}(H)$ to the *Lyapunov equation*

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle K(t)\phi,\psi\rangle_H + \langle \Lambda^{\frac{1}{2}}K(t)\phi, \Lambda^{\frac{1}{2}}\psi\rangle_H + \langle \Lambda^{\frac{1}{2}}K(t)\psi, \Lambda^{\frac{1}{2}}\phi\rangle_H = \langle R_1\phi, R_1\psi\rangle_H + \langle K(t)G\phi, G\psi\rangle_{\mathcal{L}_2^0}$$

with $K(0) = R_0^* R_0$. We show that $\langle K(T)x, x\rangle_H = \Phi(x)$ for $x \in H$ or, more generally, that

$$\langle K(t)x, y\rangle_H = \mathbb{E}\left[\int_0^t \langle R_1 X(s), R_1 Y(s)\rangle_H\, \mathrm{d}\,s + \langle R_0 X(t), R_0 Y(t)\rangle_H \,\Big|\, X_0 = x, Y_0 = y\right] \tag{23}$$

for $t \in [0,T]$ and $x, y \in H$, where Y is another solution to the same SPDE, started at another initial value.

With the finite element space V_h as in Example 3.5, we seek for $n \in \{1, \dots, N_{\Delta t}\}$ a fully discrete approximation $K_{h,\Delta t}^n \in \mathcal{L}(V_h)$ of $K(t_n)$ given by the recursion scheme

$$K_{h,\Delta t}^n - \Delta t A_h K_{h,\Delta t}^n - \Delta t K_{h,\Delta t}^n A_h = K_{h,\Delta t}^{n-1} + \Delta t P_h R_1^* R_1 P_h + \Delta t P_h G^* K_{h,\Delta t}^{n-1} P_h G P_h, \tag{24}$$

with $K_{h,\Delta t}^0 = P_h R_0^* R_0 P_h$.

We show that for all $c > 0$, $\gamma \in (2,4]$ and $\rho \in (0,\beta)$ there exists a constant $C > 0$ such that for all $h \in (0,1)$, $\Delta t \le c h^\gamma$ and $n \in \{1, \dots, N_{\Delta t}\}$

$$\left\|K_{h,\Delta t}^n P_h - K(t_n)\right\|_{\mathcal{L}(H)} \le C t_n^{-\rho-\theta} h^{2\rho}. \tag{25}$$

To the best of our knowledge, this is the first work to establish rigorous convergence rates for a fully discrete numerical approximation of (24). With this, we arrive at a way of approximating $\Phi(x)$ by the quantity $\langle K_{h,\Delta t}^{N_{\Delta t}} P_h x, P_h x\rangle_H$.

Moreover, we again consider a fully discrete approximation $X_{h,\Delta t}$ of (21) given by the backward Euler scheme (12) on V_h. We then derive a decay rate of the quantity

$$\left| \mathbb{E}\left[\left\| R_2 X_{h,\Delta t}^T \right\|_H^2 + \Delta t \sum_{k=1}^{N_{\Delta t}-1} \left\| R_1 X_{h,\Delta t}^{t_{n-k}} \right\|_H^2 \,\middle|\, X_{h,\Delta t}^{t_0} = P_h x \right] - \left\langle K_{h,\Delta t}^{N_{\Delta t}} P_h x, P_h x \right\rangle_H \right|$$

as $h, \Delta t \to 0$, and, along with (25), we deduce from this a kind of weak convergence result with respect to the functional Φ of (22).

Next, we also consider a Monte Carlo method and a covariance based method (similar to that of Paper 3) of approximating $\Phi(x)$ for all $x \in H$, both based on the discretization $X_{h,\Delta t}$ of (21). We describe in detail how these methods, along with (24), can be implemented in matrix form. Using the weak convergence results for $X_{h,\Delta t}$ and the Lyapunov approximation $K_{h,\Delta t} = (K_{h,\Delta t}^n)_{n=0,\dots,N_{\Delta t}}$, we calculate the computational complexity of the three methods and note that the Lyapunov method outperforms the other two methods.

