Thesis for the Degree of Doctor of Philosophy

Numerical Approximation of Solutions to Stochastic Partial Differential Equations and Their Moments

Kristin Kirchner

Division of Applied Mathematics and Statistics
Department of Mathematical Sciences
Chalmers University of Technology
and University of Gothenburg
Göteborg, Sweden 2018

Numerical Approximation of Solutions to Stochastic Partial Differential Equations and Their Moments
Kristin Kirchner
ISBN 978-91-7597-697-6

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 4378
ISSN 0346-718X

Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
SE-412 96 Göteborg
Sweden
Phone: +46 (0)31 772 53 57

Author e-mail: `kristin.kirchner@chalmers.se`

Cover illustration: The finite element mesh of the conterminous U.S. created with the R package INLA, which is used for the application in Paper V.

Typeset with LaTeX
Printed by Chalmers Reproservice
Göteborg, Sweden 2018

Numerical Approximation of Solutions to Stochastic Partial Differential Equations and Their Moments

Kristin Kirchner

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

Abstract

The first part of this thesis focusses on the numerical approximation of the first two moments of solutions to parabolic stochastic partial differential equations (SPDEs) with additive or multiplicative noise. More precisely, in Paper I an earlier result[1], which shows that the second moment of the solution to a parabolic SPDE driven by additive Wiener noise solves a well-posed deterministic space-time variational problem, is extended to the class of SPDEs with multiplicative Lévy noise. In contrast to the additive case, this variational formulation is not posed on Hilbert tensor product spaces as trial–test spaces, but on projective–injective tensor product spaces, i.e., on non-reflexive Banach spaces. Well-posedness of this variational problem is derived for the case when the multiplicative noise term is sufficiently small. This result is improved in Paper II by disposing of the smallness assumption. Furthermore, the deterministic equations in variational form are used to derive numerical methods for approximating the first and the second moment of solutions to stochastic ordinary and partial differential equations without Monte Carlo sampling. Petrov–Galerkin discretizations are proposed and their stability and convergence are analyzed.

In the second part the numerical solution of fractional order elliptic SPDEs with spatial white noise is considered. Such equations are particularly interesting for applications in statistics, as they can be used to approximate Gaussian Matérn fields. Specifically, in Paper III a numerical scheme is proposed, which is based on a finite element discretization in space and a quadrature for an integral representation of the fractional inverse involving only non-fractional inverses. For the resulting approximation, an explicit rate of convergence to the true solution in the strong mean-square sense is derived. Subsequently, in Paper IV weak convergence of this approximation is established. Finally, in Paper V a similar method, which exploits a rational approximation of the fractional power operator instead of the quadrature, is introduced and its performance with respect to accuracy and computing time is compared to the quadrature approach from Paper III and to existing methods for inference in spatial statistics.

Keywords: Stochastic partial differential equations, Tensor product spaces, Fractional operators, White noise, Space-time variational problems, Finite element methods, (Petrov–)Galerkin discretizations, Strong and weak convergence.

[1] A. Lang, S. Larsson, and Ch. Schwab, Covariance structure of parabolic stochastic partial differential equations, Stoch. PDE: Anal. Comp., 1(2013), pp. 351–364.

List of included papers

The following papers are included in this thesis:

I. K. Kirchner, A. Lang, and S. Larsson. *Covariance structure of parabolic stochastic partial differential equations with multiplicative Lévy noise.* J. Diff. Equations, 262(12):5896–5927, 2017.

II. K. Kirchner. *Numerical methods for the deterministic second moment equation of parabolic stochastic PDEs.* Preprint, arXiv:1611.02164.

III. D. Bolin, K. Kirchner, and M. Kovács. *Numerical solution of fractional elliptic stochastic PDEs with spatial white noise.* Preprint, arXiv:1705.06565.

IV. D. Bolin, K. Kirchner, and M. Kovács. *Weak convergence of Galerkin approximations for fractional elliptic stochastic PDEs with spatial white noise.* Preprint, arXiv:1711.05188.

V. D. Bolin and K. Kirchner. *The rational SPDE approach for Gaussian random fields with general smoothness.* Preprint, arXiv:1711.04333.

Publication not included in the thesis:

K. Kirchner, K. Urban, and O. Zeeb. *Maxwell's equations for conductors with impedance boundary conditions: Discontinuous Galerkin and Reduced Basis Methods.* M2AN, Math. Model. Numer. Anal., 50(6):1763–1787, 2016.

Author contributions

Paper I. K.K. derived the deterministic space-time variational problems, established their well-posedness on projective–injective tensor product spaces as trial–test spaces, and had the central role in the coordination and writing of the article.

Paper II. Roman Andreev assisted K.K. with the writing and the implementations for this paper.

Paper III. K.K. derived the bounds in the strong and weak-type error estimates, planned, described, and analyzed the numerical experiments, and had the central role in the coordination and writing of the paper.

Paper IV. K.K. proved weak convergence of the numerical approximation, derived the corresponding convergence rate, and did the major part of the writing.

Paper V. K.K. established the mathematical justification of the proposed approximation, including the strong error estimate, and contributed significantly to the writing of the article.

Acknowledgments

Success is not final, failure is not fatal:
it is the courage to continue that counts.[*]

Stig, my courage is a result of your constant support from the day of my interview until now. With your motivation you preserved the light and colors in my picture of math whenever I was about to *Paint It Black*. Thank you for all the fruitful discussions about mathematics and beyond and, in particular, for your advice.

I am very grateful to my co-supervisors Annika Lang and Mihály Kovács for proposing interesting research projects as well as for the careful reading of the introduction to this thesis and their valuable comments. I would like to express my gratitude to Roman Andreev for supporting me during the writing of Paper II and for assisting me with the implementations. For introducing me to the applications of stochastic partial differential equations in spatial statistics, for the good collaboration, and for assisting me with the cover image and the popular scientific description of this thesis, I thank my co-author David Bolin.

I would like to thank the directors of studies Marija Cvijovic, Peter Hegarty, and Johan Tykesson for giving me the opportunity of developing and lecturing a postgraduate course myself. It has been a valuable and nice experience.

I am grateful for the pleasant working atmosphere at the Department of Mathematical Sciences, created by my fellow Ph.D. students and colleagues. In particular, would like to thank Henrike Häbel for the nice moments that we have had together on and off campus.

Finally, I thank my family: my father for teaching me the history of rock music from AC/DC to Frank Zappa, and my mother for always being there to listen.

Kristin Kirchner
Gothenburg, April 16, 2018

[*]Often attributed to Winston Churchill, but this attribution is unproven.

To be a rock
and not to roll.

Led Zeppelin
Stairway to Heaven, 1971

Contents

Part A

Introduction

Introduction

Many models in, e.g., finance, biology, physics, and social sciences are based on ordinary or partial differential equations. In order to improve their applicability to the reality, one has to take uncertainties, such as, measurement errors or unknown fine scale structures, into account. These uncertainties influence certain parameters, the geometry of the physical domain, boundary or initial conditions, or the source terms of the mathematical model. In this work we focus on the latter scenario and consider ordinary and partial differential equations driven by an additive or multiplicative noise term. Under appropriate assumptions, existence and uniqueness of a solution to such an equation is ensured. This solution is then a square-integrable stochastic process with values in a certain state space. In particular, its first and second moment are finite.

More precisely, the present thesis consists of two parts:

(1) The numerical approximation of the first two moments of the solution process to a parabolic stochastic partial differential equation. These two moments determine the covariance and the correlation structure of the solution process. Furthermore, if the solution is Gaussian, its distribution is even completely characterized by the first two moments.
(2) The computationally efficient sampling from approximations of Gaussian Matérn fields, which is of great relevance for spatial statistics.

Existing approaches to (1) typically involve Monte Carlo sampling. However, Monte Carlo methods are, in general, computationally expensive due to the convergence order $1/2$ of the Monte Carlo estimation and the high cost for computing sample paths of solutions to stochastic partial differential equations.

An alternative approach has been suggested in [24], where the first and second moment of the solution process to a parabolic stochastic partial differential equation driven by additive Wiener noise have been described as solutions to deterministic evolution equations which can be formulated as well-posed linear space-time variational problems. Thus, instead of estimating moments from computationally expensive sample paths, one can apply numerical methods to the deterministic variational problems satisfied by the first and second moment. The main promise of this approach is in potential savings in computing time and memory through space-time compressive schemes, e.g., using adaptive wavelet methods or low-rank tensor approximations.

The first aim of this thesis is to extend the result of [24] to parabolic stochastic partial differential equation driven by multiplicative Lévy noise in Paper I.

Afterwards, in Paper II, we focus on Petrov–Galerkin discretizations for these space-time variational problems.

Stochastic partial differential equations of the form discussed in Part (2) of the thesis can, for instance, be used for defining approximations of Gaussian Matérn fields. Due to their practicality, these random fields have become very popular for modeling in spatial statistics. Namely, the variance, the practical correlation range, and the smoothness of a Gaussian Matérn field can directly be controlled via three parameters. However, the traditional covariance-based representation of the fields entails high computational costs. For example, sampling from a Gaussian Matérn field at N locations requires a matrix factorization of an $N \times N$ covariance matrix and, thus, in general, $\mathcal{O}(N^3)$ arithmetic operations. A recent approach addressing this issue is based on the idea to approximate a Gaussian Matérn field by the solution of a fractional stochastic partial differential equation with spatial white noise [25]. In Part (2) of the thesis we discuss the numerical treatment of these equations. More precisely, in Paper III we propose an explicit approximation and prove its convergence to the true solution in the strong mean-square sense. Subsequently, in Papers IV and V we discuss weak convergence of the approximation and its usage for statistical inference.

Throughout the following sections, let H and U be separable Hilbert spaces over $\mathbb{R}$ with respect to the inner products $(\,\cdot\,,\,\cdot\,)_H$ and $(\,\cdot\,,\,\cdot\,)_U$, respectively. The induced norm on H is $\|\cdot\|_H$ and similarly for U. If $(E, \|\cdot\|_E)$ is a normed vector space, then $S(E) := \{x \in E : \|x\|_E = 1\}$ denotes its unit sphere, $I\colon E \to E$ the identity on E, $\mathcal{B}(E)$ the Borel σ-algebra generated by all subsets of E which are open with respect to the norm topology, and E' the dual of E, i.e, all linear continuous mappings from E to $\mathbb{R}$.

1. Operator theory and tensor product spaces

The outline of this section is as follows. We first introduce several operator classes in §1.1, which are relevant throughout the thesis. In §1.2 we present different notions of tensor product spaces and some of their properties. These spaces appear in the deterministic variational formulations of Papers I and II. Finally, in §1.3 we establish a relation between the Schatten class operators from §1.1 on the one hand and the tensor product spaces from §1.2 on the other.

1.1. Special classes of bounded linear operators. In this subsection we present different classes of linear operators which are of relevance for our analysis. For a detailed overview of operator classes we refer to [13, 14].

1.1.1. *Bounded operators.* A linear operator $T\colon U \to H$ is called bounded or continuous if it has a finite operator norm:

$$\|T\|_{\mathcal{L}(U;H)} := \sup_{x \in S(U)} \|Tx\|_H < \infty.$$

With the above norm, the space of all continuous linear operators from U to H is a Banach space denoted by $\mathcal{L}(U; H)$. We write $\mathcal{L}(U)$ whenever $U = H$.

1.1.2. *Compact operators.* A linear operator $T\colon U \to H$ is compact if the image of any bounded set in U (or equivalently the closed unit ball in U) under T is relatively compact in H, meaning that its closure is compact. We let $\mathcal{K}(U;H)$ denote the set of all compact operators mapping from U to H and use the abbreviation $\mathcal{K}(U)$ when $U = H$.

Equivalently, cf. [43, §X.2], one can define the subspace $\mathcal{K}(U;H) \subset \mathcal{L}(U;H)$ as the closure of all finite-rank operators mapping from U to H. This means that the operator $T\colon U \to H$ is compact if and only if there exists a sequence of linear operators $T_n \in \mathcal{L}(U;H)$ with finite-dimensional range $\operatorname{Rg}(T_n)$ converging to T in the norm topology on $\mathcal{L}(U;H)$:

$$\dim \operatorname{Rg}(T_n) < \infty \quad \forall n \in \mathbb{N}, \qquad \lim_{n\to\infty} \|T - T_n\|_{\mathcal{L}(U;H)} = 0.$$

In our analysis we use the latter characterization of compact operators.

We introduce the adjoint $T^*\colon H \to U$ of a linear operator $T\colon U \to H$ by

$$(Tx, \phi)_H = (x, T^*\phi)_U \quad \forall x \in U, \quad \forall \phi \in H.$$

If $U = H$ and $T^* = T$ then T is called self-adjoint.

Self-adjoint compact operators have real-valued spectra and they generate orthonormal bases consisting of eigenvectors, see [14, Cor. X.3.5]. Since we refer to this property several times, we summarize it in the following theorem.

THEOREM 1.1 (Spectral theorem for self-adjoint compact operators). *Let $T \in \mathcal{K}(U)$ be self-adjoint. Then there exists an orthonormal basis $\{e_j\}_{j\in\mathbb{N}}$ of U and a real-valued sequence $\{\gamma_j\}_{j\in\mathbb{N}}$, which has 0 as its only accumulation point, such that $Te_j = \gamma_j e_j$ for all $j \in \mathbb{N}$.*

1.1.3. *Schatten class operators.* A continuous linear operator $T \in \mathcal{L}(U;H)$ is called a Schatten class operator of p-th order or a p-Schatten class operator for $1 \le p < \infty$ if T has a finite p-Schatten norm:

$$\|T\|_{\mathcal{L}_p(U;H)} := \left(\sum_{j\in\mathbb{N}} s_j(T)^p\right)^{1/p} < \infty,$$

where

$$s_1(T) \ge s_2(T) \ge \ldots \ge s_j(T) \ge \ldots \ge 0$$

are the singular values of T, i.e., the eigenvalues of the operator $|T| := (T^*T)^{1/2}$ (see also §2.3 for the definition of the square root for operators). The space of all Schatten class operators of p-th order mapping from U to H, denoted by $\mathcal{L}_p(U;H)$, is a Banach space with respect to $\|\cdot\|_{\mathcal{L}_p(U;H)}$. Again, we use an abbreviation when $U = H$, and write $\mathcal{L}_p(U)$ in this case. The Schatten norm is monotone in p, i.e.,

$$\|T\|_{\mathcal{L}_1(U;H)} \ge \|T\|_{\mathcal{L}_p(U;H)} \ge \|T\|_{\mathcal{L}_{p'}(U;H)} \ge \|T\|_{\mathcal{L}(U;H)}$$

for $1 \le p \le p' < \infty$ and, moreover, every Schatten class operator is compact. Therefore, the introduced operator spaces satisfy the following relation:

$$\mathcal{L}_1(U;H) \subset \mathcal{L}_p(U;H) \subset \mathcal{L}_{p'}(U;H) \subset \mathcal{K}(U;H) \subset \mathcal{L}(U;H).$$

1.1.4. *Trace class and Hilbert–Schmidt operators.* Schatten class operators of first order mapping from U into U are also called trace class operators. Their name originates from the following fact: For $T \in \mathcal{L}_1(U)$ the trace, defined by

$$\operatorname{tr}(T) := \sum_{j\in\mathbb{N}} (Te_j, e_j)_U,$$

is finite and independent of the choice of the orthonormal basis $\{e_j\}_{j\in\mathbb{N}}$ of U. Moreover, it holds $|\operatorname{tr}(T)| \leq \operatorname{tr}(|T|) = \|T\|_{\mathcal{L}_1(U)}$, cf. [10, Prop. C.1]. For self-adjoint, positive semi-definite trace class operators the trace coincides with the 1-Schatten norm, i.e., $\operatorname{tr}(T) = \|T\|_{\mathcal{L}_1(U)}$ for all $T \in \mathcal{L}_1^+(U)$, where

$$\mathcal{L}_1^+(U) := \{T \in \mathcal{L}_1(U) : T^* = T,\ (Tx, x)_U \geq 0 \quad \forall x \in U\}.$$

This is due to the equality $|T| = (T^*T)^{1/2} = T$ for $T \in \mathcal{L}_1^+(U)$.

The 2-Schatten norm of $T\colon U \to H$ satisfies

$$\|T\|^2_{\mathcal{L}_2(U;H)} = \operatorname{tr}(T^*T) = \sum_{j\in\mathbb{N}} \|Te_j\|_H^2$$

for any orthonormal basis $\{e_j\}_{j\in\mathbb{N}}$ of U. In contrast to all other Schatten norms when $p \neq 2$, this norm originates from an inner product, namely from

$$(S, T)_{\mathcal{L}_2(U;H)} := \sum_{j\in\mathbb{N}} (Se_j, Te_j)_H, \quad S, T \in \mathcal{L}_2(U; H),$$

which is referred to as the Hilbert–Schmidt inner product between S and T. Thus, the space $\mathcal{L}_2(U; H)$ is a Hilbert space. The 2-Schatten norm is called Hilbert–Schmidt norm and elements of $\mathcal{L}_2(U; H)$ are Hilbert–Schmidt operators.

1.2. Tensor product spaces. Besides H and U, let also $(\widetilde{H}, (\,\cdot\,,\,\cdot\,)_{\widetilde{H}})$ and $(\widetilde{U}, (\,\cdot\,,\,\cdot\,)_{\widetilde{U}})$ denote separable Hilbert spaces over $\mathbb{R}$. We first introduce the algebraic tensor product space $U \otimes \widetilde{U}$ as the vector space consisting of all finite sums of the form

$$\sum_{n=1}^{N} x_n \otimes \widetilde{x}_n, \quad x_n \in U,\ \widetilde{x}_n \in \widetilde{U},\ n = 1, \dots, N,$$

equipped with the obvious algebraic operations. There are several ways to define a norm on this vector space, and taking the closure with respect to the different norms yields different Banach spaces. For our purposes, the three notions of tensor product spaces below are of importance. We refer to [18, 34] for a general introduction into the theory of tensor product spaces.

1.2.1. *Hilbert tensor product space.* The Hilbert tensor product space, denoted by $U \hat{\otimes}_2 \widetilde{U}$, is the completion of the algebraic tensor product space $U \otimes \widetilde{U}$ with respect to the norm induced by the inner product

$$(\hat{x}, \hat{y})_{U\hat{\otimes}_2\widetilde{U}} := \sum_{n=1}^{N}\sum_{m=1}^{M} (x_n, y_m)_U (\widetilde{x}_n, \widetilde{y}_m)_{\widetilde{U}},$$

which is independent of the choice of the representations $\hat{x} = \sum_{n=1}^{N} x_n \otimes \widetilde{x}_n$ and $\hat{y} = \sum_{m=1}^{M} y_m \otimes \widetilde{y}_m$ of $\hat{x}$, $\hat{y} \in U \otimes \widetilde{U}$. If $U = \widetilde{U}$ we abbreviate the notation for this space by $U_2 := U \,\hat{\otimes}_2\, U$, and let $(\,\cdot\,,\,\cdot\,)_2$, $\|\cdot\|_2$ denote the inner product and the corresponding norm, respectively.

1.2.2. *Projective tensor product space.* The projective tensor product space $U \,\hat{\otimes}_\pi\, \widetilde{U}$ is obtained by taking the closure of the algebraic tensor product space $U \otimes \widetilde{U}$ with respect to the projective norm defined for $\hat{x} \in U \otimes \widetilde{U}$ by

$$\|\hat{x}\|_{U\hat{\otimes}_\pi\widetilde{U}} := \inf\left\{\sum_{n=1}^{N} \|x_n\|_U \|\widetilde{x}_n\|_{\widetilde{U}} : \hat{x} = \sum_{n=1}^{N} x_n \otimes \widetilde{x}_n\right\}.$$

We write $U_\pi := U \,\hat{\otimes}_\pi\, U$ and $\|\cdot\|_\pi := \|\cdot\|_{U\hat{\otimes}_\pi U}$, whenever $U = \widetilde{U}$.