From the representation (23), we can see that the operator $K(T)$ shares many features with a covariance operator, such as it being self-adjoint and positive semidefinite. Like a covariance operator, it admits a representation in the form of an integral operator. There exists a function $\vartheta \in L^2(\mathcal{D} \times \mathcal{D}, \mathbb{R})$ such that

$$\langle K(T)u, v\rangle_H = \int_{\mathcal{D}\times\mathcal{D}} \vartheta(x,y)u(x)v(y)\,\mathrm{d}x\,\mathrm{d}y \tag{26}$$

for all $u, v \in H$. The reason for this is that $K(T) \in \mathcal{L}(\dot{H}^{-r}, H)$ for all $r < 2$, which we show in the paper. This implies that $K(T) \in \mathcal{L}_2(H) = \mathcal{L}(\mathcal{D}\times\mathcal{D}, \mathbb{R})$ for $d = 1, 2, 3$ (cf. Example 2.6). We can approximate ϑ by expanding $K_{h,\Delta t}^{N_{\Delta t}}$ on $V_h^{(2)}$. This is how Figure 5 was obtained, where we show an approximation of ϑ on $\mathcal{D}\times\mathcal{D}$ for $T = 1$. The parameters of the underlying SPDE were chosen as in Figure 3(b) with $R_0 = I$ and $R_1 = 0$ so that

$$\int_{\mathcal{D}\times\mathcal{D}} \vartheta(x,y)u(x)v(y)\,\mathrm{d}x\,\mathrm{d}y = \mathbb{E}\left[\langle X(t), Y(t)\rangle_H \,\middle|\, X_0 = u, Y_0 = v \right].$$

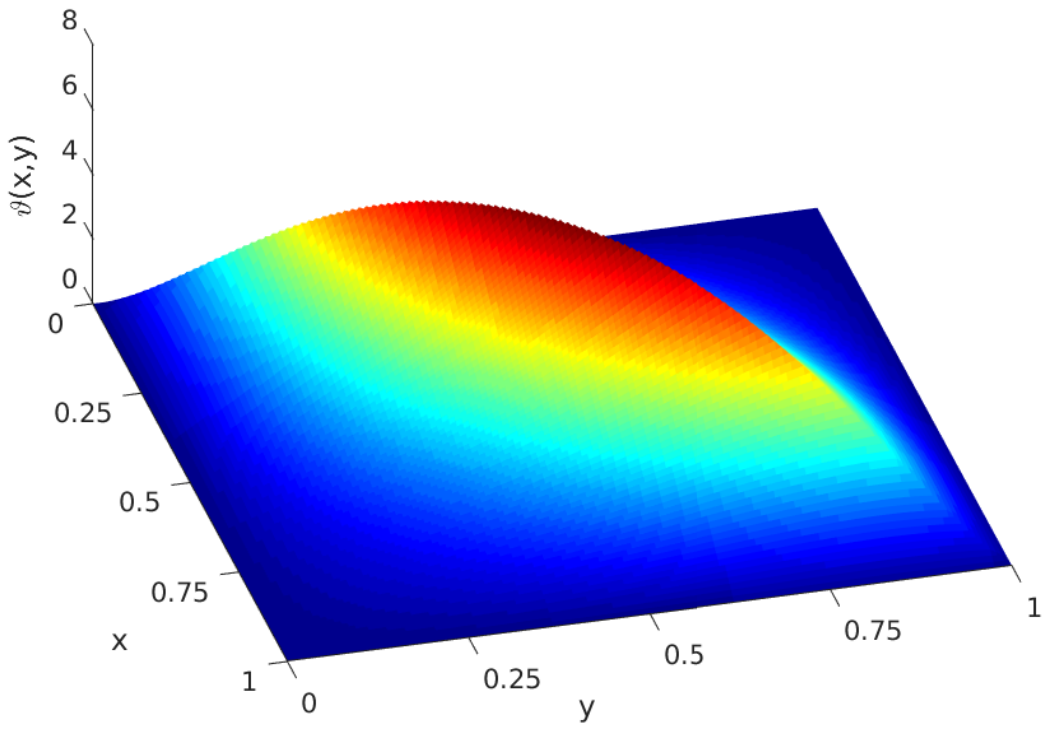


Figure 5: Estimate of the function ϑ in (26) with $R_0 = I, R_1 = 0$ and $T = 1$, for the SPDE in Example 3.4.

References

[1] A. Abdulle and A. Blumenthal. Stabilized multilevel Monte Carlo method for stiff stochastic differential equations. *J. Comput. Phys.*, 251:445–460, 2013.

[2] M. Alns et al. The FEniCS Project version 1.5. *Archive of Numerical Software*, 3(100), 2015.

[3] A. Andersson, R. Kruse, and S. Larsson. Duality in refined Sobolev–Malliavin spaces and weak approximations of SPDE. *Stoch. PDE: Anal. Comp.*, 4(1):113–149, 2016.

[4] A. Andersson, A. Lang, A. Petersson, and L. Schroer. Finite element approximation of Lyapunov equations for the computation of quadratic functionals of SPDE. Preprint at arXiv:1910.05261, 2019.