1.2.3. *Injective tensor product space.* The injective norm of an element $\hat{x}$ in the algebraic tensor product space $U \otimes \widetilde{U}$ is defined as

$$\|\hat{x}\|_{U\hat{\otimes}_\varepsilon\widetilde{U}} := \sup\left\{\Big|\sum_{n=1}^{N} f(x_n)\, g(\widetilde{x}_n)\Big| : f \in S(U'),\, g \in S(\widetilde{U}')\right\},$$

where $\sum_{n=1}^{N} x_n \otimes \widetilde{x}_n$ is any representation of $\hat{x} \in U \otimes \widetilde{U}$. Note that the value of the supremum is independent of the choice of the representation of $\hat{x}$, see [34, p. 45]. The completion of $U \otimes \widetilde{U}$ with respect to this norm is called injective tensor product space and denoted by $U \,\hat{\otimes}_\varepsilon\, \widetilde{U}$. If $U = \widetilde{U}$, the abbreviations $U_\epsilon := U \,\hat{\otimes}_\varepsilon\, U$ as well as $\|\cdot\|_\epsilon := \|\cdot\|_{U\hat{\otimes}_\varepsilon U}$ are used.

1.2.4. *Some remarks.* The projective and injective tensor product spaces in §1.2.2 and §1.2.3 can also be defined for Banach spaces E, $\widetilde{E}$. Note that, even if E, $\widetilde{E}$ are reflexive, the tensor product spaces $E \,\hat{\otimes}_\pi\, \widetilde{E}$ and $E \,\hat{\otimes}_\varepsilon\, \widetilde{E}$ are Banach spaces, which are, in general, not reflexive, cf. [34, Thm. 4.21].

An immediate consequence of the above definitions of the different tensor norms is the following chain of continuous embeddings, see [34, Prop. 6.1(a)]:

$$U \,\hat{\otimes}_\pi\, \widetilde{U} \hookrightarrow U \,\hat{\otimes}_2\, \widetilde{U} \hookrightarrow U \,\hat{\otimes}_\varepsilon\, \widetilde{U}.$$

Here, the embedding constants are all equal to 1.

Another important fact when dealing with linear operators on tensor product spaces is the following: For $T \in \mathcal{L}(U; H)$ and $S \in \mathcal{L}(\widetilde{U}; \widetilde{H})$ setting

$$(T \otimes S)(x \otimes \widetilde{x}) = (Tx) \otimes (S\widetilde{x}), \quad x \in U,\ \widetilde{x} \in \widetilde{U},$$

and extending this definition by linearity to elements in $U \otimes \widetilde{U}$ yields a well-defined linear operator $T \otimes S$ mapping between the algebraic tensor product spaces $U \otimes \widetilde{U}$ and $H \otimes \widetilde{H}$. This operator admits a unique extension to a continuous linear operator $T\hat{\otimes}_\iota S\colon U\hat{\otimes}_\iota\widetilde{U} \to H\hat{\otimes}_\iota\widetilde{H}$ and it holds

$$\|T\hat{\otimes}_\iota S\|_{\mathcal{L}(U\hat{\otimes}_\iota\widetilde{U};H\hat{\otimes}_\iota\widetilde{H})} = \|T\|_{\mathcal{L}(U;H)}\|S\|_{\mathcal{L}(\widetilde{U};\widetilde{H})}$$

for all types of tensor spaces $\iota \in \{2, \pi, \epsilon\}$ considered above, see [34, Propositions 2.3 & 3.2] and Lemma 3.1(ii) in Paper I.

1.3. Relating tensor product spaces and Schatten class operators. In the following, we establish a connection between the tensor product spaces U_π and U_2 on the one hand, and Schatten class operators of order 1 and 2 on the other hand. In addition, we show that compact operators in $\mathcal{K}(U)$ are related to the elements in the injective tensor product space U_ϵ.

For this purpose, we first define a kernel k on U as an element in the algebraic tensor product space $U \otimes U$, which is, in addition, symmetric, i.e.,

$$(k, x \otimes y)_2 = (k, y \otimes x)_2 \quad \forall x, y \in U.$$

Owing to the Riesz representation theorem, we can define the action of the linear operator $T_k \colon U \to U$ associated with the kernel k on $x \in U$ as the unique element $T_k x \in U$ satisfying

$$(T_k x, y)_U = (k, x \otimes y)_2 \quad \forall y \in U.$$

The next proposition illustrates the relation between the introduced tensor product spaces and compact / 1-Schatten class / 2-Schatten class operators.

PROPOSITION 1.2. *Let k be a kernel on U. The linear operator $T_k \colon U \to U$ associated with the kernel k is self-adjoint and it holds*

$$\|k\|_\epsilon = \|T_k\|_{\mathcal{L}(U)}, \quad \|k\|_\pi = \|T_k\|_{\mathcal{L}_1(U)}, \quad \textit{and} \quad \|k\|_2 = \|T_k\|_{\mathcal{L}_2(U)}.$$

In other words, the mapping $\mathcal{J} \colon k \mapsto T_k$ extends to an isometric isomorphism between the spaces

$$U_\epsilon^{\mathrm{sym}} \overset{\mathcal{J}}{\cong} \mathcal{K}^{\mathrm{sym}}(U), \quad U_\pi^{\mathrm{sym}} \overset{\mathcal{J}}{\cong} \mathcal{L}_1^{\mathrm{sym}}(U), \quad \textit{and} \quad U_2^{\mathrm{sym}} \overset{\mathcal{J}}{\cong} \mathcal{L}_2^{\mathrm{sym}}(U),$$

where the superscript $^{\mathrm{sym}}$ indicates, for a tensor product space, to take the closure of only the symmetric elements in the algebraic tensor product space $U \otimes U$ with respect to the tensor norm, and, for an operator space, the closed subspace of self-adjoint operators.

PROOF. Self-adjointness of T_k follows from the symmetry of the kernel k. In order to prove the norm equalities, let $k = \sum_{n=1}^N k_n^1 \otimes k_n^2$ be any representation of $k \in U \otimes U$.

Then we obtain for the operator norm of the induced operator T_k:

$$\begin{aligned}\|T_k\|_{\mathcal{L}(U)} &= \sup_{x \in S(U)} \|T_k x\|_U = \sup_{x,y \in S(U)} (T_k x, y)_U \\ &= \sup_{x,y \in S(U)} \sum_{n=1}^N (k_n^1, x)_U (k_n^2, y)_U = \sup_{f,g \in S(U')} \sum_{n=1}^N f(k_n^1) g(k_n^2) = \|k\|_\epsilon.\end{aligned}$$

In this calculation the Riesz representation theorem justifies taking the supremum over $f, g \in S(U')$ instead of over $x, y \in S(U)$. Therefore, the self-adjoint linear operator $T_k \colon U \to U$ is continuous if and only if its kernel k is an element of the tensor product space $U_\epsilon^{\mathrm{sym}}$. The identity

$$(T_k x, y)_U = \sum_{n=1}^N (k_n^1, x)_U (k_n^2, y)_U = \Big(\sum_{n=1}^N (k_n^1, x)_U \, k_n^2, y \Big)_U \quad \forall y \in U$$

shows that $T_k x = \sum_{n=1}^{N} (k_n^1, x)_U \, k_n^2$ for all $x \in U$, i.e., T_k is a finite-rank operator and thus compact if $k \in U \otimes U$. In the more general case when $k \in U_\epsilon^{\mathrm{sym}}$, we can find a sequence of kernels in $U \otimes U$ converging to k with respect to the injective norm $\|\cdot\|_\epsilon$. Due to the isometry property derived above, also T_k can be approximated by self-adjoint finite-rank operators in $\mathcal{L}(U)$ and, hence, $T_k \in \mathcal{K}^{\mathrm{sym}}(U)$, see §1.1.2.

The application of Theorem 1.1 to $T_k \in \mathcal{K}^{\mathrm{sym}}(U)$ yields the existence of an orthonormal basis $\{e_j\}_{j\in\mathbb{N}}$ of U consisting of eigenvectors of T_k with corresponding eigenvalues $\{\gamma_j\}_{j\in\mathbb{N}} \subset \mathbb{R}$. The observation $(k, e_i \otimes e_j)_2 = (T_k e_i, e_j)_U = \gamma_j \delta_{ij}$, where δ_{ij} denotes the Kronecker delta, shows that k can be expanded in U_2 as $k = \sum_{j\in\mathbb{N}} \gamma_j (e_j \otimes e_j)$ and we obtain the estimate

$$\|k\|_\pi \le \sum_{j\in\mathbb{N}} |\gamma_j| = \mathrm{tr}(|T_k|) = \|T_k\|_{\mathcal{L}_1(U)}.$$

The reverse inequality follows from the Cauchy–Schwarz inequality for sums and Parseval's identity:

$$\begin{aligned}\|T_k\|_{\mathcal{L}_1(U)} &= \sum_{j\in\mathbb{N}} |\gamma_j| = \sum_{j\in\mathbb{N}} |(k, e_j \otimes e_j)_2| = \sum_{j\in\mathbb{N}} \Big| \sum_{n=1}^{N} (k_n^1, e_j)_U (k_n^2, e_j)_U \Big| \\ &\le \sum_{n=1}^{N} \Big(\sum_{j\in\mathbb{N}} (k_n^1, e_j)_U^2 \Big)^{1/2} \Big(\sum_{j\in\mathbb{N}} (k_n^2, e_j)_U^2 \Big)^{1/2} = \sum_{n=1}^{N} \|k_n^1\|_U \|k_n^2\|_U.\end{aligned}$$

Since the representation of k is arbitrary, we may take the infimum over all representations of k in $U \otimes U$ and conclude $\|T_k\|_{\mathcal{L}_1(U)} \le \|k\|_\pi$. Thus, $\mathcal{J}$ extends to an isometric isomorphism between the spaces U_π^{sym} and $\mathcal{L}_1^{\mathrm{sym}}(U)$.

For the Hilbert–Schmidt norm of T_k we find

$$\begin{aligned}\|T_k\|_{\mathcal{L}_2(U)}^2 &= \sum_{i\in\mathbb{N}} \|T_k e_i\|_U^2 = \sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} (T_k e_i, e_j)_U^2 \\ &= \sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} \sum_{m=1}^{N} \sum_{n=1}^{N} (k_m^1, e_i)_U (k_n^1, e_i)_U (k_m^2, e_j)_U (k_n^2, e_j)_U \\ &= \sum_{m=1}^{N} \sum_{n=1}^{N} (k_m^1, k_n^1)_U (k_m^2, k_n^2)_U = \|k\|_2^2.\end{aligned}$$

The space of self-adjoint Hilbert–Schmidt operators $\mathcal{L}_2^{\mathrm{sym}}(U)$ is therefore isometrically isomorphic to the subspace U_2^{sym} of the Hilbert tensor product space. □

In the literature, kernels in U_π and U_2, which are not necessarily symmetric, are often called Fredholm kernels [17] and Hilbert–Schmidt kernels [43, §VII.3, Example 1], respectively. We note that Proposition 1.2 can be seen as a generalization of Mercer's theorem [28], which considers continuous real-valued kernels $k \in C^{\mathrm{sym}}([a,b] \times [a,b]) \subset U_\epsilon^{\mathrm{sym}}$ for $-\infty < a < b < \infty$, where $U = L_2(a,b)$ and

$$C^{\mathrm{sym}}([a,b] \times [a,b]) := \{f \in C([a,b] \times [a,b]) : f(s,t) = f(t,s),\ a \le s,t \le b\}.$$

As a by-product of Proposition 1.2, we obtain an explicit way to calculate the projective norm of kernels in U_π^{sym} which are positive semi-definite, i.e., in

$$U_\pi^+ := \{k \in U_\pi^{\mathrm{sym}} : (k, x \otimes x)_2 \geq 0 \quad \forall x \in U\}.$$

To this end, we first introduce the real-valued linear operator δ on the algebraic tensor product space $U \otimes U$ as follows: If $k = \sum_{n=1}^N k_n^1 \otimes k_n^2$ is any representation of $k \in U \otimes U$, we set $\delta(k) := \sum_{n=1}^N (k_n^1, k_n^2)_U$. The equalities

$$\delta(k) = \sum_{n=1}^N \sum_{j\in\mathbb{N}} (k_n^1, e_j)_U (e_j, k_n^2)_U = \sum_{j\in\mathbb{N}} (k, e_j \otimes e_j)_U = \sum_{j\in\mathbb{N}} (k, \widetilde{e}_j \otimes \widetilde{e}_j)_U,$$

where $\{e_j\}_{j\in\mathbb{N}}$, $\{\widetilde{e}_j\}_{j\in\mathbb{N}}$ are any two orthonormal bases of U, shows that δ is well-defined, independently of the representation of k. Furthermore, the operator δ is bounded with respect to the projective norm, since the estimate

$$|\delta(k)| \leq \sum_{n=1}^N \|k_n^1\|_U \|k_n^2\|_U$$

holds for any representation of $k \in U \otimes U$. For this reason and owing to the density of $U \otimes U$ in U_π, there exists a unique linear continuous extension of δ to a functional $\delta \in (U_\pi)'$ on the projective tensor product space U_π.

If the kernel k is positive semi-definite, $k \in U_\pi^+$, then the associated linear operator $T_k \colon U \to U$ is an element of $\mathcal{L}_1^+(U)$ and we may choose the orthonormal basis $\{e_j\}_{j\in\mathbb{N}} \subset U$ as the eigenbasis of T_k with corresponding eigenvalues $\gamma_j \geq 0$. We then obtain the identity

$$\|T_k\|_{\mathcal{L}_1(U)} = \sum_{j\in\mathbb{N}} \gamma_j = \sum_{j\in\mathbb{N}} (T_k e_j, e_j)_U = \sum_{j\in\mathbb{N}} (k, e_j \otimes e_j)_2 = \delta(k),$$

and the projective norm of a positive semi-definite kernel $k \in U_\pi^+$ is given by

$$\|k\|_\pi = \|T_k\|_{\mathcal{L}_1(U)} = \delta(k).$$

For $U := L_2(a,b)$, the functional δ is equal to the L_1-norm on the diagonal:

$$\|k\|_\pi = \delta(k) = \int_a^b k(t,t)\,\mathrm{d}t \quad \forall k \in L_2(a,b)_\pi^+.$$

2. Analytic tools for evolution equations

In this section we first recall the concepts of differentiating vector-valued functions in §2.1, as well as the notions of semigroups on Hilbert spaces and their generators in §2.2. We then define fractional powers of closed operators in §2.3. This definition is particularly important for Papers III–V in the second part of the thesis. The specific class of closed, possibly unbounded (differential) operators considered throughout all Papers I–V is discussed in §2.4. Finally, in §2.5, the spaces of Bochner-integrable vector-valued functions are introduced, which appear in the variational problems of Papers I and II.

For general introductions to differentiation and integration of vector-valued functions as well as semigroup theory, we refer to [15, 19, 26, 31, 43].

2.1. Differentiation in Hilbert spaces. In this subsection we introduce the two main concepts of differentiation for vector-valued functions: Gâteaux derivatives and Fréchet derivatives. To this end, let $U_0 \subset U$ be an open subset of the Hilbert space U.

2.1.1. *Gâteaux derivative.* The Gâteaux derivative or Gâteaux differential of a function $f\colon U_0 \to H$ at $x_0 \in U_0$ in the direction $y \in U$ is defined as

$$Df(x_0; y) := \lim_{\tau \to 0} \frac{f(x_0 + \tau y) - f(x_0)}{\tau}.$$

If this limit exists along all directions $y \in U$, we say that f is Gâteaux differentiable in $x_0 \in U_0$.

2.1.2. *Fréchet derivative.* A function $f\colon U_0 \to H$ is called Fréchet differentiable at $x_0 \in U_0$ if there exists a bounded linear operator $T\colon U \to H$ such that

$$\lim_{\|y\|_U \to 0} \frac{\|f(x_0 + y) - f(x_0) - Ty\|_H}{\|y\|_U} = 0,$$

or equivalently, in Laudau notation,

$$f(x_0 + y) = f(x_0) + Ty + o(\|y\|_U) \quad \text{as} \quad \|y\|_U \to 0.$$

We call $Df(x_0) := T$ the Fréchet derivative or the Fréchet differential of f at x_0. If f is Fréchet differentiable at every $x_0 \in U_0$, its derivative

$$Df\colon U_0 \to \mathcal{L}(U; H), \qquad x_0 \mapsto Df(x_0),$$

is an operator-valued function, taking values in the space of bounded linear operators mapping from U to H. We note that, sometimes, Df is also called the strong derivative of f (in particular, when $U = \mathbb{R}$).

Clearly, the definitions of Gâteaux and Fréchet differentiability coincide if $U = \mathbb{R}$, but already for $U = \mathbb{R}^2$ and $H = \mathbb{R}$ one can construct functions which are Gâteaux differentiable, e.g., in the origin, but not Fréchet differentiable. An example is the function $f\colon \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x_1, x_2) := x_1^3(x_1^2 + x_2^2)^{-1}$ if $(x_1, x_2) \neq (0, 0)$ and $f(0, 0) := 0$.

In fact, another way of defining Fréchet differentiability of f at $x_0 \in U_0$ is to require that f is Gâteaux differentiable at x_0 with $Df(x_0; \cdot\,) \in \mathcal{L}(U; H)$ and that, in addition, the difference quotients converge uniformly for all directions $y \in U$. In this case, $Df(x_0) = Df(x_0; \cdot\,)$. From this definition, it is evident that every function, which is Fréchet differentiable at x_0, is also Gâteaux differentiable there, and that the converse is not true.

For our purposes the notion of Fréchet differentiability of real-valued functions $f\colon U \to \mathbb{R}$ will be particularly important, i.e., $U_0 = U$ and $H = \mathbb{R}$. In this case, the Fréchet derivative $Df\colon U \to U'$ takes by definition values in the dual space U'. By the Riesz representation theorem, it can be identified with a mapping $Df\colon U \to U$ and, thus, the second Fréchet derivative of f at $x_0 \in U$ with a bounded linear operator on U, i.e.,

$$D^2 f(x_0) := D(Df)(x_0) \in \mathcal{L}(U).$$

Consequently, the spaces of once resp. twice continuously Fréchet differentiable real-valued functions on U can be indentified as follows,

$$C^1(U;\mathbb{R}) := \{f \in C(U;\mathbb{R}) : Df \in C(U;U)\},$$
$$C^2(U;\mathbb{R}) := \{f \in C^1(U;\mathbb{R}) : D^2 f \in C(U;\mathcal{L}(U))\},$$

where $C(U;H)$ denotes the vector space of all continuous mappings $f\colon U \to H$.

2.2. Semigroups and generators. A family $(S(t), t \geq 0)$ of bounded linear operators from H to H is called a C_0-semigroup (or strongly continuous one-parameter semigroup) on H if

(i) $S(0) = I$ and $S(t+t') = S(t)S(t')$ for all $t, t' \geq 0$,
(ii) $\lim_{t\to 0^+} S(t)\phi = \phi$ for all $\phi \in H$, where $\lim_{t\to 0^+}$ denotes the one-sided limit from above 0 in H.

The semigroup $(S(t), t \geq 0)$ is uniformly bounded if

(iii) there exists $M \geq 1$ such that $\|S(t)\|_{\mathcal{L}(H)} \leq M$ for all $t \geq 0$.

In the case $M = 1$, it is called a semigroup of contractions. Finally, if one can replace $t, t' \geq 0$ in (i)–(ii) by $z, z' \in \Sigma$ for a sector

$$\Sigma := \{z \in \mathbb{C} : \varphi_1 < \arg(z) < \varphi_2,\ \varphi_1 < 0 < \varphi_2\}$$

containing the nonnegative real axis, and if, in addition,

(iv) the map $z \mapsto S(z)$ is analytic in Σ,

we say that the semigroup $(S(t), t \geq 0)$ is analytic.

The linear operator $\widehat{A}$, defined on the domain

$$\mathcal{D}(\widehat{A}) := \left\{\phi \in H : \lim_{t\to 0^+} \frac{S(t)\phi - \phi}{t} \text{ exists}\right\}$$

by

$$\widehat{A}\phi := \lim_{t\to 0^+} \frac{S(t)\phi - \phi}{t} \quad \forall \phi \in \mathcal{D}(\widehat{A}),$$

is the infinitesimal generator of the semigroup $(S(t), t \geq 0)$.

The following famous theorem named after the mathematicians Einar Hille and Kôsaku Yosida characterizes generators of C_0-semigroups of contractions, see [31, Thm. 3.1]. It ensures that, for instance, an unbounded operator A of the form presented in §2.4 and considered throughout all Papers I–V is related to a generator $\widehat{A}$ of a C_0-semigroup via $\widehat{A} = -A$.

Theorem 2.1 (Hille–Yosida). *A linear (unbounded) operator $\widehat{A}$ is the infinitesimal generator of a C_0-semigroup of contractions $(S(t), t \geq 0)$ if and only if*

(i) $\widehat{A}$ is closed and densely defined,
(ii) the resolvent set $\varrho(\widehat{A}) := \{\lambda \in \mathbb{C} : \lambda I - \widehat{A}$ is invertible$\}$ contains all positive real numbers and

$$\|(\lambda I - \widehat{A})^{-1}\|_{\mathcal{L}(H)} \leq \lambda^{-1} \quad \forall \lambda > 0.$$

2.3. Fractional powers of closed operators. Let A be a linear operator for which $\widehat{A} := -A$ is the infinitesimal generator of an analytic C_0-semigroup $(S(t), t \geq 0)$ on the Hilbert space H. For $\beta > 0$, we use the integral representation of $A^{-\beta}$ by A.V. Balakrishnan [2] to define the negative fractional power operator in terms of the semigroup by

$$A^{-\beta} := \frac{1}{\Gamma(\beta)} \int_0^\infty t^{\beta-1} S(t) \, \mathrm{d}t,$$

where Γ is the gamma function, and we set $A^{-0} := I$. Owing to the identity

$$\frac{1}{\Gamma(\beta)} \frac{\pi}{\sin(\pi\beta)} = \int_0^\infty u^{-\beta} e^{-u} \, \mathrm{d}u, \quad 0 < \beta < 1,$$

Fubini's theorem, and the change of variables $u = st$, we can reformulate this definition for $0 < \beta < 1$ as

$$\begin{aligned} A^{-\beta} &= \frac{\sin(\pi\beta)}{\pi} \int_0^\infty \int_0^\infty u^{-\beta} e^{-u} s^{\beta-1} S(s) \, \mathrm{d}u \, \mathrm{d}s \\ &= \frac{\sin(\pi\beta)}{\pi} \int_0^\infty S(s) \int_0^\infty t^{-\beta} e^{-st} \, \mathrm{d}t \, \mathrm{d}s = \frac{\sin(\pi\beta)}{\pi} \int_0^\infty t^{-\beta} (tI + A)^{-1} \, \mathrm{d}t. \end{aligned}$$

Here, we have used the integral representation of the resolvent,

$$(tI + A)^{-1} = \int_0^\infty e^{-st} S(s) \, \mathrm{d}s,$$

in the last step. For our purposes, both representations of $A^{-\beta}$ will be relevant: The first one is used in the derivation of the weak error bound in Paper IV. The latter representation, which holds only for $0 < \beta < 1$, is the starting point for the quadrature used to compute the numerical approximation in Papers III–V.

The so-defined negative fractional power operators satisfy the following properties, see [31, §2.6]:

(a) There exists $C > 0$ such that $\|A^{-\beta}\|_{\mathcal{L}(H)} \leq C$ for $0 \leq \beta \leq 1$.
(b) $A^{-(\beta+\gamma)} = A^{-\beta} A^{-\gamma}$ for all $\beta, \gamma \geq 0$.
(c) $\lim_{\beta \to 0^+} A^{-\beta} \phi = \phi$ for all $\phi \in H$.
(d) $A^{-\beta}$ is injective for all $\beta \geq 0$.

The characteristics (a)–(c) show that $(A^{-\beta}, \beta \geq 0)$ is a C_0-semigroup on H. Property (d) allows us to define the fractional power operator A^β for $\beta \geq 0$ on the domain $\mathcal{D}(A^\beta) := \mathrm{Rg}(A^{-\beta})$ by

$$A^\beta := (A^{-\beta})^{-1} \quad \text{if } \beta > 0, \quad \text{and} \quad A^0 := I.$$

Below, we collect some properties of fractional power operators which are used in the thesis without further mentioning, see [31, Thm. 6.8]:

(a) A^β is a closed operator.
(b) $\mathcal{D}(A^\beta) \subset \mathcal{D}(A^\gamma)$ if $\beta \geq \gamma \geq 0$.
(c) For every $\beta \geq 0$, A^β is densely defined, i.e., $\overline{\mathcal{D}(A^\beta)} = H$.
(d) $A^{\beta+\gamma} \phi = A^\beta A^\gamma \phi$ for all $\phi \in \mathcal{D}(A^\theta)$, where $\theta := \max\{\beta, \gamma, \beta + \gamma\}$, and for every $\beta, \gamma \in \mathbb{R}$.

2.4. A class of operators generating analytic semigroups. In what follows, we let $A\colon \mathcal{D}(A) \to H$ be a linear operator defined on a dense subspace $\mathcal{D}(A)$ of H. In addition, we assume that A is self-adjoint and positive definite,

$$(A\phi, \psi)_H = (\phi, A\psi)_H, \quad (A\vartheta, \vartheta)_H > 0 \quad \forall \phi, \psi, \vartheta \in \mathcal{D}(A),\ \vartheta \neq 0,$$

and that A has a compact inverse $A^{-1} \in \mathcal{K}(H)$. The application of Theorem 1.1 to A^{-1} shows that there exists an orthonormal basis $\{e_j\}_{j\in\mathbb{N}}$ of H consisting of eigenvectors of A and an nondecreasing sequence of corresponding eigenvalues $\lambda_j > 0$, i.e., $Ae_j = \lambda_j e_j$ for all $j \in \mathbb{N}$. In particular, the operator $\widehat{A} := -A$ is closed, densely defined on $\mathcal{D}(\widehat{A}) = \mathcal{D}(A)$, and its spectrum consists only of negative real numbers. By the Hille–Yosida theorem, Theorem 2.1, and by [15, Cor. 4.7], $-A$ is thus the generator of an analytic C_0-semigroup of contractions $(S(t), t \geq 0)$ and fractional powers of the operator A are defined as in §2.3.

Furthermore, in this case, the fractional power operator $A^\beta\colon \mathcal{D}(A^\beta) \to H$, defined in §2.3, has the spectral representation

$$A^\beta \phi := \sum_{j\in\mathbb{N}} \lambda_j^\beta (\phi, e_j)_H\, e_j, \quad \beta \geq 0,$$

on the domain

$$\mathcal{D}(A^\beta) = \left\{ \phi \in H : \sum_{j\in\mathbb{N}} \lambda_j^{2\beta} (\phi, e_j)_H^2 < \infty \right\}.$$

The subspace $\dot{H}^r := \mathcal{D}(A^{r/2}) \subset H$, $r \geq 0$, is itself a Hilbert space with respect to the inner product

$$(\phi, \psi)_r := (A^{r/2}\phi, A^{r/2}\psi)_H.$$

For $r > 0$, we let the negative-indexed space $\dot{H}^{-r}$ be defined as the dual space of $\dot{H}^r$. By identifying $\dot{H}^{-0} := H'$ with $\dot{H}^0 = H$ (via the Riesz map), we obtain the following scale of continuously and densely embedded Hilbert spaces,

$$\dot{H}^s \hookrightarrow \dot{H}^r \hookrightarrow \dot{H}^0 = H \cong \dot{H}^{-0} \hookrightarrow \dot{H}^{-r} \hookrightarrow \dot{H}^{-s}$$

for $s \geq r \geq 0$. The duality pairing between $\dot{H}^{-r}$ and $\dot{H}^r$ (or vice versa) is denoted by $\langle \cdot, \cdot \rangle_r$. If $r = 1$, we omit the subscript and write $\langle \cdot, \cdot \rangle$. The norm on $\dot{H}^{-r}$ can then be expressed by [40]

$$\|g\|_{-r} = \sup_{\phi \in \dot{H}^r \setminus \{0\}} \frac{\langle g, \phi \rangle_r}{\|\phi\|_r} = \left(\sum_{j\in\mathbb{N}} \lambda_j^{-r} \langle g, e_j \rangle_r^2 \right)^{1/2}.$$

It is an immediate consequence of this representation of the dual norm and the definition of the $\dot{H}^r$ spaces above that there exists a unique continuous extension of $A^\beta\colon \dot{H}^{2\beta} \to H$ to an isometric isomorphism $A^\beta\colon \dot{H}^r \to \dot{H}^{r-2\beta}$ for all $\beta \geq 0$ and any $r \in \mathbb{R}$, cf. Lemma 2.1 of Paper III.

If we let $H^* := \dot{H}^{-0}$, $V := \dot{H}^1$, and $V^* := \dot{H}^{-1}$, the operator $A\colon V \to V^*$ is bounded and we obtain the following Gelfand triple,

$$V \hookrightarrow H \cong H^* \hookrightarrow V^*.$$

Finally, we stress the difference between the spaces V' and V^*: V' denotes the dual of V in its original sense, i.e., all continuous linear mappings from V to $\mathbb{R}$, while V^* is the identification of the dual with respect to the pivot space H. Therefore, we have the relation: $g \in V^*$ if and only if $\langle g, \cdot \rangle \in V'$.

2.5. Bochner spaces. In order to define the trial and test spaces for the variational formulations of the evolution equations in the first part of the thesis, Papers I and II, we introduce a special class of function spaces: Bochner (or Lebesgue–Bochner) spaces. For the definitions of measurability and the Bochner integral, we refer to [43, §V.5].

For finite $T > 0$ and $1 \leq p < \infty$, we consider the time interval $J := (0, T)$ and the Bochner space $L_p(J; H)$ of Bochner-measurable, p-integrable functions mapping from J to the Hilbert space H, which is itself a Banach space with respect to the norm

$$\|u\|_{L_p(J;H)} := \left(\int_J \|u(t)\|_H^p \, \mathrm{d}t \right)^{1/p}.$$

If $p = 2$, it is a Hilbert space with respect to the obvious inner product.

Let $u \in L_1(J; H)$ be an H-valued Bochner-integrable function. Following [12, Ch. XVIII, §1, Def. 3] we define the distributional (weak) derivative $\partial_t u$ of u as the H-valued distribution satisfying

$$((\partial_t u)(v), \phi)_H = -\int_J \frac{\mathrm{d}v}{\mathrm{d}t}(t) \, (u(t), \phi)_H \, \mathrm{d}t \quad \forall (v, \phi) \in C_0^\infty(J; \mathbb{R}) \times H.$$

Note that this notion of differentiability is, in general, weaker than the concepts presented in §2.1 (strong derivative vs. weak derivative).

Recall the Gelfand triple $V \hookrightarrow H \cong H^* \hookrightarrow V^*$ from §2.4. The definition of the distributional derivative of an H-valued function above implies that for a function $u \in L_2(J; V^*)$ its distributional derivative $\partial_t u$ satisfies

$$\langle (\partial_t u)(v), \phi \rangle = -\int_J \frac{\mathrm{d}v}{\mathrm{d}t}(t) \, \langle u(t), \phi \rangle \, \mathrm{d}t \quad \forall (v, \phi) \in C_0^\infty(J; \mathbb{R}) \times V.$$

After having defined the Bochner space $L_2(J; V^*)$ and the distributional derivative we can now introduce the vector-valued Sobolev space

$$H^1(J; V^*) := \{ u \in L_2(J; V^*) : \partial_t u \in L_2(J; V^*) \}$$

and equip it with the norm

$$\|u\|_{H^1(J;V^*)} := \left(\|u\|_{L_2(J;V^*)}^2 + \|\partial_t u\|_{L_2(J;V^*)}^2 \right)^{1/2}.$$

With respect to the corresponding inner product, $H^1(J; V^*)$ is a Hilbert space.

The formulations of the variational problems in Papers I and II involve trial and test spaces originating from the spaces

$$\mathcal{X} := L_2(J; V), \qquad \widehat{\mathcal{Y}} := L_2(J; V) \cap H^1(J; V^*).$$

These spaces are Hilbert spaces: $\mathcal{X}$ with respect to the Bochner inner product $(\,\cdot\,,\,\cdot\,)_{\mathcal{X}} = (\,\cdot\,,\,\cdot\,)_{L_2(J;V)}$ and $\widehat{\mathcal{Y}}$ being equipped with the graph norm

$$\|v\|_{\mathcal{Y}} := \Big(\|v\|^2_{L_2(J;V)} + \|\partial_t v\|^2_{L_2(J;V^*)}\Big)^{1/2},$$

and the obvious corresponding inner product.

It is a well-known result [12, Ch. XVIII, §1, Thm. 1] that $\widehat{\mathcal{Y}} \hookrightarrow C(\bar{J};H)$, where $C(\bar{J};H)$ denotes the space of continuous H-valued functions on the closure $\bar{J} := [0,T]$ of J. Therefore, for $v \in \widehat{\mathcal{Y}}$, the values $v(0)$ and $v(T)$ are well-defined in H, and the following are closed subspaces of $\widehat{\mathcal{Y}}$:

$$\mathcal{Y}_0 := \big\{v \in \widehat{\mathcal{Y}} : v(0) = 0 \text{ in } H\big\}, \qquad \mathcal{Y} := \big\{v \in \widehat{\mathcal{Y}} : v(T) = 0 \text{ in } H\big\},$$

which are equipped with the same norm $\|\cdot\|_{\mathcal{Y}}$ as $\widehat{\mathcal{Y}}$. We note that

$$|||v|||_{\mathcal{Y}} := \Big(\|v\|^2_{L_2(J;V)} + \|\partial_t v\|^2_{L_2(J;V^*)} + \|v(0)\|^2_H\Big)^{1/2}$$

defines an equivalent norm on $\mathcal{Y}$, which has the following advantages:

- The embedding constant in $\mathcal{Y} \hookrightarrow C(\bar{J};H)$ is smaller: Fix $t \in J$. The integration of $\frac{\mathrm{d}}{\mathrm{d}t}\|v(t)\|^2_H = 2\,\langle \partial_t v(t), v(t)\rangle$ for $v \in \mathcal{Y}$ over the interval $[t,T]$ yields (recall that $v(T) = 0$ for $v \in \mathcal{Y}$):
$$\begin{aligned}\|v(t)\|^2_H &\le 2\,\|\partial_t v\|_{L_2((t,T);V^*)}\|v\|_{L_2((t,T);V)}\\ &\le \|v\|^2_{L_2((t,T);V)} + \|\partial_t v\|^2_{L_2((t,T);V^*)}.\end{aligned}$$
This already shows that the embedding constant with respect to $\|\cdot\|_{\mathcal{Y}}$ is bounded by 1. By integrating from 0 to t instead we obtain:
$$\|v(t)\|^2_H \le \|v(0)\|^2_H + \|v\|^2_{L_2((0,t);V)} + \|\partial_t v\|^2_{L_2((0,t);V^*)}.$$
Adding the two inequalities shows that $\|v(t)\|_H \le \frac{1}{\sqrt{2}}|||v|||_{\mathcal{Y}}$. For sharpness of these bounds we refer to the example in §2.2 of Paper II.
- If we define the evolution operator $b\colon \mathcal{X} \to \mathcal{Y}'$ by
$$(bu)(v) := \int_J \langle u(t), (-\partial_t + A)v(t)\rangle\,\mathrm{d}t,$$
then $b \in \mathcal{L}(\mathcal{X};\mathcal{Y}')$ is an isometry, i.e., $|||bu|||_{\mathcal{Y}'} = \|u\|_{\mathcal{X}}$, where we set $|||f|||_{\mathcal{Y}'} := \sup_{v\in\mathcal{Y}\setminus\{0\}} \frac{|f(v)|}{|||v|||_{\mathcal{Y}}}$, see §3.3.

Since, in particular, the latter property is useful for the error analysis of numerical methods, the results of Paper II are formulated with respect to the norm $|||\cdot|||_{\mathcal{Y}}$, while in, e.g., [24, 36, 37] and also in Paper I the norm $\|\cdot\|_{\mathcal{Y}}$ is used.

3. Deterministic initial value problems

Subject of this section is the abstract inhomogeneous initial value problem

$$\text{(IVP)} \qquad u'(t) + Au(t) = f(t), \quad t \in J = (0,T), \qquad u(0) = u_0,$$

for a right-hand side $f \in L_1(J;H)$ and an initial data $u_0 \in H$. Here, $u'\colon J \to H$ denotes the strong derivative of the H-valued function u, see §2.1. We assume

that $-A\colon \mathcal{D}(A) \to H$ is the infinitesimal generator of an analytic C_0-semigroup $(S(t), t \geq 0)$, cf. §2.2. For instance, A can be of the form discussed in §2.4.

In the homogeneous case when $f = 0$, Problem (IVP) is often referred to as the Cauchy problem (relative to the operator $-A$) in the literature.

In the following, we first state the classical definition of a solution to (IVP) in §3.1. We then present three different concepts widening the notion of solutions to (IVP): strong and mild solutions in §3.2, and weak solutions obtained via variational formulations of (IVP) in §3.3.

3.1. Classical solutions. A *classical solution* u to (IVP) on $[0,T)$ is an H-valued function which is continuous on $[0,T)$, continuously differentiable on J, takes values in $\mathcal{D}(A)$ on J, i.e.,

$$u \in C([0,T);H) \cap C((0,T);\mathcal{D}(A)), \quad u' \in C((0,T);H),$$

and satisfies (IVP).

Since $-A$ is assumed to be the infinitesimal generator of an analytic C_0-semigroup, one knows that the homogeneous initial value problem has a unique classical solution for every initial data $u_0 \in H$ [31, §4, Cor. 3.3]. It is given by $u(t) = S(t)u_0$ for all $t \in [0,T)$. From this result, it is evident that the inhomogeneous problem in (IVP) has at most one classical solution.

However, for a general right-hand side $f \in L_1(J;H)$, the definition of a classical solution is often too restrictive in order to ensure existence. Consider, e.g., the simple example when $A = 0$. Then the initial value problem (IVP) does not have a classical solution unless f is continuous. But also continuity of f on $\bar{J} = [0,T]$ is not sufficient to guarantee existence of a classical solution to (IVP) when $-A$ generates a C_0-semigroup, see [31, §4.2] for a counterexample. For this reason, the generalized solution concepts presented in the following subsections have been introduced.

3.2. Strong and mild solutions. Note that for $f \in L_1(J;H)$ and $A = 0$, the initial value problem (IVP) always has a solution, which is differentiable almost everywhere and satisfies $u'(t) = f(t)$ for almost every $t \in J$. Namely, the function given by $u(t) = u_0 + \int_0^t f(s)\,\mathrm{d}s$ has these properties. This motivates the following definition of a strong solution to (IVP), see [31, §4.2].

A function $u\colon \bar{J} \to H$, which is differentiable almost everywhere on $\bar{J}$ with $u' \in L_1(J;H)$, is called a *strong solution* of (IVP) if $u(0) = u_0$ and

$$u'(t) + Au(t) = f(t) \quad \text{for a.e. } t \in \bar{J}.$$

We note that this definition implies integrability of the strong solution u itself and of Au on J, i.e.,

$$\int_J \big(\|u(t)\|_H + \|Au(t)\|_H\big)\,\mathrm{d}t < \infty.$$

In addition, the following hold for almost every $t \in \bar{J}$:

$$u(t) \in \mathcal{D}(A), \qquad u(t) = u_0 - \int_0^t Au(s)\,\mathrm{d}s + \int_0^t f(s)\,\mathrm{d}s.$$

Since $-A$ is the generator of an analytic C_0-semigroup one knows that the initial value problem (IVP) has a unique strong solution for $u_0 \in H$ if f is locally α-Hölder continuous on $(0,T]$ with exponent $\alpha > 0$, see [31, §4, Cor. 3.3].

However, a solution concept for (IVP) which admits a unique solution for every $f \in L_1(J;H)$ would be preferable. The fact that the unique classical solution to the homogeneous problem for $u_0 \in H$ is given by $u(t) = S(t)u_0$ motivates to define generalized solutions in terms of the analytic C_0-semigroup $(S(t), t \geq 0)$ generated by $-A$. More precisely, for $f \in L_1(J;H)$, we let the function $u\colon \bar{J} \to H$ be defined by

$$u(t) := S(t)u_0 + \int_0^t S(t-s)f(s)\,\mathrm{d}s, \quad t \in \bar{J}.$$

Then u is continuous, i.e., $u \in C(\bar{J};H)$, but in general not differentiable and it does not necessarily take values in $\mathcal{D}(A)$ on J, see [26, 31]. Furthermore, if (IVP) has a classical solution, then it is given by this function. In this way, it may be considered as a generalized solution called the *mild solution* of (IVP).

As already mentioned, every classical solution is a mild solution. Since we assume that the semigroup $(S(t), t \geq 0)$ is analytic, it is furthermore ensured that for every $u_0 \in H$, the mild solution is also a classical solution if f is locally α-Hölder continuous on $(0,T]$ with exponent $\alpha > 0$, see [31, §4, Cor. 3.3].

3.3. The variational approach: weak solutions. In [36, 37] it has been proposed to treat the initial value problem (IVP) with variational formulations posed on Bochner and vector-valued Sobolev spaces as trial and test spaces, cf. the spaces introduced in §2.5. These spaces have to be balanced in such a way that the resulting solution operator is a bijection between the dual of the test space and the trial space, since then existence and uniqueness of a variational solution in the trial space is ensured.