[5] A. Andersson and S. Larsson. Weak convergence for a spatial approximation of the nonlinear stochastic heat equation. *Math. Comput.*, 85(299):1335–1358, 2016.

[6] A. Barth and A. Lang. Milstein approximation for advection-diffusion equations driven by multiplicative noncontinuous martingale noises. *Appl. Math. Optim.*, 66(3):387–413, 2012.

[7] D. Conus, A. Jentzen, and R. Kurniawan. Weak convergence rates of spectral Galerkin approximations for SPDEs with nonlinear diffusion coefficients. *Ann. Appl. Probab.*, 29(2):653–716, 2019.

[8] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*, volume 152. Cambridge University Press, Cambridge, 2nd edition, 2014.

[9] R. C. Dalang. The stochastic wave equation. In D. Khoshnevisan and F. Rassoul-Agha, editors, *A Minicourse on Stochastic Partial Differential Equations*, pages 39–71, Berlin, Heidelberg, 2009. Springer.

[10] A. Debussche. Weak approximation of stochastic partial differential equations: the nonlinear case. *Math. Comput.*, 80(273):89–117, 2011.

[11] A. Debussche and J. Printems. Weak order for the discretization of the stochastic heat equation. *Math. Comput.*, 78(266):845–863, 2009.

[12] M. Kovács, A. Lang, and A. Petersson. Weak convergence of fully discrete finite element approximations of semilinear hyperbolic SPDE with additive noise. Submitted. Preprint at arXiv:1909.04571, 2019.

[13] M. Kovács, S. Larsson, and F. Lindgren. Weak convergence of finite element approximations of linear stochastic evolution equations with additive noise II. Fully discrete schemes. *BIT Numerical Mathematics*, 413(2):497, 2013.

[14] M. Kovács, F. Lindner, and R. Schilling. Weak convergence of finite element approximations of linear stochastic evolution equations with additive Lévy noise. *SIAM/ASA J. on Uncert. Quant.*, 3(1):1159–1199, 2015.

[15] R. Kruse. *Strong and weak approximation of semilinear stochastic evolution equations*, volume 2093 of *Lecture Notes in Mathematics*. Springer, Cham, 2014.

[16] A. Lang. A note on the importance of weak convergence rates for SPDE approximations in multilevel Monte Carlo schemes. In R. Cools and D. Nuyens, editors, *Monte Carlo and Quasi-Monte Carlo Methods, MCQMC, Leuven, Belgium, April 2014*, volume 163 of *Springer Proceedings in Mathematics & Statistics*, pages 489–505, 2016.

[17] A. Lang and A. Petersson. Monte Carlo versus multilevel Monte Carlo in weak error simulations of SPDE approximations. *Math. Comput. Simulation*, 143:99 – 113, 2018.

[18] A. Lang, A. Petersson, and A. Thalhammer. Mean-square stability analysis of approximations of stochastic differential equations in infinite dimensions. *BIT Numer. Math.*, 57(4):963–990, 2017.

[19] S. Larsson and V. Thomée. *Partial Differential Equations with Numerical Methods*, volume 45 of *Texts in Applied Mathematics*. Springer-Verlag, Berlin, 2003.

[20] F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc.*, 73(4):423–498, 2011.

[21] G. J. Lord, C. E. Powell, and T. Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2014.

[22] S. V. Lototsky and B. L. Rozovsky. *Stochastic Partial Differential Equations*. Springer, Cham, 2017.

[23] S. Peszat and J. Zabczyk. *Stochastic Partial Differential Equations with Lévy Noise. An Evolution Equation Approach*, volume 113 of *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, Cambridge, 2007.

[24] A. Petersson. Rapid covariance-based sampling of linear SPDE approximations in the multilevel Monte Carlo method. To appear in *Monte Carlo and Quasi-Monte Carlo Methods, MCQMC, Rennes, France, July 2018*. Preprint at arXiv:1806.11523v3, 2019.

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

[26] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer, Berlin, Heidelberg, 2nd edition, 2006.

[27] X. Wang. An exponential integrator scheme for time discretization of nonlinear stochastic wave equation. *Journal of Scientific Computing*, 64(1):234–263, Jul 2015.

[28] A. Yagi. H^∞-functional calculus and characterization of domains of fractional powers. In T. Ando, R. E. Curto, I. B. Jung, and W. Y. Lee, editors, *Recent Advances in Operator Theory and Applications*, pages 217–235, Basel, 2008. Birkhäuser.