In order to introduce the trial and test spaces of the variational problems below, we assume that A is a densely defined, self-adjoint, positive definite linear operator with a compact inverse as in §2.4. In addition, we recall the Gelfand triple $V \hookrightarrow H \cong H^* \hookrightarrow V^*$, as well as the Hilbert spaces from §2.5:

$$\mathcal{X} = L_2(J;V), \qquad \widehat{\mathcal{Y}} = L_2(J;V) \cap H^1(J;V^*),$$

and the closed linear subspaces $\mathcal{Y}_0, \mathcal{Y} \subset \widehat{\mathcal{Y}}$ of functions vanishing at time $t = 0$ and $t = T$, respectively.

There are different ways to derive a well-posed variational formulation of the initial value problem (IVP), but all of them have in common that the resulting solution concept is less restrictive compared to the notions of the classical and strong solutions presented in §§3.1–3.2.

For the first approach, let $f \in L_2(J;V^*)$. Furthermore, we assume that

(i) the solution u to (IVP) has a square-integrable V^*-valued distributional derivative $\partial_t u$,
(ii) u takes values in $V = \mathcal{D}(A^{1/2})$ almost everywhere in $\bar{J}$, and
(iii) u is square-integrable on J with respect to V.

In other words, u is a well-defined element of the space $\widehat{\mathcal{Y}}$. If the initial value u_0 equals 0, a variational formulation of (IVP) is given by [37]:

(VP1) $\qquad$ Find $\quad u \in \mathcal{Y}_0 \quad$ s.t. $\quad b_0(u,v) = \ell_0(v) \quad \forall v \in \mathcal{X},$

where the bilinear form $b_0 \colon \mathcal{Y}_0 \times \mathcal{X} \to \mathbb{R}$ is defined by

$$b_0(w,v) := \int_J \langle \partial_t w(t) + Aw(t), v(t) \rangle \, \mathrm{d}t, \quad w \in \mathcal{Y}_0, \ v \in \mathcal{X},$$

and $\ell_0(v) := \int_J \langle f(t), v(t) \rangle \, \mathrm{d}t$ for $v \in \mathcal{X}$.

In order to cope with non-vanishing initial data $u(0) = u_0 \neq 0$, one can, e.g., choose a function $u_p \in \widehat{\mathcal{Y}}$ with $u_p(0) = u_0$ and consider the problem

$$\text{Find} \quad \widetilde{u} \in \mathcal{Y}_0 \quad \text{s.t.} \quad b_0(\widetilde{u}, v) = \widetilde{\ell}_0(v) \quad \forall v \in \mathcal{X},$$

where $\widetilde{\ell}_0(v) := \ell_0(v) - b_0(u_p, v)$ for $v \in \mathcal{X}$. Then, the function $u := u_p + \widetilde{u} \in \widehat{\mathcal{Y}}$ solves the initial value problem (IVP) in the variational sense.

A problem arising with this approach is that the function $u_p \in \widehat{\mathcal{Y}}$ has to be constructed and, in particular, this function has to have V-regularity almost everywhere. Thus, e.g., the choice $u_p(t) := e^{-t} u_0$ is only admissible for $u_0 \in V$.

If $u_0 \in H \setminus V$ one can, e.g., enforce the initial value with a multiplier in the variational problem, see [36]. To this end, we consider the Banach space

$$\widehat{\mathcal{X}} := \{(v,\phi) : v \in \mathcal{X}, \ \phi \in H\}, \qquad \|(v,\phi)\|_{\widehat{\mathcal{X}}} := \left(\|v\|_{\mathcal{X}}^2 + \|\phi\|_H^2\right)^{1/2},$$

equipped with the following algebraic operations:

$$(v,\phi) + (w,\psi) := (v+w, \phi+\psi), \qquad \lambda(v,\phi) := (\lambda v, \lambda\phi) \quad \forall \lambda \in \mathbb{R}.$$

We define the bilinear form $\widehat{b} \colon \widehat{\mathcal{Y}} \times \widehat{\mathcal{X}} \to \mathbb{R}$ for $w \in \widehat{\mathcal{Y}}$ and $(v,\phi) \in \widehat{\mathcal{X}}$ by

$$\widehat{b}(w,(v,\phi)) := \int_J \langle \partial_t w(t) + Aw(t), v(t) \rangle \, \mathrm{d}t + (w(0), \phi)_H,$$

where again $\partial_t w$ denotes the distributional derivative of w. A variational solution to (IVP) is then given by the function u satisfying

(VP2) $\qquad$ Find $\quad u \in \widehat{\mathcal{Y}} \quad$ s.t. $\quad \widehat{b}(u,(v,\phi)) = \widehat{\ell}(v,\phi) \quad \forall (v,\phi) \in \widehat{\mathcal{X}},$

where $\widehat{\ell}(v,\phi) := \int_J \langle f(t), v(t) \rangle \, \mathrm{d}t + (u_0, \phi)_H$ for $(v,\phi) \in \widehat{\mathcal{X}}$.

An alternative approach, presented in [37], is to choose the trial and test spaces in such a way that the variational problem incorporates the initial condition as a "natural boundary condition". For this purpose, we first note that the following integration by parts formula holds for functions $w, v \in \widehat{\mathcal{Y}}$ with V^*-valued distributional derivatives $\partial_t w$ and $\partial_t v$:

$$\int_J \langle \partial_t w(t), v(t) \rangle \, \mathrm{d}t = -\int_J \langle w(t), \partial_t v(t) \rangle \, \mathrm{d}t + (w(T), v(T))_H - (w(0), v(0))_H.$$

After multiplying the initial value problem (IVP) with a test function $v \in \mathcal{Y}$ (i.e., $v(T) = 0$) and integrating over J, the application of the above identity

yields the following variational problem:

(VP3) Find $u \in \mathcal{X}$ s.t. $b(u,v) = \ell(v) \quad \forall v \in \mathcal{Y}$,

with the bilinear form $b\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ defined by

$$b(w,v) := \int_J \langle w(t), -\partial_t v(t) + Av(t) \rangle \,\mathrm{d}t, \quad w \in \mathcal{X}, v \in \mathcal{Y},$$

and the right-hand side $\ell(v) := \int_J \langle f(t), v(t) \rangle \,\mathrm{d}t + (u_0, v(0))_H$. In this way, we have incorporated the term arising from the initial value in the functional ℓ. Furthermore, compared to (VP1)–(VP2), we have moved the distributional derivative ∂_t from the trial function to the test function. Therefore, we call (VP3) a weak variational formulation and its solution a *weak (variational) solution*.

In the following theorem we address well-posedness of the three presented variational problems.

THEOREM 3.1. *The bilinear forms $b_0\colon \mathcal{Y}_0 \times \mathcal{X} \to \mathbb{R}$, $\widehat{b}\colon \widehat{\mathcal{Y}} \times \widehat{\mathcal{X}} \to \mathbb{R}$, and $b\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ in* (VP1), (VP2), *and* (VP3) *are continuous and there exist constants $\gamma_0, \widehat{\gamma}, \gamma > 0$ such that the following inf-sup and surjectivity conditions are satisfied:*

$$\inf_{w \in S(\mathcal{Y}_0)} \sup_{v \in S(\mathcal{X})} b_0(w,v) \geq \gamma_0, \qquad \forall v \in \mathcal{X} \setminus \{0\} : \quad \sup_{w \in S(\mathcal{Y}_0)} b_0(w,v) > 0,$$

$$\inf_{w \in S(\widehat{\mathcal{Y}})} \sup_{v \in S(\widehat{\mathcal{X}})} \widehat{b}(w,v) \geq \widehat{\gamma}, \qquad \forall v \in \widehat{\mathcal{X}} \setminus \{0\} : \quad \sup_{w \in S(\widehat{\mathcal{Y}})} \widehat{b}(w,v) > 0,$$

$$\inf_{w \in S(\mathcal{X})} \sup_{v \in S(\mathcal{Y})} b(w,v) \geq \gamma, \qquad \forall v \in \mathcal{Y} \setminus \{0\} : \quad \sup_{w \in S(\mathcal{X})} b(w,v) > 0.$$

Furthermore, for any $f \in L_2(J;V^*)$, the functionals ℓ_0 in (VP1) *and $\widehat{\ell}$ in* (VP2) *are linear and continuous on $\mathcal{X}$ and $\widehat{\mathcal{X}}$, respectively. More precisely, the following estimates hold:*

$$\|\ell_0\|_{\mathcal{X}'} \leq \|f\|_{L_2(J;V^*)}, \qquad \|\widehat{\ell}\|_{\widehat{\mathcal{X}}'} \leq \|f\|_{L_2(J;V^*)} + \|u_0\|_H.$$

The functional ℓ in (VP3) *is continuous on $\mathcal{Y}$ for every $f \in \mathcal{Y}^*$ and $u_0 \in H$ with $\|\ell\|_{\mathcal{Y}'} \leq \|f\|_{\mathcal{Y}^*} + \|u_0\|_H$, where $\mathcal{Y}^*$ denotes the identification of the dual of $\mathcal{Y}$ via the inner product on $L_2(J;H)$.*

PROOF. For the proof of the inf-sup and surjectivity conditions, see [36, Thm. 5.1] and [37, Thm. 2.2].

The bounds for $\|\ell_0\|_{\mathcal{X}'}$ and $\|\widehat{\ell}\|_{\widehat{\mathcal{X}}'}$ are readily seen. In order to derive the bound for $\|\ell\|_{\mathcal{Y}'}$, we recall that the embedding constant in $\mathcal{Y} \hookrightarrow C(\bar{J};H)$ equals one, see §2.5. Thus, we obtain for $v \in \mathcal{Y}$:

$$|\ell(v)| \leq \|f\|_{\mathcal{Y}^*} \|v\|_{\mathcal{Y}} + \|u_0\|_H \|v(0)\|_H \leq (\|f\|_{\mathcal{Y}^*} + \|u_0\|_H) \|v\|_{\mathcal{Y}}. \qquad \square$$

We close this section by drawing some conclusions from Theorem 3.1.

- The bilinear forms b_0, $\widehat{b}$, and b induce boundedly invertible continuous linear operators $b_0 \in \mathcal{L}(\mathcal{Y}_0; \mathcal{X}')$, $\widehat{b} \in \mathcal{L}(\widehat{\mathcal{Y}}; \widehat{\mathcal{X}}')$, and $b \in \mathcal{L}(\mathcal{X}; \mathcal{Y}')$, where we use the same notation for the operators as for the bilinear forms, since it will

be evident from the context to which we refer. Therefore, the variational problems (VP1), (VP2), and (VP3) are uniquely solvable.

For (VP3), the data-to-solution mapping $(f, u_0) \mapsto u$, where $u \in \mathcal{X}$ denotes the solution to (VP3), satisfies the stability bound

$$\|u\|_{\mathcal{X}} \leq \gamma^{-1}\|\ell\|_{\mathcal{Y}'} \leq \gamma^{-1}(\|f\|_{\mathcal{Y}^*} + \|u_0\|_H),$$

and analogous results hold for (VP1) and (VP2).

- Recall the equivalent norm $|||\cdot|||_{\mathcal{Y}}$ on $\mathcal{Y}$ from §2.5:

$$|||v|||_{\mathcal{Y}} = \Big(\|v\|^2_{L_2(J;V)} + \|\partial_t v\|^2_{L_2(J;V^*)} + \|v(0)\|^2_H\Big)^{1/2}.$$

With respect to this norm on $\mathcal{Y}$, the induced operator $b\colon \mathcal{X} \to \mathcal{Y}'$ is an isometric isomorphism.

To verify the isometry property, we first emphasize the identity

$$|||v|||_{\mathcal{Y}} = \| - \partial_t v + Av\|_{L_2(J;V^*)} \quad \forall v \in \mathcal{Y}.$$

Thus, $|||bw|||_{\mathcal{Y}'} := \sup_{v\in\mathcal{Y}\setminus\{0\}} \frac{|b(w,v)|}{|||v|||_{\mathcal{Y}}} \leq \|w\|_{\mathcal{X}}$ for $w \in \mathcal{X}$ follows from the definition of b. It remains to verify $|||bw|||_{\mathcal{Y}'} \geq \|w\|_{\mathcal{X}}$ or, equivalently, that $\|b^{-1}\ell\|_{\mathcal{X}} \leq |||\ell|||_{\mathcal{Y}'}$ holds for every $\ell \in \mathcal{Y}'$.

Since the inf-sup constant γ of b is positive and $|||\cdot|||_{\mathcal{Y}}$ defines an equivalent norm on $\mathcal{Y}$, one knows that the inf-sup constant $\gamma_{|||}$ of b with respect to $|||\cdot|||_{\mathcal{Y}}$ on $\mathcal{Y}$ inherits the positivity: $\gamma_{|||} > 0$. Moreover, the stability bound $\|b^{-1}\ell\|_{\mathcal{X}} \leq \gamma_{|||}^{-1}|||\ell|||_{\mathcal{Y}'}$, for $\ell \in \mathcal{Y}'$, and the following equalities hold:

$$\gamma_{|||} = \inf_{w\in\mathcal{X}} \sup_{v\in\mathcal{Y}} \frac{b(w,v)}{\|w\|_{\mathcal{X}}|||v|||_{\mathcal{Y}}} = \inf_{v\in\mathcal{Y}} \sup_{w\in\mathcal{X}} \frac{b(w,v)}{\|w\|_{\mathcal{X}}|||v|||_{\mathcal{Y}}}.$$

For $v \in \mathcal{Y}$, we set $w := v - (A^*)^{-1}\partial_t v$ to obtain $\|w\|_{\mathcal{X}} = |||v|||_{\mathcal{Y}}$ and

$$\begin{aligned} b(w,v) &= \|v\|^2_{L_2(J;V)} + \|\partial_t v\|^2_{L_2(J;V^*)} - 2\int_J \langle \partial_t v(t), v(t)\rangle \,\mathrm{d}t \\ &= \|v\|^2_{\mathcal{Y}} + \|v(0)\|^2_H = |||v|||^2_{\mathcal{Y}} = \|w\|_{\mathcal{X}}|||v|||_{\mathcal{Y}}. \end{aligned}$$

This shows that $\gamma_{|||} \geq 1$ and the isometry property of $b\colon \mathcal{X} \to \mathcal{Y}'$ with respect to the norms $\|\cdot\|_{\mathcal{X}}$ and $|||\cdot|||_{\mathcal{Y}'}$ follows.

Note also that this result in conjunction with the relation $\|v\|_{\mathcal{Y}} \leq |||v|||_{\mathcal{Y}}$, $v \in \mathcal{Y}$, implies the following lower bound for the inf-sup constant of the bilinear form b with respect to the norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Y}}$:

$$\inf_{w\in S(\mathcal{X})} \sup_{v\in S(\mathcal{Y})} b(w,v) = \inf_{w\in\mathcal{X}} \sup_{v\in\mathcal{Y}} \frac{b(w,v)}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} \geq 1.$$

We conclude that $\gamma \geq \gamma_{|||} = 1$.

- The assumptions on the source term f and the initial value u_0 made in the latter variational approach (VP3) for a well-defined weak solution of (IVP) are not more restrictive than needed for a mild solution in §3.2, since $\ell \in \mathcal{Y}'$ for every $f \in L_1(J;H)$ and $u_0 \in H$.

4. Stochastic calculus in Hilbert spaces

In order to introduce the stochastic differential equations of interest, in this section we first recall certain notions and concepts from probability theory. Specifically, in §4.1 we start by defining random variables with values in Hilbert spaces, as well as Gaussian measures, and Gaussian white noise on Hilbert spaces. We then proceed with classes of vector-valued stochastic processes in §4.2. In addition, we summarize basic definitions and results from Itô integration and Itô calculus in §§4.3–4.4. This establishes the framework for defining solutions to parabolic stochastic differential equations in §4.5.

From here on, let $(\Omega, \mathcal{A}, \mathbb{P})$ denote a complete probability space equipped with the filtration $\mathcal{F} := (\mathcal{F}_t, t \in \mathcal{I})$ which satisfies the "ususal conditions", i.e.,

(i) $\mathcal{F}$ is right continuous, i.e., $\mathcal{F}_t = \mathcal{F}_{t^+} := \bigcap_{s>t} \mathcal{F}_s$ for all $t \in \mathcal{I}$;
(ii) $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets of $\mathcal{A}$.

For our purposes, the index set $\mathcal{I}$ is either the nonnegative part of the real axis $\mathcal{I} := \{t \in \mathbb{R} : t \geq 0\}$ or the closed finite time interval $\mathcal{I} := \bar{J} = [0, T]$.

Throughout this section, we write $s \wedge t := \min\{s, t\}$ for $s, t \in \mathbb{R}$, and we mark equalities which hold $\mathbb{P}$-almost surely with $\mathbb{P}$-a.s.

4.1. Random variables in Hilbert spaces. The purpose of this subsection is to define random variables which take values in the separable Hilbert space U. In order to introduce the notion of Gaussian white noise on U, which is necessary for Papers III–V in the second part of the thesis, we furthermore generalize this definition to the concept of weak random variables. For an overview of vector-valued random variables, Gaussian measures on Hilbert spaces, and white noise theory, we refer to [3, 10, 32].

4.1.1. *Random variables.* Any measurable function $Z\colon (\Omega, \mathcal{A}) \to (U, \mathcal{B}(U))$ is called a U-valued random variable. The distribution of Z, defined as the image measure $\mathbb{P} \circ Z^{-1}$ of $\mathbb{P}$ under Z, is a probability measure on $(U, \mathcal{B}(U))$.

For any integrable U-valued random variable, i.e., $\int_\Omega \|Z(\omega)\|_U \,\mathrm{d}\mathbb{P}(\omega) < \infty$, we call the Bochner integral

$$\mathbb{E}[Z] := \int_\Omega Z(\omega) \,\mathrm{d}\mathbb{P}(\omega) \in U$$

the *expectation* or the *mean* of Z.

In addition, if $Z \in L_2(\Omega; U)$, i.e., $\mathbb{E}[\|Z\|_U^2] < \infty$, the *covariance* of Z,

$$\mathrm{Cov}(Z) := \mathbb{E}[(Z - \mathbb{E}[Z]) \otimes (Z - \mathbb{E}[Z])],$$

is an element of the projective tensor space U_π^{sym}, see §§1.2–1.3, since

$$\|\mathrm{Cov}(Z)\|_\pi \leq \mathbb{E}[\|(Z - \mathbb{E}[Z]) \otimes (Z - \mathbb{E}[Z])\|_\pi] = \mathbb{E}[\|Z - \mathbb{E}[Z]\|_U^2] < \infty.$$

By Proposition 1.2, it can thus be identified with an operator $Q \in \mathcal{L}_1^{\mathrm{sym}}(U)$, which is referred to as the covariance operator of Z. The covariance operator Q inherits the positive semi-definiteness from the covariance,

$$(\mathrm{Cov}(Z), x \otimes x)_2 = \mathbb{E}[(Z - \mathbb{E}[Z], x)_U^2] \geq 0 \quad \forall x \in U \quad \Rightarrow \quad Q \in \mathcal{L}_1^+(U).$$

Recall the subspace U_π^+ of the projective tensor product space U_π from §1.3. Proposition 1.2 also implies that the spaces U_π^+ and $\mathcal{L}_1^+(U)$ are isometrically isomorphic. Therefore, there is a one-to-one correspondence between covariances (covariance kernels) in U_π^+ and covariance operators in $\mathcal{L}_1^+(U)$.

We remark that, historically, the name "kernel" originates from the special but frequently used case $U = L_2(D)$ for some domain $D \subset \mathbb{R}^d$, $d \in \mathbb{N}$. By the definitions above, the covariance $k_Z := \operatorname{Cov}(Z)$ of $Z \in L_2(\Omega; L_2(D))$ satisfies

$$\begin{aligned}(k_Z, f \otimes g)_2 &= \int_{D\times D} k_Z(\mathbf{x},\mathbf{y}) f(\mathbf{x}) g(\mathbf{y})\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \\ &= \mathbb{E}\left[\int_D (Z(\mathbf{x}) - \mathbb{E}[Z(\mathbf{x})]) f(\mathbf{x})\,\mathrm{d}\mathbf{x} \int_D (Z(\mathbf{y}) - \mathbb{E}[Z(\mathbf{y})]) g(\mathbf{y})\,\mathrm{d}\mathbf{y}\right]\end{aligned}$$

for all $f, g \in L_2(D)$. In particular, if the random variable Z is ($\mathbb{P}$-a.s.) continuous on the closure $\bar{D}$ of the domain D, then its covariance is continuous on $\bar{D} \times \bar{D}$, and

$$k_Z(\mathbf{x},\mathbf{y}) = \mathbb{E}[(Z(\mathbf{x}) - \mathbb{E}[Z(\mathbf{x})])(Z(\mathbf{y}) - \mathbb{E}[Z(\mathbf{y})])] = \operatorname{Cov}(Z(\mathbf{x}), Z(\mathbf{y}))$$

for every $\mathbf{x}, \mathbf{y} \in \bar{D}$. Therefore, the covariance k_Z has the form of the kernels considered, e.g., in Mercer's theorem [28].

4.1.2. *Gaussian measures.* A probability measure μ on $(U, \mathcal{B}(U))$ is called a Gaussian measure if, for every $x \in U$, the measurable function $(x, \cdot\,)_U$ is normally distributed, i.e., there exist numbers $m_x \in \mathbb{R}$ and $\sigma_x^2 \geq 0$ such that, for all $a \in \mathbb{R}$,

$$\mu(\{y \in U : (x,y)_U \leq a\}) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \int_{-\infty}^{a} e^{-\frac{(r-m_x)^2}{2\sigma_x^2}}\,\mathrm{d}r.$$

For $\sigma_x^2 = 0$, the normal distribution is degenerate and

$$\mu(\{y \in U : (x,y)_U \leq a\}) = \begin{cases} 1 & \text{if } m_x \leq a, \\ 0 & \text{otherwise.} \end{cases}$$

If μ is a Gaussian measure on U, then there exist a unique element $m \in U$ and a unique self-adjoint, positive semi-definite trace class operator $Q \in \mathcal{L}_1^+(U)$ such that, for all $x, y \in U$,

$$\int_U (x,z)_U\,\mu(\mathrm{d}z) = (x,m)_U,$$

$$\int_U (x,z)_U (y,z)_U\,\mu(\mathrm{d}z) - (x,m)_U (y,m)_U = (Qx,y)_U.$$

We call m the mean and Q the covariance operator of the measure μ.

Another characterization of Gaussian measures can be made in terms of their characteristic functions. For a probability measure μ on $(U, \mathcal{B}(U))$, its characteristic function φ_μ is defined by

$$\varphi_\mu(x) := \int_U e^{i(x,y)_U}\,\mu(\mathrm{d}y), \quad x \in U.$$

A probability measure μ on $(U, \mathcal{B}(U))$ with mean $m \in U$ and covariance operator $Q \in \mathcal{L}_1^+(U)$ is then a Gaussian measure if and only if

$$\varphi_\mu(x) = e^{i(x,m)_U - \frac{1}{2}(Qx,x)_U}, \quad x \in U.$$

Thus, a Gaussian measure μ is uniquely determined by its mean $m \in U$ and its covariance operator $Q \in \mathcal{L}_1^+(U)$, and we let $\mathcal{N}(m, Q)$ denote its distribution.

Finally, a mapping $Z\colon \Omega \to U$ is called a U-valued Gaussian random variable if there exist $m \in U$ and $Q \in \mathcal{L}_1^+(U)$ such that its distribution is a Gaussian measure, i.e., $\mathbb{P} \circ Z^{-1} \sim \mathcal{N}(m, Q)$. In this case, for all $x \in U$, the real-valued random variable $(x, Z)_U$ is normally distributed with

$$\begin{aligned} \mathbb{E}[(x, Z)_U] &= (x, m)_U, \\ \mathbb{E}[(x, Z - m)_U (y, Z - m)_U] &= (Qx, y)_U, \\ \mathbb{E}[\|Z - m\|_U^2] &= \operatorname{tr}(Q). \end{aligned}$$

The characteristic function of Z, i.e., $\varphi_Z(x) := \mathbb{E}[e^{i(x,Z)_U}]$, is then given by

$$\varphi_Z(x) = e^{i(x,m)_U - \frac{1}{2}(Qx,x)_U}, \quad x \in U,$$

and the distribution of Z is uniquely determined by its mean m and its covariance operator Q, and we write $Z \sim \mathcal{N}(m, Q)$.

Let $Z \sim \mathcal{N}(0, Q)$ be a zero-mean U-valued Gaussian random variable, and let $Q^{-1/2}$ denote the operator pseudo-inverse of the well-defined fractional power covariance operator $Q^{1/2}$, cf. §2.3. Note that the operators $Q^{1/2}$ and $Q^{-1/2}$ allow for the spectral expansion

$$Q^{\pm 1/2} x := \sum_{j \in \mathbb{J}} \gamma_j^{\pm 1/2} (x, e_j)_U \, e_j, \quad x \in U,$$

with respect to an orthonormal basis $\{e_j\}_{j \in \mathbb{N}} \subset U$ consisting of eigenvectors of $Q \in \mathcal{L}_1^+(U)$ with corresponding nonnegative eigenvalues $\{\gamma_j\}_{j \in \mathbb{N}}$, see Theorem 1.1. The index set is given by $\mathbb{J} := \{j \in \mathbb{N} : \gamma_j \neq 0\}$. The reproducing kernel Hilbert space $\mathcal{H}$ (RKHS for short) of $Z \sim \mathcal{N}(0, Q)$ is then defined by

$$\mathcal{H} := Q^{1/2} U = \operatorname{Rg}(Q^{1/2}), \qquad (x, y)_{\mathcal{H}} := (Q^{-1/2} x, Q^{-1/2} y)_U, \quad x, y \in \mathcal{H},$$

cf. [10, §2.2.2]. This definition is motivated by the fact that

$$(x, Z)_U \sim \mathcal{N}(0, \|x\|_{\mathcal{H}^*}^2) \quad \forall x \in U, \quad \text{where} \quad \|x\|_{\mathcal{H}^*} := \sup_{y \in S(\mathcal{H})} (x, y)_U,$$

i.e., the RKHS describes the values of U that Z attains.

4.1.3. *Weak random variables and Gaussian white noise.* In §§4.1.1–4.1.2 we have discussed random variables and Gaussian measures on $(U, \mathcal{B}(U))$ with self-adjoint, positive semi-definite trace class covariance operators. As observed in §4.1.1, the covariance operator of a U-valued random variable is necessarily of this form. Similarly, one can show that, if μ is a Gaussian measure on U, its covariance operator Q has to be of trace class, since otherwise μ fails to be countably additive with respect to the Borel σ-algebra $\mathcal{B}(U)$.

It is, however, possible to consider the generalization to bounded covariance operators,

$$Q \in \mathcal{L}^+(U) := \{T \in \mathcal{L}(U) : T = T^*,\ (Tx, x)_U \geq 0 \quad \forall x \in U\}$$

for probability measures on $(U, \mathcal{C}(U))$, where $\mathcal{C}(U)$ is the σ-algebra generated by the cylinder sets of U. Here $C \subset U$ is a (Borel) cylinder set if there exist finitely many elements $x_1, \ldots, x_n \in U$ and a Borel set B in $\mathbb{R}^n$ such that

$$C = \left\{y \in U : ((y, x_1)_U, \ldots, (y, x_n)_U)^\mathsf{T} \in B\right\}.$$

Measures on $(U, \mathcal{C}(U))$ are called cylinder measures on the Hilbert space U. Note that they are by definition countably additive with respect to $\mathcal{C}(U)$, but not necessarily on the Borel sets $\mathcal{B}(U)$. Thus, every measure with respect to the Borel σ-algebra $\mathcal{B}(U)$ induces a cylinder measure, but the converse is, in general, not true.

A measurable mapping $Z\colon (\Omega, \mathcal{A}) \to (U, \mathcal{C}(U))$, whose distribution $\mathbb{P} \circ Z^{-1}$ is a cylinder measure, may be seen as a generalized U-valued random variable, often referred to as a weak random variable in the literature [3]. This name originates from the fact that a weak random variable Z is weakly measurable, i.e., $g \circ Z\colon (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a real-valued random variable for every functional $g \in U'$, as cylinder sets form neighborhoods in the weak topology of U.

Since the characteristic function of a probability measure involves only one-dimensional projections of elements in U, it is possible to define the characteristic function φ_μ for a cylinder measure μ in the same way as for probability measures on $(U, \mathcal{B}(U))$. In particular, a probability measure μ on $(U, \mathcal{C}(U))$ is a Gaussian cylinder measure with mean $m \in U$ and covariance operator $Q \in \mathcal{L}^+(U)$ if and only if its characteristic function is given by

$$\varphi_\mu(x) = e^{i(x,m)_U - \frac{1}{2}(Qx,x)_U}, \quad x \in U.$$

It is a consequence of Sazanov's theorem [3, Thm. 6.3.1] that a Gaussian cylinder measure μ can be extended to a Gaussian measure on $(U, \mathcal{B}(U))$ if and only if its covariance operator is of trace class, i.e., $Q \in \mathcal{L}_1^+(U)$. We call a weak random variable $Z\colon (\Omega, \mathcal{A}) \to (U, \mathcal{C}(U))$, whose distribution is a Gaussian cylinder measure with covariance operator $Q \in \mathcal{L}^+(U) \setminus \mathcal{L}_1^+(U)$, a U-valued Gaussian weak random variable. Finally, we define Gaussian white noise $\mathcal{W}$ on U as a zero-mean Gaussian weak random variable, whose characteristic function is given by

$$\varphi_{\mathcal{W}}(x) = \mathbb{E}[e^{i(x,\mathcal{W})_U}] = e^{-\frac{1}{2}\|x\|_U^2}, \quad x \in U,$$

i.e., $m = 0$ and the covariance operator of $\mathcal{W}$ is the identity, $Q = I \in \mathcal{L}^+(U)$.

Note that the RKHS $\mathcal{H}$ of $\mathcal{W}$ coincides with U, and that, for every $x \in U$, the real-valued random variable $(x, \mathcal{W})_U$ is normally distributed with

$$\mathbb{E}[(x, \mathcal{W})_U] = 0, \qquad \mathbb{E}[(x, \mathcal{W})_U (y, \mathcal{W})_U] = (x, y)_U \quad \forall y \in U.$$

For this reason, white noise on U is sometimes also defined as the isometry $\mathcal{J}_{\mathcal{W}}\colon U \to L_2(\Omega; \mathbb{R})$ such that, for every $x \in U$, $\mathcal{J}_{\mathcal{W}}(x)$ is a centered Gaussian random variable satisfying $\mathbb{E}[\mathcal{J}_{\mathcal{W}}(x)\mathcal{J}_{\mathcal{W}}(y)] = (x, y)_U$, see [29, Def. 1.1.1].

A third interpretation of Gaussian white noise $\mathcal{W}$ is to regard it as a limit $\mathcal{W} = \lim_{N\to\infty} \mathcal{W}_N$ of well-defined U-valued Gaussian random variables $\mathcal{W}_N$, where this convergence holds in a weak sense that has to be specified. To this end, let $\{e_j\}_{j\in\mathbb{N}}$ be any orthonormal basis of the separable Hilbert space U and let $\{\xi_j\}_{j\in\mathbb{N}}$ be a sequence of independent $\mathcal{N}(0,1)$-distributed real-valued random variables. For $N \in \mathbb{N}$, we set

$$\mathcal{W}_N := \sum_{j=1}^{N} \xi_j e_j, \qquad \mathcal{J}_N(x) := (x, \mathcal{W}_N)_U, \quad x \in U.$$

Then, for every finite N, the U-valued random variable $\mathcal{W}_N$ is Gaussian with zero-mean and a covariance operator Q_N satisfying $\operatorname{tr}(Q_N) = N < \infty$. Furthermore, for every $x \in U$, the sequence $\{\mathcal{J}_N(x)\}_{N\in\mathbb{N}}$ of real-valued random variables converges $\mathbb{P}$-a.s. and in any L^p-sense to an $\mathcal{N}(0, \|x\|_U^2)$-distributed random variable. Letting $\mathcal{J}_{\mathcal{W}}(x) := \lim_{N\to\infty} \mathcal{J}_N(x)$ is an explicit way of constructing the isometry $\mathcal{J}_{\mathcal{W}} \colon U \to L_2(\Omega;\mathbb{R})$ mentioned above. The series expansion

$$\mathcal{W} = \sum_{j\in\mathbb{N}} \xi_j \, e_j,$$

which converges in this weak sense, is called a (formal) Karhunen–Loève expansion of the white noise $\mathcal{W}$. Note that this series does not converge in $L^2(\Omega;U)$ if U is an infinite-dimensional Hilbert space.

We close this section with the classical idea of Gaussian white noise on a domain $D \subset \mathbb{R}^d$. This is to construct noise $\mathcal{W}$ on D such that the real-valued random variables $\mathcal{W}(O_1) := \int_{O_1} \mathcal{W}(\mathbf{x})\,\mathrm{d}\mathbf{x}$ and $\mathcal{W}(O_2)$ are independent and Gaussian distributed for every distinct open non-trivial subsets $O_1, O_2 \subset D$. By letting $\mathcal{W}$ be Gaussian white noise on $U = L_2(D)$, we obtain this property:

$$\mathbb{E}[\mathcal{W}(O_1)\mathcal{W}(O_2)] = \mathbb{E}[(\mathcal{W}, \mathbb{1}_{O_1})_{L_2(D)}(\mathcal{W}, \mathbb{1}_{O_2})_{L_2(D)}] = (\mathbb{1}_{O_1}, \mathbb{1}_{O_2})_{L_2(D)} = 0,$$

where $\mathbb{1}_O$ denotes the indicator function of a set $O \subset D$.

4.2. Stochastic processes. In this subsection we present the classes of stochastic processes as well as their characteristics which are of interest for our investigations. Since not needed in greater generality throughout the thesis, we restrict this presentation to Wiener processes and square-integrable Lévy processes with trace class covariance operators. We refer the reader to, e.g., [32, §§3, 4] for a detailed introduction to stochastic processes taking values in Hilbert spaces and, in particular, to Lévy processes.

A U-valued stochastic process is defined as a family $(X(t), t \in \mathcal{I})$ of U-valued random variables. For our purposes, the following characterizations of a stochastic process $X := (X(t), t \in \mathcal{I})$ taking values in U will be important:

- *Integrability*: X is said to be integrable if

$$\|X(t)\|_{L_1(\Omega;U)} := \mathbb{E}[\|X(t)\|_U] < \infty \quad \forall t \in \mathcal{I},$$

and square-integrable if

$$\|X(t)\|_{L_2(\Omega;U)} := \mathbb{E}[\|X(t)\|_U^2]^{1/2} < \infty \quad \forall t \in \mathcal{I}.$$

- *Mean*: If X is integrable, the U-valued mapping
$$m\colon \mathcal{I} \to H, \qquad t \mapsto \mathbb{E}[X(t)]$$
is well-defined and it is called the mean or the first moment of X.
- *Second moment and covariance*: Assuming that X is square-integrable, the tensor-space-valued functions $M, C\colon \mathcal{I} \times \mathcal{I} \to U_\pi$ defined by
$$M(s,t) := \mathbb{E}[X(s) \otimes X(t)],$$
$$C(s,t) := \mathbb{E}[(X(s) - \mathbb{E}[X(s)]) \otimes (X(t) - \mathbb{E}[X(t)])], \quad s,t \in \mathcal{I},$$
are called the second moment and the covariance of X, respectively.

We note that the covariance can be expressed in terms of the second moment and the mean:
$$C(s,t) = M(s,t) - m(s) \otimes m(t).$$
Furthermore, the second moment and the covariance are well-defined mappings to the projective tensor product space U_π (cf. §4.1.1), since
$$\begin{aligned}\|\mathbb{E}[X(s) \otimes X(t)]\|_\pi &\le \mathbb{E}[\|X(s) \otimes X(t)\|_\pi] = \mathbb{E}[\|X(s)\|_U \|X(t)\|_U] \\ &\le \|X(s)\|_{L_2(\Omega;U)} \|X(t)\|_{L_2(\Omega;U)},\end{aligned}$$
for all $s,t \in \mathcal{I}$, where the first estimate holds by the properties of the Bochner integral and the last one by Hölder's inequality.

In the following we present certain classes of stochastic processes which we are going to refer to in the course of the thesis.

4.2.1. *Martingales.* An integrable stochastic process $(X(t), t \in \mathcal{I})$ taking values in U is called a U-valued martingale with respect to $\mathcal{F}$ if it is $\mathcal{F}$-adapted, i.e., $X(t)$ is $\mathcal{F}_t$-measurable for all $t \in \mathcal{I}$, and it satisfies the martingale property: the conditional expectation of $X(t)$ with respect to the σ-field $\mathcal{F}_s$ for $s \le t$ is given by $\mathbb{E}[X(t)|\mathcal{F}_s] = X(s)$.

4.2.2. *Lévy processes.* A U-valued stochastic process $L := (L(t), t \in \mathcal{I})$ is said to be a Lévy process if the following conditions are satisfied:

(i) L has independent increments, i.e., the U-valued random variables $L(t_1) - L(t_0)$, $L(t_2) - L(t_1)$, …, $L(t_n) - L(t_{n-1})$ are independent for all $t_0, \dots, t_n \in \mathcal{I}$, $0 \le t_0 < t_1 < \dots < t_n$;
(ii) L has stationary increments, i.e., the distribution of $L(t) - L(s)$, $s \le t$, $s,t \in \mathcal{I}$, depends only on the difference $t - s$;
(iii) $L(0) = 0$ ($\mathbb{P}$-a.s.);
(iv) L is stochastically continuous in t for all $t \in \mathcal{I}$, i.e.,
$$\lim_{\substack{s \to t \\ s \in \mathcal{I}}} \mathbb{P}(\|L(t) - L(s)\|_U > \epsilon) = 0 \quad \forall \epsilon > 0.$$

We often make some or all of the following assumptions on a Lévy process L:

(a) L is adapted to the filtration $\mathcal{F}$;
(b) for $t > s$ the increment $L(t) - L(s)$ is independent of $\mathcal{F}_s$;
(c) L is integrable;
(d) L has zero-mean, i.e., $\mathbb{E}[L(t)] = 0$ for all $t \in \mathcal{I}$;
(e) L is square-integrable.

Note that the Lévy process L satisfies Assumptions (a)–(b), e.g., for the filtration $\mathcal{F} := (\bar{\mathcal{F}}_t^L, t \in \mathcal{I})$, where $\bar{\mathcal{F}}_t^L$ denotes the smallest σ-field containing the σ-field $\mathcal{F}_t^L := \sigma(L(s) : s \leq t)$ generated by L and all $\mathbb{P}$-null sets of $\mathcal{A}$, [32, Rem. 4.43]. Under Assumptions (a)–(d) L is a martingale, see [32, Prop. 3.25].

If, in addition, L satisfies Assumption (e), then there exists a self-adjoint, positive semi-definite trace class operator $Q \in \mathcal{L}_1^+(U)$ such that, for all $s, t \in \mathcal{I}$,

$$\mathbb{E}\left[(L(s) \otimes L(t), x \otimes y)_2\right] = (s \wedge t)\,(Qx, y)_U \quad \forall x, y \in U,$$

see [32, Thm. 4.44]. This operator is called the covariance operator of the Lévy process L, cf. covariance operators of U-valued random variables in §4.1.1. The RKHS of L is then by definition the RKHS of the random variable $L(1)$ [32, Def. 7.2], which is identically defined as for a U-valued Gaussian random variable with covariance operator Q, see §4.1.2.

In the following, we illustrate how the covariance of a Lévy process relates to the concept of tensor product spaces from §1.2. Suppose that $\mathcal{I} = \bar{J} = [0, T]$ and that L is a U-valued Lévy process satisfying Assumptions (a)–(e) above. Since L has zero-mean, the second moment and the covariance C of L coincide and we obtain

$$\begin{aligned}\int_{J\times J} (C(s,t), w(s) \otimes v(t))_2 \,\mathrm{d}s\,\mathrm{d}t &= \int_{J\times J} (s \wedge t)\,(Qw(s), v(t))_U \,\mathrm{d}s\,\mathrm{d}t \\ &= \int_{J\times J} \Big(\sum_{j\in\mathbb{N}} (s \wedge t)\,\gamma_j\,(e_j \otimes e_j), w(s) \otimes v(t)\Big)_2 \,\mathrm{d}s\,\mathrm{d}t\end{aligned}$$

for all $w, v \in \mathcal{Z} := L_2(J; U)$, where $\{e_j\}_{j\in\mathbb{N}}$ is an orthonormal basis of U consisting of eigenvectors of Q with corresponding nonnegative eigenvalues $\{\gamma_j\}_{j\in\mathbb{N}}$, cf. Theorem 1.1. Therefore, the covariance of L can be represented for almost every $s, t \in \bar{J}$ by

$$C(s,t) = \sum_{j\in\mathbb{N}} (s \wedge t)\,\gamma_j\,(e_j \otimes e_j),$$

with convergence of this series in U_2, since $\operatorname{tr}(Q) = \sum_{j\in\mathbb{N}} \gamma_j < \infty$. Furthermore, the covariance C is even an element of $\mathcal{Z}_\pi$ with $\|C\|_\pi \leq \frac{1}{2} T^2 \operatorname{tr} Q$.

In order to derive the latter bound, set $w_\wedge(s,t) := s \wedge t$, and let $q \in U_\pi$ be the kernel associated with the operator Q. Then $w_\wedge$ is an element of the projective tensor space $L_2(J;\mathbb{R})_\pi$ and $q = \sum_{j\in\mathbb{N}} \gamma_j (e_j \otimes e_j)$. By the characterization of the projective norm derived in §1.3 we find

$$\|w_\wedge\|_\pi = \delta(w_\wedge) = \tfrac{1}{2} T^2, \qquad \|q\|_\pi = \|Q\|_{\mathcal{L}_1(U)} = \operatorname{tr}(Q),$$

since $w_\wedge \in L_2(J;\mathbb{R})_\pi$ and $q \in U_\pi$ are positive semi-definite kernels. Thus, the covariance of L satisfies

$$C = w_\wedge \otimes q \in L_2(J;\mathbb{R})_\pi \otimes U_\pi, \qquad \|C\|_{L_2(J;\mathbb{R})_\pi \hat{\otimes}_\pi U_\pi} = \tfrac{1}{2} T^2 \operatorname{tr}(Q),$$

and the same bound holds for the projective norm on $\mathcal{Z}_\pi$ due to

$$L_2(J;\mathbb{R})_\pi \,\hat{\otimes}_\pi\, U_\pi \cong (L_2(J;\mathbb{R}) \,\hat{\otimes}_\pi\, U)_\pi \hookrightarrow (L_2(J;\mathbb{R}) \,\hat{\otimes}_2\, U)_\pi \cong \mathcal{Z}_\pi,$$

where the first identification is due to the associativity of the projective tensor product [21, Ch. 33]. The embedding holds, since the projective tensor product inherits the continuous embedding from $L_2(J;\mathbb{R}) \hat{\otimes}_\pi U \hookrightarrow L_2(J;\mathbb{R}) \hat{\otimes}_2 U$ and the last identification is a consequence of the definitions of the Hilbert tensor product space $L_2(J;\mathbb{R}) \hat{\otimes}_2 U$ and the Bochner space $\mathcal{Z} = L_2(J;U)$.

4.2.3. *Wiener processes.* An important subclass of Lévy processes is formed by Wiener processes. Here, a zero-mean Lévy process $W := (W(t), t \in \mathcal{I})$ is said to be an U-valued Wiener process if it has ($\mathbb{P}$-a.s.) continuous trajectories in U. Then, for every $t \in \mathcal{I}$, the U-valued random variable $t^{-1/2}W(t)$ is Gaussian distributed with zero-mean and covariance operator $Q \in \mathcal{L}_1^+(U)$, see §4.1.2.

In the finite-dimensional case $U = \mathbb{R}^n$ the covariance is given by

$$\mathbb{E}\left[W(s)W(t)^\mathsf{T}\right] = (s \wedge t)\mathbf{Q} \quad \forall s,t \in \mathcal{I}$$

for a symmetric positive semi-definite matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$. If $\mathbf{Q}$ equals the identity matrix in $\mathbb{R}^{n\times n}$ then W is called Wiener white noise, cf. §4.1.3. In one dimension ($U = \mathbb{R}$) W is called a real-valued Brownian motion if $Q = 1$.

4.3. Stochastic integration. The purpose of this section is to make sense of the H-valued stochastic integral

$$\int_0^t \Psi(s)\,\mathrm{d}L(s), \quad t \in \bar{J} = [0,T],$$

where the noise $L = (L(t), t \in \bar{J})$ is a Lévy process taking values in a separable Hilbert space U and Ψ is a stochastic process taking values in an appropriate space of operators mapping from U to the separable Hilbert space H. Thus, for fixed $t \in \bar{J}$, the stochastic integral itself becomes an H-valued random variable.

To this end, following the lines of [32, §8.2], we first define the stochastic Itô integral for $\mathcal{L}(U;H)$-valued processes Ψ on $\bar{J}$, which are simple, i.e., there exist finite sequences

- of nonnegative numbers $0 = t_0 < t_1 < \ldots < t_N \le T$,
- of operators $\Psi_1, \ldots, \Psi_N \in \mathcal{L}(U;H)$, and
- of events $A_n \in \mathcal{F}_{t_{n-1}}$, $1 \le n \le N$,

such that

$$\Psi(s) = \sum_{n=1}^N \mathbb{1}_{A_n}\, \mathbb{1}_{(t_{n-1},t_n]}(s)\, \Psi_n, \quad s \in \bar{J},$$

where $\mathbb{1}_{\,\cdot}$ denotes the indicator function of a set (here, an event or an interval).

Let the U-valued Lévy process L satisfy Assumptions (a)–(e) from §4.2.2. Then the stochastic integral with respect to the simple process Ψ and the Lévy noise L is defined by

$$\int_0^t \Psi(s)\,\mathrm{d}L(s) := \sum_{n=1}^N \mathbb{1}_{A_n} \Psi_n (L(t_n \wedge t) - L(t_{n-1} \wedge t)) \quad \forall t \in \bar{J}.$$

The so-constructed stochastic integral is called an Itô integral. Recall the RKHS $\mathcal{H} = Q^{1/2}U$ of the Lévy process L, equipped with $(\,\cdot\,,\,\cdot\,)_\mathcal{H} = (Q^{-1/2}\,\cdot\,, Q^{-1/2}\,\cdot\,)_U$.

The Itô integral has an important property—the Itô isometry, see [32, Prop. 8.6]:

$$\mathbb{E}\left[\left\|\int_0^t \Psi(s)\,\mathrm{d}L(s)\right\|_H^2\right] = \mathbb{E}\left[\int_0^t \|\Psi(s)\|_{\mathcal{L}_2(\mathcal{H};H)}^2\,\mathrm{d}s\right] \quad \forall t \in \bar{J}.$$

In order to extend the space of admissible integrands, we take the closure of the vector space of all $\mathcal{L}(U;H)$-valued simple processes with respect to the following norm,

$$\|\Psi\|_{L^2_{\mathcal{H},T}(H)}^2 := \mathbb{E}\left[\int_0^T \|\Psi(s)\|_{\mathcal{L}_2(\mathcal{H};H)}^2\,\mathrm{d}s\right].$$

The resulting Banach space denoted by $L^2_{\mathcal{H},T}(H)$ is in fact a Hilbert space, namely the space of predictable processes Ψ taking values in the space of Hilbert–Schmidt operators $\mathcal{L}_2(\mathcal{H};H)$ such that the $L^2_{\mathcal{H},T}(H)$-norm defined above is finite, i.e.,

$$\begin{aligned} L^2_{\mathcal{H},T}(H) &:= \{\Psi\colon \Omega \times \bar{J} \to \mathcal{L}_2(\mathcal{H};H) : \Psi \text{ predictable}, \|\Psi\|_{L^2_{\mathcal{H},T}(H)} < \infty\} \\ &= L_2(\Omega \times \bar{J}, \mathcal{P}_{\bar{J}}, \mathbb{P} \times \mathrm{d}t; \mathcal{L}_2(\mathcal{H};H)). \end{aligned}$$

Here, $\mathcal{P}_{\bar{J}}$ denotes the σ-field of all predictable sets in $\Omega \times \bar{J}$, that is the smallest σ-field of subsets of $\Omega \times \bar{J}$ containing all sets of the form $A \times (s,t]$, where $0 \le s < t \le T$ and $A \in \mathcal{F}_s$, see [32, Thm. 8.7, Cor. 8.17]. Recall that the process Ψ is called predictable if Ψ is measurable with respect to $\mathcal{P}_{\bar{J}}$.

The construction described above yields a well-defined stochastic integral $\int_0^t \Psi(s)\,\mathrm{d}L(s) \in L_2(\Omega;H)$ for integrands $\Psi \in L^2_{\mathcal{H},T}(H)$ for all $t \in \bar{J}$. Moreover, this is by definition the largest class of integrands satisfying the Itô isometry.

For Paper I we furthermore need the notion of the weak stochastic integral. In order to introduce it, let Ψ be a stochastic process in the space of admissible integrands $L^2_{\mathcal{H},T}(H)$, and let $(X(t), t \in \bar{J})$ be a predictable H-valued stochastic process which is ($\mathbb{P}$-a.s.) continuous on $\bar{J}$. We then define the $\mathcal{L}_2(\mathcal{H};\mathbb{R})$-valued stochastic process Ψ_X, for $t \in \bar{J}$, by

$$\Psi_X(t)\colon z \mapsto (X(t), \Psi(t)z)_H \quad \forall z \in \mathcal{H}.$$

The predictability of X and of Ψ imply that Ψ_X is predictable. In addition, it can be seen from $\|\Psi\|_{L^2_{\mathcal{H},T}(H)} < \infty$ and from the ($\mathbb{P}$-a.s.) continuity of the trajectories of X that Ψ_X is an admissible integrand, $\Psi_X \in L^2_{\mathcal{H},T}(\mathbb{R})$. Therefore, we can define the real-valued weak stochastic integral $\int_0^t (X(s), \Psi(s)\,\mathrm{d}L(s))_H$ as the stochastic integral with respect to the integrand Ψ_X, i.e.,

$$\int_0^t (X(s), \Psi(s)\,\mathrm{d}L(s))_H := \int_0^t \Psi_X(s)\,\mathrm{d}L(s) \quad \forall t \in \bar{J} \quad \mathbb{P}\text{-a.s.}$$

We note that the Itô isometry for the original stochastic integral implies an isometry for the weak stochastic integral,

$$\mathbb{E}\left[\left|\int_0^t (X(s), \Psi(s)\,\mathrm{d}L(s))_H\right|^2\right] = \mathbb{E}\left[\int_0^t \|\Psi_X(s)\|_{\mathcal{L}_2(\mathcal{H};\mathbb{R})}^2\,\mathrm{d}s\right] \quad \forall t \in \bar{J},$$

which is particularly important for the analysis in Paper I.

4.4. The Itô formula. A vector-valued stochastic process, which can be written as a sum of a Bochner and a stochastic Itô integral, is often referred to as an Itô process in the literature [9]. More precisely, the H-valued stochastic process $X := (X(t), t \in \bar{J})$ defined by

$$X(t) := X_0 + \int_0^t Y(s)\,\mathrm{d}s + \int_0^t \Psi(s)\,\mathrm{d}W(s), \quad t \in \bar{J},$$

is called an Itô process generated by the quadruple (X_0, Y, Ψ, W), where

(i) X_0 is an $\mathcal{F}_0$-measurable H-valued random variable,
(ii) $Y = (Y(t), t \in \bar{J})$ is a predictable H-valued stochastic process which is ($\mathbb{P}$-a.s.) Bochner-integrable on $\bar{J}$,
(iii) $\Psi \in L^2_{\mathcal{H},T}(H)$ is an admissible stochastic integrand, see §4.3, and
(iv) $(W(t), t \in \bar{J})$ is a U-valued Wiener process with covariance operator $Q \in \mathcal{L}_1^+(U)$.

From this definition it is evident that every Itô process X is predictable. If, in addition, $X_0 \in L_2(\Omega; H)$ then X is square-integrable.

Suppose now that $f\colon \bar{J} \times H \to \mathbb{R}$ is a real-valued function and that the process $X = (X(t), t \in \bar{J})$ is an Itô process. Then, under appropriate assumptions on f and X, the real-valued stochastic process $(f(t, X(t)), t \in \bar{J})$ also allows for an explicit integral representation, i.e., it is again an Itô process. A version of this important result, the Itô formula, is formulated in the next theorem, cf. [6, 7]. It will play a central role in the weak error analysis of Paper IV.

THEOREM 4.1 (Itô formula). *Let $f\colon \bar{J} \times H \to \mathbb{R}$ be such that the real-valued functions f, $\partial_t f$, and the first two Fréchet derivatives $\partial_\phi f$, $\partial_{\phi\phi} f$ with respect to $\phi \in H$ with values in H and $\mathcal{L}(H)$, respectively, are all continuous on $\bar{J} \times H$. Assume that $X = (X(t), t \in \bar{J})$ is an Itô process generated by (X_0, Y, Ψ, W). Then it holds ($\mathbb{P}$-a.s.), for all $t \in \bar{J}$,*

$$\begin{aligned} f(t, X(t)) = f(0, X_0) &+ \int_0^t \partial_t f(s, X(s))\,\mathrm{d}s + \int_0^t (\partial_\phi f(s, X(s)), Y(s))_H\,\mathrm{d}s \\ &+ \frac{1}{2}\int_0^t \operatorname{tr}(\partial^2_{\phi\phi} f(s, X(s))\Psi(s)Q\Psi(s)^*)\,\mathrm{d}s \\ &+ \int_0^t (\partial_\phi f(s, X(s)), \Psi(s)\mathrm{d}W(s))_H, \end{aligned}$$

where the latter term is a weak stochastic integral, see §4.3. In other words, the real-valued stochastic process $(f(t, X(t)), t \in \bar{J})$ is an Itô process generated by the quadruple $(\widetilde{X}_0, \widetilde{Y}, \widetilde{\Psi}, W)$, where $\widetilde{X}_0 := f(0, X_0)$ and, for $t \in \bar{J}$,

$$\widetilde{Y}(t) := \partial_t f(t, X(t)) + (\partial_\phi f(t, X(t)), Y(t))_H + \frac{1}{2}\operatorname{tr}(\partial^2_{\phi\phi} f(t, X(t))\Psi(t)Q\Psi(t)^*),$$

$$\widetilde{\Psi}(t)\colon z \mapsto (\partial_\phi f(t, X(t)), \Psi(t)z)_H \quad \forall z \in \mathcal{H}.$$

4.5. Strong and mild solutions of stochastic partial differential equations. We consider equations of the form

$$\text{(SDE)} \qquad \mathrm{d}X(t) + AX(t)\mathrm{d}t = G[X(t)]\,\mathrm{d}L(t), \quad t \in \bar{J}, \qquad X(0) = X_0,$$

in the Hilbert space H, where

- $A\colon \mathcal{D}(A) \subset H \to H$ is a densely defined, self-adjoint, positive definite, possibly unbounded linear operator with a compact inverse as in §2.4;
- L is a U-valued Lévy process satisfying Assumptions (a)–(e) made in §4.2.2 and $\mathcal{H} = Q^{1/2}U$ denotes its RKHS, see §4.2.2;
- $G\colon H \to \mathcal{L}_2(\mathcal{H}; H)$ is an affine operator, i.e., there exist operators

$$G_1 \in \mathcal{L}(H; \mathcal{L}_2(\mathcal{H}; H)) \quad \text{and} \quad G_2 \in \mathcal{L}_2(\mathcal{H}; H)$$

  such that $G[\phi] = G_1[\phi] + G_2$ for all $\phi \in H$;
- X_0 is an $\mathcal{F}_0$-measurable, square-integrable H-valued random variable.

Equation (SDE) is said to be a stochastic differential equation, SDE for short. More precisely, it is called a stochastic *ordinary* differential equation (SODE) if $H = \mathbb{R}^n$ and $U = \mathbb{R}^m$ for finite dimensions $m, n \in \mathbb{N}$ and a stochastic *partial* differential equation (SPDE) if H is an infinite dimensional function space and A a differential operator. For vanishing G_1, (SDE) is said to have additive noise, otherwise it is called an SDE with multiplicative noise.

The purpose of this section is to make sense of the notion of solutions to this kind of equations. In fact, as for the deterministic initial value problem (IVP) in §3, there exist also different solution concepts for (SDE). In the following, we present their definitions and how they relate. For an introduction to stochastic ordinary differential equations and stochastic partial differential equations the reader is referred to [20, 22, 30] and to [10, 23, 32, 33], respectively.

An H-valued predictable process $X = (X(t), t \in \bar{J})$ taking values in $\mathcal{D}(A)$ $\mathcal{P}_{\bar{J}}$-a.s. is called a strong solution to (SDE) if

$$\int_J \big(\|X(s)\|_H + \|AX(s)\|_H + \|G(X(s))\|^2_{\mathcal{L}_2(\mathcal{H};H)}\big)\,\mathrm{d}s < \infty \quad \mathbb{P}\text{-a.s.},$$

and the following integral equation is satisfied for all $t \in \bar{J}$:

$$X(t) = X_0 - \int_0^t AX(s)\,\mathrm{d}s + \int_0^t G(X(s))\,\mathrm{d}L(s) \quad \mathbb{P}\text{-a.s.}$$

We emphasize the close relation between the notions of a strong solution to the deterministic initial value problem (IVP) on the one hand, and of a strong solution to the stochastic differential equation (SDE) on the other hand.

Since we assume that the operators A and G are linear and affine, respectively, in the case of an SODE with $H = \mathbb{R}^n$ and $U = \mathbb{R}^m$ we have

$$A \in \mathbb{R}^{n\times n}, \quad G_1 \in \mathcal{L}(\mathbb{R}^n; \mathbb{R}^{n\times m}), \quad \text{and} \quad G_2 \in \mathbb{R}^{n\times m}.$$

It is well-known [30, Thm. 5.2.1] that the resulting SODE admits a strong solution when driven by $\mathbb{R}^m$-valued Wiener white noise, cf. §4.2.3. Moreover,

this solution is unique up to modification, i.e., if X_1 and X_2 are two strong solutions, then

$$\mathbb{P}(X_1(t) = X_2(t)) = 1 \quad \forall t \in \bar{J}.$$

Under the additional assumption that the mapping $t \mapsto X(t)$ is ($\mathbb{P}$-a.s.) continuous (t-continuity), the solution is pathwise unique in the sense of [22], i.e.,

$$\mathbb{P}(X_1(t) = X_2(t) \quad \forall t \in \bar{J}) = 1$$

for any two t-continuous strong solutions X_1 and X_2.

As a further illustration of the concept of strong solutions we take as explicit examples the real-valued ($m = n = 1$) model SODEs considered in Paper II: with additive noise

$$(\text{SODE}_+) \quad \mathrm{d}X(t) + \lambda X(t)\,\mathrm{d}t = \mu\,\mathrm{d}W(t), \quad t \in \bar{J}, \qquad X(0) = X_0,$$

and with multiplicative noise

$$(\text{SODE}_*) \quad \mathrm{d}X(t) + \lambda X(t)\,\mathrm{d}t = \rho X(t)\,\mathrm{d}W(t), \quad t \in \bar{J}, \qquad X(0) = X_0,$$

for an initial value $X_0 \in L_2(\Omega;\mathbb{R})$ and constant parameters $\lambda, \mu, \rho > 0$.

As stated above, there exist strong solutions to these SODEs. Indeed, for additive noise the Ornstein–Uhlenbeck process defined by

$$X(t) := e^{-\lambda t}X_0 + \mu \int_0^t e^{-\lambda(t-s)}\,\mathrm{d}W(s), \quad t \in \bar{J},$$

and in the multiplicative case the geometric Brownian motion given by

$$X(t) := X_0 e^{-(\lambda+\rho^2/2)t+\rho W(t)}, \quad t \in \bar{J},$$

satisfy the conditions of being strong solutions to the model SODEs above. We note that in both cases the integral equation can be verified by an application of the Itô formula from §4.4—to the process $(e^{\lambda t}X(t), t \in \bar{J})$ in the additive case and to the geometric Brownian motion X itself in the multiplicative case. Moreover, these solutions are the unique t-continuous strong solutions.

Due to the availability of existence and uniqueness results for strong solutions to SODEs of the kind above and, more generally, with global Lipschitz coefficients (see [22, 30] for Wiener noise and [1] for Lévy noise), this solution concept is usually sufficient in the finite-dimensional case when $\dim(H) = n < \infty$ and $\dim(U) = m < \infty$.

However, as for deterministic initial value problems, it is often unsatisfactory when considering equations in infinite dimensions, since—depending on the operator A—the condition "X takes values in $\mathcal{D}(A)$ $\mathbb{P}$-a.s." may be too restrictive. Recall that for the deterministic problem (IVP) existence of a strong solution is only ensured, if the source term f is Hölder continuous. In the terminology of the deterministic framework, the noise term generated by the Lévy process L in (SDE) takes the role of the source term. Since a Lévy process is in general not pathwise differentiable (e.g., in the case when L is a Wiener process), it is usually irregular with respect to t. For this reason, strong solutions rarely exist and a less restrictive solution concept is needed.

As in the deterministic case, the semigroup $(S(t), t \geq 0)$ generated by $-A$ can be used to define mild solutions of (SDE), see [32, Def. 9.5].

Let $X = (X(t), t \in \bar{J})$ be an H-valued predictable process with

$$\sup_{t \in \bar{J}} \|X(t)\|_{L^2(\Omega;H)} < \infty.$$

Then X is said to be a mild solution to (SDE) if

$$X(t) = S(t)X_0 + \int_0^t S(t-s)G(X(s))\,\mathrm{d}L(s), \quad \forall t \in \bar{J}, \quad \mathbb{P}\text{-a.s.}$$

In contrast to strong solutions, one knows [32, Thms. 9.15 & 9.29] that under the assumptions on A and G made above, there exists a mild solution to (SDE), which is unique up to modification. Moreover, the mild solution has a càdlàg modification, which is pathwise unique. Whenever a strong solution to (SDE) exists, it coincides with the mild solution.

5. Numerical methods for variational problems

The aim of this section is to introduce two basic numerical approaches to solve partial differential equations approximately: Galerkin methods and Petrov–Galerkin approximations. More precisely, we first introduce conforming Galerkin discretizations for strongly elliptic problems on Hilbert spaces and some of their approximation properties in §5.1. These are needed for the numerical schemes discussed in the second part of the thesis, Papers III–V. Note that, in general, Galerkin methods are based on variational formulations with coinciding trial–test spaces. This is not the case for the variational problems derived in the first part of the thesis, Papers I and II. In fact, the trial and test spaces there are the projective and injective tensor product spaces $\mathcal{X}_\pi$ and $\mathcal{Y}_\epsilon$, respectively, where $\mathcal{X}$ and $\mathcal{Y}$ are the vector-valued function spaces introduced in §2.5. Recall from §1.2 that these tensor product spaces are non-reflexive Banach spaces. For this reason, we discuss Petrov–Galerkin discretizations of linear variational problems posed on general normed vector spaces in §5.2. For introductions to (Petrov–)Galerkin methods we refer to [5, 16, 39].

5.1. Galerkin approximations. Let us assume that $A\colon \mathcal{D}(A) \subset H \to H$ is a densely defined, self-adjoint, positive definite linear operator with a compact inverse as in §2.4. Recall from §2.4 the Gelfand triple

$$V \hookrightarrow H \cong H^* \hookrightarrow V^*, \quad \text{where} \quad V := \mathcal{D}(A^{1/2}),$$

as well as the eigenvalues $\{\lambda_j\}_{j\in\mathbb{N}}$ of A, which are arranged in nondecreasing order with corresponding eigenvectors $\{e_j\}_{j\in\mathbb{N}}$, which are orthonormal in H. As pointed out in §2.4, the operator A extends continuously to a bounded linear operator $A \in \mathcal{L}(V; V^*)$. Furthermore, $\langle A\psi, \psi\rangle \geq \lambda_1 \|\psi\|_H^2$ for all $\psi \in V$. Thus, A induces a continuous and strongly elliptic (or coercive) bilinear form

$$a\colon V \times V \to \mathbb{R}, \qquad a(\phi, \psi) := \langle A\phi, \psi\rangle,$$

which we have used in §2.4 to define the inner product on V.

The Riesz representation theorem ensures that, given $f \in V^*$, there exists a unique element in V solving the variational problem

$$\text{Find} \quad \phi \in V \quad \text{s.t.} \quad a(\phi, \psi) = \langle f, \psi \rangle \quad \forall \psi \in V.$$

We are interested in approximations of this solution $\phi \in V$.

For this purpose, we assume that $(V_h)_{h\in(0,1)} \subset V$ is a family of subspaces of V with finite dimensions $N_h := \dim(V_h) < \infty$. The Galerkin discretization of the operator A is then defined by

$$A_h \colon V_h \to V_h, \quad (A_h\phi_h, \psi_h)_H = \langle A\phi_h, \psi_h \rangle = a(\phi_h, \psi_h) \quad \forall \phi_h, \psi_h \in V_h,$$

and the Galerkin approximation of the above variational solution $\phi \in V$ is the unique element $\phi_h \in V_h$ satisfying

$$\text{Find} \quad \phi_h \in V_h \quad \text{s.t.} \quad a(\phi_h, \psi_h) = \langle f, \psi_h \rangle \quad \forall \psi_h \in V_h.$$

This approximation is often referred to as the Ritz projection of $\phi = A^{-1}f$ onto V_h. If $f \in H$, it can be expressed in terms of the discretized operator A_h and the H-orthogonal projection Π_h onto V_h by $\phi_h = A_h^{-1}\Pi_h f$. If $f \in V^* \setminus H$, one has to take the unique continuous linear extension $\Pi_h \colon V^* \to V_h$ instead.

For the error analysis in Papers III–V, it is particularly important that the eigenvalues $\{\lambda_{j,h}\}_{j=1}^{N_h}$ (again assumed to be in nondecreasing order) as well as the corresponding H-orthonormal eigenvectors $\{e_{j,h}\}_{j=1}^{N_h}$ of the discretized operator A_h approximate the first N_h eigenvalues and eigenvectors of A sufficiently well. For this reason, we state the following result on the approximation properties of Galerkin (finite element) approximations for elliptic problems induced by linear differential operators in $\mathbb{R}^d$, which in this generality can be found, e.g., in [39, Thms. 6.1 & 6.2].

THEOREM 5.1. *Let $D \subset \mathbb{R}^d$ be a bounded, convex, polygonal domain and $A \colon \mathcal{D}(A) \subset L_2(D) \to L_2(D)$ be a (strongly) elliptic linear differential operator of order $2m \in \mathbb{N}$, i.e., there exists a constant $c > 0$ such that*

$$\langle Av, v \rangle \geq c \, \|v\|^2_{H^m(D)} \quad \forall v \in V = \mathcal{D}(A^{1/2}).$$

Assume that $(V_h)_{h\in(0,1)}$ is a quasi-uniform family of admissible finite element spaces $V_h \subset V$ of polynomial degree $p \in \mathbb{N}$. Then there exist constants $h_0 \in (0,1)$ and $C_1, C_2 > 0$, independent of h, such that

$$\lambda_j \leq \lambda_{j,h} \leq \lambda_j + C_1 h^{2(p+1-m)} \lambda_j^{\frac{p+1}{m}},$$
$$\|e_j - e_{j,h}\|_{L_2(D)} \leq C_2 h^{\min\{p+1, 2(p+1-m)\}} \lambda_j^{\frac{p+1}{2m}},$$

for all $h \in (0, h_0)$ and $j \in \{1, \ldots, N_h\}$.

In particular, if $m = 1$ and the family $(V_h)_{h\in(0,1)}$ of finite element spaces is quasi-uniform with continuous, piecewise linear basis functions, it holds

$$\lambda_j \leq \lambda_{j,h} \leq \lambda_j + C_1 h^2 \lambda_j^2 \quad \text{and} \quad \|e_j - e_{j,h}\|_{L_2(D)} \leq C_2 h^2 \lambda_j, \quad 1 \leq j \leq N_h.$$

5.2. Petrov–Galerkin approximations. We now proceed with collecting results for Petrov–Galerkin discretizations of the generic linear variational problem

$$\text{Find} \quad u \in \mathcal{U} \quad \text{s.t.} \quad \mathscr{B}(u, v) = \ell(v) \quad \forall v \in \mathcal{V},$$

posed on normed vector spaces $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$ and $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$, with a continuous linear right-hand side $\ell \in \mathcal{V}'$, and a bilinear form $\mathscr{B}\colon \mathcal{U} \times \mathcal{V} \to \mathbb{R}$. We assume that $\mathscr{B}$ is continuous on $\mathcal{U} \times \mathcal{V}$, so that the operator induced by $\mathscr{B}$ (again denoted by $\mathscr{B}$) is linear and bounded, i.e., $\mathscr{B} \in \mathcal{L}(\mathcal{U}; \mathcal{V}')$. The generality of considering normed vector spaces as trial–test spaces instead of Hilbert spaces will allow us to address the variational problem satisfied by the second moment of the solution to (SDE) with multiplicative noise.

We assume that $\mathcal{U}_h \times \mathcal{V}_h \subset \mathcal{U} \times \mathcal{V}$ is a fixed pair of non-trivial subspaces with equal finite dimension $\dim \mathcal{U}_h = \dim \mathcal{V}_h < \infty$. We aim at approximating the solution $u \in \mathcal{U}$ of the variational problem above by a function $u_h \in \mathcal{U}_h$ and quantifying the error $\|u - u_h\|_{\mathcal{U}}$. For this purpose, suppose that the operator $\bar{\mathscr{B}}\colon \mathcal{U} \to \mathcal{V}'$ is an approximation of $\mathscr{B}$, which again is continuous. We introduce the notation

$$\|\ell\|_{\mathcal{V}_h'} := \sup_{v \in S(\mathcal{V}_h)} |\ell(v)|$$

for functionals ℓ which are defined on $\mathcal{V}_h$, and assume that the approximation $\bar{\mathscr{B}}$ admits a constant $\bar{\gamma}_h > 0$ such that

$$\|\bar{\mathscr{B}} w_h\|_{\mathcal{V}_h'} \geq \bar{\gamma}_h \|w_h\|_{\mathcal{U}} \quad \forall w_h \in \mathcal{U}_h.$$

In other words, the corresponding bilinear form $\bar{\mathscr{B}}$ satisfies the discrete inf-sup condition

$$\inf_{w_h \in S(\mathcal{U}_h)} \sup_{v_h \in S(\mathcal{V}_h)} \bar{\mathscr{B}}(w_h, v_h) \geq \bar{\gamma}_h > 0.$$

We then define the approximate solution u_h as the solution of the discrete variational problem:

$$\text{Find} \quad u_h \in \mathcal{U}_h \quad \text{s.t.} \quad \bar{\mathscr{B}}(u_h, v_h) = \ell(v_h) \quad \forall v_h \in \mathcal{V}_h.$$

Recall that we only assume that the discrete trial–test spaces $\mathcal{U}_h$ and $\mathcal{V}_h$ are of the same dimension, and that they may differ. In this case when $\mathcal{U}_h \neq \mathcal{V}_h$, the discrete variational problem is said to be a Petrov–Galerkin discretization and its solution $u_h \in \mathcal{U}_h$ is called a Petrov–Galerkin approximation.

The following proposition ensures existence and uniqueness of the Petrov–Galerkin approximation. In addition, it quantifies the error $\|u - u_h\|_{\mathcal{U}}$, which is of importance for the convergence analysis of the Petrov–Galerkin discretizations discussed in Paper II.

Proposition 5.2. *Fix $u \in \mathcal{U}$. Under the above assumptions there exists a unique $u_h \in \mathcal{U}_h$ such that*

$$\bar{\mathscr{B}}(u_h, v_h) = \mathscr{B}(u, v_h) \quad \forall v_h \in \mathcal{V}_h.$$

The mapping $u \mapsto u_h$ is linear with $\|u_h\|_{\mathcal{U}} \leq \bar{\gamma}_h^{-1}\|\mathscr{B}u\|_{\mathcal{V}_h'}$ and satisfies the quasi-optimality estimate

$$\|u-u_h\|_{\mathcal{U}} \leq \big(1+\bar{\gamma}_h^{-1}\|\bar{\mathscr{B}}\|_{\mathcal{L}(\mathcal{U};\mathcal{V}')}\big) \inf_{w_h\in\mathcal{U}_h} \|u-w_h\|_{\mathcal{U}} + \bar{\gamma}_h^{-1}\|(\mathscr{B}-\bar{\mathscr{B}})u\|_{\mathcal{V}_h'}.$$

Proof. Injectivity of the operator $\bar{\mathscr{B}}$ on $\mathcal{U}_h$ follows from the discrete inf-sup condition imposed above. Since $\dim\mathcal{U}_h = \dim\mathcal{V}_h$, the operator $\bar{\mathscr{B}}\colon \mathcal{U}_h \to \mathcal{V}_h'$ is an isomorphism and existence and uniqueness of u_h follow.

In order to derive the quasi-optimality estimate, fix $w_h \in \mathcal{U}_h$. By the triangle inequality, $\|u-u_h\|_{\mathcal{U}} \leq \|u-w_h\|_{\mathcal{U}} + \|w_h-u_h\|_{\mathcal{U}}$. Due to the discrete inf-sup condition we can bound the second term as follows:

$$\begin{aligned}\bar{\gamma}_h\|w_h-u_h\|_{\mathcal{U}} &\leq \sup_{v_h\in\mathcal{V}_h} \bar{\mathscr{B}}(w_h-u_h,v_h) = \sup_{v_h\in\mathcal{V}_h} [\bar{\mathscr{B}}(w_h,v_h)-\mathscr{B}(u,v_h)]\\ &\leq \sup_{v_h\in\mathcal{V}_h} \bar{\mathscr{B}}(w_h-u,v_h) + \sup_{v_h\in\mathcal{V}_h} [\bar{\mathscr{B}}(u,v_h)-\mathscr{B}(u,v_h)]\\ &\leq \|\bar{\mathscr{B}}\|_{\mathcal{L}(\mathcal{U};\mathcal{V}')}\|u-w_h\|_{\mathcal{U}} + \|(\mathscr{B}-\bar{\mathscr{B}})u\|_{\mathcal{V}_h'}.\end{aligned}$$

Therefore, for arbitrary $w_h \in \mathcal{U}_h$ we may bound the error $\|u-u_h\|_{\mathcal{U}}$ by

$$\|u-u_h\|_{\mathcal{U}} \leq \big(1+\bar{\gamma}_h^{-1}\|\bar{\mathscr{B}}\|_{\mathcal{L}(\mathcal{U};\mathcal{V}')}\big)\|u-w_h\|_{\mathcal{U}} + \bar{\gamma}_h^{-1}\|(\mathscr{B}-\bar{\mathscr{B}})u\|_{\mathcal{V}_h'},$$

and taking the infimum with respect to $w_h \in \mathcal{U}_h$ proves the assertion. □

6. Summary of the appended papers

6.1. Papers I & II: Variational methods for moments of solutions to SPDEs. In Paper I we pursue the study of [24], where the second moment and the covariance of the mild solution to a parabolic SPDE driven by additive Wiener noise have been described as solutions to well-posed space-time variational problems posed on Hilbert tensor products of Bochner spaces. More precisely, in [24] parabolic SPDEs of the form (SDE) have been considered, where the noise term is driven by an H-valued Wiener process, see §4.2.3, and, for all $\phi \in H$, the operator $G[\phi]$ is the identity on H, i.e., $G_1[\,\cdot\,] = 0$, $G_2 = I$. In this case, the state space of the Wiener noise and of the mild solution X to (SDE) coincide (both are H).

With the notation and definitions of the vector-valued function spaces $\mathcal{X}$ and $\mathcal{Y}$, as well as the bilinear form $b\colon \mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ from §2.5 and §3.3, the mean $m = \mathbb{E}[X]$ of the mild solution X satisfies the deterministic variational problem:

(VPm) Find $m \in \mathcal{X}$ s.t. $b(m,v) = (\mathbb{E}[X_0], v(0))_H \quad \forall v \in \mathcal{Y}$.

Well-posedness of this problem was subject of the analysis of (VP3) in §3.3.

In [24] the tensorized bilinear form $B\colon \mathcal{X}_2\times\mathcal{Y}_2\to\mathbb{R}$ has been introduced on the Hilbert tensor product spaces $\mathcal{X}_2$ and $\mathcal{Y}_2$ (see §1.2) as $B := b\otimes b$, or explicitly as

$$B(w,v) := \int_J\int_J \langle w(s,t), (-\partial_s + A)\otimes(-\partial_t + A)v(s,t)\rangle \,\mathrm{d}s\,\mathrm{d}t.$$

Here, $\langle \cdot, \cdot \rangle$ denotes the duality pairing between V_2 and V_2^*. It has been proven that the second moment $M = \mathbb{E}[X \otimes X]$, see §4.2, of the mild solution X satisfies the following deterministic variational problem:

$$\text{(VPM}_+\text{)} \qquad \text{Find} \quad M \in \mathcal{X}_2 \quad \text{s.t.} \quad B(M, v) = \ell_+(v) \quad \forall v \in \mathcal{Y}_2,$$

where $\ell_+(v) := (\mathbb{E}[X_0 \otimes X_0], v(0,0))_2 + \delta_q(v)$. The functional $\delta_q \colon \mathcal{Y}_2 \to \mathbb{R}$ is defined by (recall from §1.2 that $(\cdot, \cdot)_2$ abbreviates the inner product on H_2, and from §1.3 and §4.1.1 that there exists a unique kernel $q \in H_\pi^+$ corresponding to the covariance operator $Q \in \mathcal{L}_+^1(H)$ of the Wiener process)

$$\delta_q(v) := \int_J (q, v(t,t))_2 \, \mathrm{d}t \quad \forall v \in \mathcal{Y}_2.$$

It has been shown that under the assumption $\operatorname{tr}(AQ) < \infty$, the functional δ_q and, thus, the right-hand side ℓ_+ of the variational problem (VPM$_+$) are elements of the dual $(\mathcal{Y}_2)'$. As remarked in §3.3, the operator $b \in \mathcal{L}(\mathcal{X}; \mathcal{Y}')$ is an isomorphism, so that $B = b \otimes b \in \mathcal{L}(\mathcal{X}_2; \mathcal{Y}_2')$ inherits this property and the variational problem (VPM$_+$) for the second moment is well-posed.

In Paper I we prove that also in the case of multiplicative Lévy noise the second moment and the covariance of the square-integrable mild solution to (SDE) satisfy deterministic space-time variational problems posed on tensor products of vector-valued function spaces. In contrast to the case of additive Wiener noise considered in [24], the pair of trial–test spaces is not given by Hilbert tensor product spaces, but by projective–injective tensor product spaces. In addition, the resulting bilinear form in the variational problem involves a non-separable form on these tensor spaces. Therefore, well-posedness does not readily follow from the isomorphism property of b, and a careful analysis is needed to prove existence and uniqueness of a solution to the derived variational problem.

To specify this non-separable form, we first introduce, besides the vector-valued function spaces $\mathcal{X}$ and $\mathcal{Y}$, the Bochner space $\mathcal{Z} := L_2(J; H)$. We then define the bilinear form Δ, referred to as the trace product, for $w \otimes \widetilde{w} \in \mathcal{Z} \otimes \mathcal{Z}$ and $v \otimes \widetilde{v} \in \mathcal{Y} \otimes \mathcal{Y}$ by

$$\Delta(w \otimes \widetilde{w}, v \otimes \widetilde{v}) := \int_J (w(t), v(t))_H (\widetilde{w}(t), \widetilde{v}(t))_H \, \mathrm{d}t,$$

extending this definition by bilinearity to the algebraic tensor product spaces $\mathcal{Z} \otimes \mathcal{Z}$ and $\mathcal{Y} \otimes \mathcal{Y}$. We prove that the trace product Δ admits a unique continuous extension to a bilinear form $\Delta \colon \mathcal{Z}_\pi \times \mathcal{Y}_\epsilon \to \mathbb{R}$, and that the induced operator $\Delta \in \mathcal{L}(\mathcal{Z}_\pi; \mathcal{Y}_\epsilon')$ satisfies $\|\Delta\|_{\mathcal{L}(\mathcal{Z}_\pi; \mathcal{Y}_\epsilon')} \leq 1$, where $\mathcal{Y}_\epsilon'$ denotes the dual of the injective tensor product space $\mathcal{Y}_\epsilon$ ($\mathcal{Y}$ being equipped with $\|\cdot\|_{\mathcal{Y}}$, see §2.5).

The covariance operator $Q \in \mathcal{L}_1^+(U)$ of the Lévy noise L in (SDE) and the linear part $G_1 \in \mathcal{L}(H; \mathcal{L}_2(\mathcal{H}; H))$ of the multiplicative term interacting with L enter the deterministic equation for the second moment through an operator denoted by $\mathcal{G}_1$. We let $q \in U_\pi^+$ be the kernel corresponding to Q, i.e., $q = \sum_{j \in \mathbb{N}} \gamma_j (e_j \otimes e_j)$, where $\{\gamma_j\}_{j \in \mathbb{N}}$ are the nonnegative eigenvalues of Q with U-orthonormal eigenvectors $\{e_j\}_{j \in \mathbb{N}}$. For $w, \widetilde{w} \in \mathcal{X}$, we then define the

operator $\mathcal{G}_1$ via

$$\mathcal{G}_1[w \otimes \widetilde{w}] := (G_1[w] \otimes G_1[\widetilde{w}])q = \sum_{j\in\mathbb{N}} \gamma_j (G_1[w]e_j \otimes G_1[\widetilde{w}]e_j),$$

and prove existence and uniqueness of a continuous extension $\mathcal{G}_1 \in \mathcal{L}(\mathcal{X}_\pi; \mathcal{Z}_\pi)$ with $\|\mathcal{G}_1\|_{\mathcal{L}(\mathcal{X}_\pi;\mathcal{Z}_\pi)} \le \|G_1\|^2_{\mathcal{L}(V;\mathcal{L}_2(\mathcal{H};H))}$.

Having defined the trace product Δ and the operator $\mathcal{G}_1$, we introduce the continuous bilinear form

$$\mathcal{B}\colon \mathcal{X}_\pi \times \mathcal{Y}_\epsilon \to \mathbb{R}, \qquad \mathcal{B}(w,v) := B(w,v) - \Delta(\mathcal{G}_1[w], v).$$

We show that the second moment of the solution X to (SDE) with multiplicative Lévy noise satisfies the deterministic variational problem:

$$\text{(VPM}_*\text{)} \qquad \text{Find} \quad M \in \mathcal{X}_\pi \quad \text{s.t.} \quad \mathcal{B}(M,v) = \ell_*(v) \quad \forall v \in \mathcal{Y}_\epsilon$$

with the right-hand side (recall $m \in \mathcal{X}$ denotes the mean of X)

$$\begin{aligned}\ell_*(v) := (\mathbb{E}[X_0 \otimes X_0], v(0,0))_2 &+ \Delta((G_1[m] \otimes G_2)q, v) \\ &+ \Delta((G_2 \otimes G_1[m])q, v) + \Delta((G_2 \otimes G_2)q, v).\end{aligned}$$

Well-posedness of this problem is proven under an appropriate assumption on the operator G_1. More precisely, the lower bound $\gamma \ge 1$ for the inf-sup constant of the bilinear form $b\colon (\mathcal{X}, \|\cdot\|_{\mathcal{X}}) \times (\mathcal{Y}, \|\cdot\|_{\mathcal{Y}}) \to \mathbb{R}$ discussed in §3.3 along with the observation that $\mathcal{Y}'_\epsilon = (\mathcal{Y}')_\pi$, see [35, Thms. 2.5 & 5.13], implies that the bilinear form $B\colon \mathcal{X}_\pi \times \mathcal{Y}_\epsilon \to \mathbb{R}$ satisfies the following inf-sup condition

$$\inf_{w\in S(\mathcal{X}_\pi)} \sup_{v\in S(\mathcal{Y}_\epsilon)} B(w,v) \ge 1.$$

Owing to the bounds $\|\Delta\|_{\mathcal{L}(\mathcal{Z}_\pi;\mathcal{Y}'_\epsilon)} \le 1$ and $\|\mathcal{G}_1\|_{\mathcal{L}(\mathcal{X}_\pi;\mathcal{Z}_\pi)} \le \|G_1\|^2_{\mathcal{L}(V;\mathcal{L}_2(\mathcal{H};H))}$ mentioned above, we thereby find

$$\inf_{w\in S(\mathcal{X}_\pi)} \sup_{v\in S(\mathcal{Y}_\epsilon)} \mathcal{B}(w,v) \ge 1 - \|G_1\|^2_{\mathcal{L}(V;\mathcal{L}_2(\mathcal{H};H))}.$$

Thus, the operator $\mathcal{B}\colon \mathcal{X}_\pi \to \mathcal{Y}'_\epsilon$ induced by the bilinear form $\mathcal{B}$ in (VPM$_*$) is injective if

$$\text{(G1)} \qquad \|G_1\|_{\mathcal{L}(V;\mathcal{L}_2(\mathcal{H};H))} < 1.$$

Finally, a well-posed variational problem for the covariance function of the mild solution X to (SDE), again posed on $\mathcal{X}_\pi$ and $\mathcal{Y}_\epsilon$ as trial–test spaces, follows from the results for the mean and the second moment.

The purpose of Paper II is to introduce numerical methods for the variational problems derived in Paper I, and to discuss their stability and convergence. To this end, we first consider the canonical examples of stochastic ODEs (i.e., $V = H = \mathbb{R}$) with additive or multiplicative Wiener noise, namely the Ornstein–Uhlenbeck process (SODE$_+$) and the geometric Brownian motion (SODE$_*$) from §4.5. As pointed out in [24] and Paper I, the equations for the second moment and the covariance are posed on tensor products of function spaces. In the additive case (VPM$_+$) they can be taken as the Hilbert tensor product spaces $\mathcal{X}_2$ and $\mathcal{Y}_2$ [24], and well-posedness is readily seen, since

$B\colon (\mathcal{X}_2, \|\cdot\|_{\mathcal{X}_2}) \to (\mathcal{Y}_2, |||\cdot|||_{\mathcal{Y}_2})$ is an isometric isomorphism. In the multiplicative case (VPM$_*$), however, the pair $\mathcal{X}_\pi \times \mathcal{Y}_\epsilon$ of projective–injective tensor product spaces as trial–test spaces is required and well-posedness, proven in Paper I under Assumption (G1), is not an immediate consequence anymore due to the presence of the trace product Δ in the bilinear form $\mathcal{B}$. For the considered example (SODE$_*$) with multiplicative noise, we prove well-posedness of the deterministic variational problem for the second moment for all $\lambda, \rho > 0$, i.e., beyond the smallness assumption (G1) on $G_1[\,\cdot\,] = \rho\,\cdot$ made in Paper I.

Afterwards, we focus on deriving numerical approximations for the mean and the second moment in the scalar (ODE) case. We start by discussing different Petrov–Galerkin discretizations for the variational problem (VPm) satisfied by the first moment. From these, Petrov–Galerkin discretizations based on tensor product piecewise polynomials are constructed, which are then applied to the variational problems (VPM$_+$) and (VPM$_*$) for the second moments. We discuss briefly the discretization of (VPM$_+$) and focus on the more sophisticated multiplicative case (VPM$_*$). We prove stability of the discrete solution with respect to the projective norm on $\mathcal{X}_\pi$ and conclude that a discrete inf-sup condition is satisfied. Therefore, Proposition 5.2 is applicable, which yields a quasi-optimality estimate. From this, convergence of the discrete solution to the exact solution in $\mathcal{X}_\pi$ is derived.

Finally, we consider the vector-valued (PDE) situation when $\dim(H) = \infty$, focussing on the more sophisticated situation of multiplicative noise. The transition from convolutions of real-valued functions to semigroup theory on tensor product spaces allows us to prove well-posedness of the deterministic second moment equation also in the vector-valued case even beyond the smallness assumption (G1) on the multiplicative term made in Paper I. In fact, the analysis of Paper II reveals that the natural condition on G_1 for well-posedness of (VPM$_*$) is that $\|\mathcal{G}_1\|_{\mathcal{L}(V_\pi)} < \infty$.

We then propose (tensorized) space-time discretizations of (VPM$_*$) based on finite element subspaces $V_h \subset V$ in space, and the temporal discretizations which we have introduced and discussed for the scalar case. We prove that, if the functional $\ell \in \mathcal{Y}_\epsilon'$ on the right-hand side of (VPM$_*$) is symmetric and positive semi-definite, these approximations are stable in the sense that their projective norm on $\mathcal{X}_\pi$ can be bounded in terms of the dual norm $\|\ell\|_{\mathcal{Y}_\epsilon'}$. Note that the right-hand side $\ell_* \in \mathcal{Y}_\epsilon'$ in (VPM$_*$) is symmetric and positive semi-definite. From this result, we again deduce discrete inf-sup and quasi-optimality estimates.

In both (scalar and vector-valued) parts of Paper II, numerical experiments verify the theoretical outcomes, showing linear convergence with respect to the discretization parameter in time.

6.2. Papers III–V: Fractional elliptic SPDEs with spatial white noise. With the notions from §4.1, a real-valued Gaussian random field u defined on a spatial domain $D \subset \mathbb{R}^d$ can be seen as an $L_2(D)$-valued Gaussian random variable, which is ($\mathbb{P}$-a.s.) continuous on the closure $\bar{D}$. For this reason,

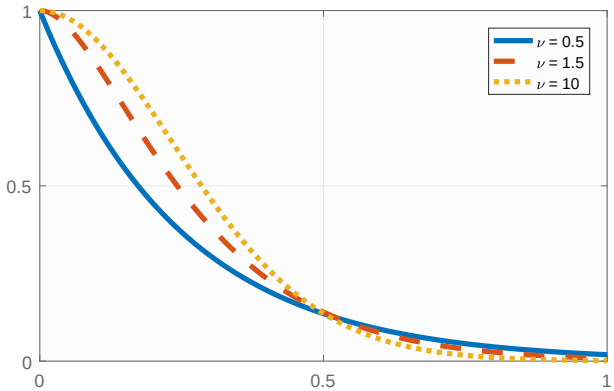

FIGURE 1. The function C_0 defining the Matérn covariances for $\nu = 0.5$ (solid line), $\nu = 1.5$ (dashed line), and $\nu = 10$ (dotted line), with $\sigma = 1$ and $\kappa = 2\sqrt{8\nu}$.

the pointwise covariances,

$$C(\mathbf{x}_1, \mathbf{x}_2) := \mathrm{Cov}(u(\mathbf{x}_1), u(\mathbf{x}_2)), \quad \mathbf{x}_1, \mathbf{x}_2 \in D,$$

are well-defined, and we call the resulting function $C\colon D \times D \to \mathbb{R}$ the covariance function of u. If this covariance function is, for $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$, given by

$$C(\mathbf{x}_1, \mathbf{x}_2) = C_0(\|\mathbf{x}_1 - \mathbf{x}_2\|), \qquad C_0(x) := \frac{2^{1-\nu}\sigma^2}{\Gamma(\nu)}(\kappa x)^\nu K_\nu(\kappa x), \quad x \in \mathbb{R}_{\geq 0},$$

for some constants $\nu, \sigma, \kappa > 0$, then u is called a Gaussian Matérn field, named after the Swedish forestry statistician Bertil Matérn [27]. Here, $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$ and Γ, K_ν denote the gamma function and the modified Bessel function of the second kind, respectively. A key feature of zero-mean Gaussian random fields with Matérn covariance functions is that they are characterized by the three parameters $\nu, \sigma, \kappa > 0$, which determine the smoothness, the variance, and the practical correlation range of the field. Because of this practicality, these fields are often used for modeling in spatial statistics, see [38]. In Figure 1 the function $C_0\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, which defines the Matérn covariances, is displayed for $\nu \in \{0.5, 1.5, 10\}$. In all three cases, the variance is normalized to $\sigma = 1$ and $\kappa = 2\sqrt{8\nu}$ is chosen such that the practical correlation range (i.e., the distance when the correlation is approximately 0.1) equals 0.5. Note that, for $\nu = 0.5$, the function C_0 is given by $C_0(x) = \sigma^2 e^{-\kappa x}$, i.e., this is the case of an exponential covariance function.

If $D = \mathbb{R}^d$ is the whole d-dimensional space, it is a well-known result [41, 42] that, for $\tau > 0$ and $\beta > d/4$, the Gaussian solution u to the fractional order

elliptic SPDE

$$(\mathrm{SDE}_\beta) \qquad (\kappa^2 - \Delta)^\beta (\tau u(\mathbf{x})) = \mathcal{W}(\mathbf{x}), \quad \mathbf{x} \in D,$$

has a Matérn covariance function with

$$\nu = 2\beta - d/2 \quad \text{and} \quad \sigma^2 = \frac{\kappa^{d-4\beta}\Gamma(2\beta - d/2)}{\tau^2(4\pi)^{d/2}\Gamma(2\beta)}.$$

Here, $\mathcal{W}$ is Gaussian white noise on $\mathbb{R}^d$ (i.e., on $L_2(\mathbb{R}^d)$, see §4.1.3), and we use Δ to denote the Laplacian, since the trace product does not occur in the second part of the thesis. In particular, the fractional order $\beta > d/4$ in (SDE_β) defines the differentiability of u.

This relation between Gaussian Matérn fields and SPDEs has been used in [25] to construct Markov random field approximations of Matérn fields on bounded domains $D \subsetneq \mathbb{R}^d$ if $2\beta \in \mathbb{N}$ by augmenting the differential operator $\kappa^2 - \Delta$ with Neumann boundary conditions on ∂D and using a finite element method for the numerical solution of the resulting problem. Due to the computational benefits, e.g., for inference, this approach has become widely used in spatial statistics. However, since the smoothness of the process, controlled by β, is particularly important for spatial predictions, the restriction $2\beta \in \mathbb{N}$ is a considerable drawback of the proposed method.

In the second part of the thesis, Papers III–V, we consider therefore the fractional order equation (recall the definition of a fractional power operator from §2.3)

$$A^\beta u = g + \mathcal{W},$$

for an elliptic operator $A\colon \mathcal{D}(A) \subset H \to H$ as in §2.4, $g \in H$, and Gaussian white noise $\mathcal{W}$ on the separable Hilbert space H, see §4.1.3. Note that (SDE_β) is a member of this class of equations.

In Paper III we propose a numerical scheme that, for any $\beta \in (0, 1)$, generates samples of an approximation $u_{h,k}^Q$ to the Gaussian solution process u, with values in a finite-dimensional subspace $V_h \subset V := \mathcal{D}(A^{1/2})$. Recall from §5.1 that A_h denotes the Galerkin discretization of A with respect to V_h and that Π_h denotes the H-orthogonal projection onto V_h. The presented approximation,

$$u_{h,k}^Q := Q_{h,k}^\beta(\Pi_h g + \mathcal{W}_h^\Phi),$$

is based on the following two components:

- The operator $Q_{h,k}^\beta$ is the quadrature approximation for $A_h^{-\beta}$ of [4]:

  $$Q_{h,k}^\beta := \frac{2k\sin(\pi\beta)}{\pi} \sum_{\ell=-K^-}^{K^+} e^{2\beta y_\ell}(I + e^{2y_\ell}A_h)^{-1}.$$

  The quadrature nodes $\{y_\ell = \ell k : \ell \in \mathbb{Z}, -K^- \leq \ell \leq K^+\}$ are equidistant with distance $k > 0$ and $K^- := \lceil \frac{\pi^2}{4\beta k^2} \rceil$, $K^+ := \lceil \frac{\pi^2}{4(1-\beta)k^2} \rceil$.
- The white noise $\mathcal{W}$ on H is approximated by the V_h-valued random variable $\mathcal{W}_h^\Phi \in L_2(\Omega; V_h)$ given by $\mathcal{W}_h^\Phi := \sum_{j=1}^{N_h} \xi_j \, \phi_{j,h}$, where $\Phi := \{\phi_{j,h}\}_{j=1}^{N_h}$ is any basis of the finite element space V_h. The vector $\boldsymbol{\xi}$ is multivariate Gaussian

distributed with mean zero and covariance matrix $\mathbf{M}^{-1}$, where $\mathbf{M}$ denotes the mass matrix with respect to the basis Φ, i.e., $M_{ij} = (\phi_{i,h}, \phi_{j,h})_H$.

Note that this method exploits the second integral representation of negative fractional power operators from §2.3:

$$A^{-\beta} = \frac{\sin(\pi\beta)}{\pi} \int_0^\infty t^{-\beta}(tI + A)^{-1} \, \mathrm{d}t = \frac{2\sin(\pi\beta)}{\pi} \int_{-\infty}^\infty e^{2\beta y}(I + e^{2y}A)^{-1} \, \mathrm{d}y,$$

where the latter equality holds due to the change of variables $t = e^{-2y}$.

For deriving an explicit rate of convergence for the strong mean-square error $\|u - u_{h,k}^Q\|_{L_2(\Omega;H)}$ of the proposed approximation, we make some assumptions on the operator A and the approximation properties of the finite-dimensional subspaces $(V_h)_{h\in(0,1)}$ of V. We recall the eigenvalue–eigenvector pairs $\{(\lambda_j, e_j)\}_{j\in\mathbb{N}}$ and $\{(\lambda_{j,h}, e_{j,h})\}_{j=1}^{N_h}$ of A and A_h from §2.4 and §5.1, respectively, and summarize the assumptions below.

(i) There exists an exponent $\alpha > 0$ with $\lambda_j \propto j^\alpha$, i.e., there are constants $c_\lambda, C_\lambda > 0$ such that $c_\lambda j^\alpha \leq \lambda_j \leq C_\lambda j^\alpha$ for all $j \in \mathbb{N}$, and we assume that α and the fractional power $\beta \in (0, 1)$ are such that $2\alpha\beta > 1$;
(ii) $N_h := \dim(V_h) \propto h^{-d}$, i.e., the family $(V_h)_{h\in(0,1)}$ is quasi-uniform;
(iii) $\lambda_{j,h}$ converges to λ_j with rate $r > 0$ for all $1 \leq j \leq N_h$.
(iv) $e_{j,h}$ converges in H to e_j with rate $s > 0$ for all $1 \leq j \leq N_h$.

We emphasize that these are technical assumptions, which we need to prove convergence of the approximation $u_{h,k}^Q$ to the exact solution u, but no explicit knowledge of the spectrum or the eigenvectors of A and A_h is necessary to compute $u_{h,k}^Q$ in practice. This is a major advantage compared to methods, which exploit the formal Karhunen–Loève expansion of the white noise from §4.1.3,

$$A^{-\beta}\mathcal{W} = \sum_{j\in\mathbb{N}} \lambda_j^{-\beta} \xi_j \, e_j, \qquad \xi_j \sim \mathcal{N}(0, 1) \text{ i.i.d.},$$

and truncate this series after finitely many terms. Note that, in contrast to the white noise $\mathcal{W}$ itself, the above series representation of $A^{-\beta}\mathcal{W}$ converges in $L_2(\Omega; H)$ if $2\alpha\beta > 1$.

The main outcome of Paper III is strong convergence of the approximation $u_{h,k}^Q$ to u in $L_2(\Omega; H)$ at the explicit rate

$$\min\{d(\alpha\beta - 1/2), r, s\},$$

given that the quadrature step size $k > 0$ is calibrated appropriately with the spatial discretization parameter h. In particular, for the motivating example (SDE$_\beta$) of Matérn approximations on $D := (0, 1)^d$ we have $A = \kappa^2 - \Delta$ and $\alpha = 2/d$, see [8, Ch. VI.4]. Thus, if $(V_h)_{h\in(0,1)}$ is a quasi-uniform family of finite element spaces with piecewise linear basis functions, we have $r = s = 2$ by Theorem 5.1 and obtain the convergence rate $\min\{(2\beta - d/2), 2\}$ for the approximation $u_{h,k}^Q$. This result is verified by numerical experiments in $d = 1, 2, 3$ spatial dimensions.

Subsequently, in Paper IV we focus on weak approximations based on $u_{h,k}^Q$, i.e., we investigate the weak error

$$\big|\mathbb{E}[\varphi(u)] - \mathbb{E}[\varphi(u_{h,k}^Q)]\big|,$$

where $\varphi \in C^2(H;\mathbb{R})$ is a twice continuously Fréchet differentiable real-valued function, see §2.1. Furthermore, we assume that the second Fréchet derivative $D^2\varphi$ has at most polynomial growth of degree $p \in \mathbb{N}$, i.e., there exist a constant $K > 0$ such that

$$\|D^2\varphi(x)\|_{\mathcal{L}(H)} \leq K\,(1 + \|x\|_H^p) \quad \forall x \in H.$$

For the error analysis, we introduce two time-dependent stochastic processes, $Y := (Y(t), t \in [0,1])$ and $\widetilde{Y} := (\widetilde{Y}(t), t \in [0,1])$, which at time $t = 1$ have the same probability distribution as the exact solution and the approximation, respectively,

$$Y(1) \overset{d}{=} u \qquad \text{and} \qquad \widetilde{Y}(1) \overset{d}{=} u_{h,k}^Q.$$

These processes are Itô processes (see §4.4) driven by the same H-valued Wiener process $W^\beta := (W^\beta(t), t \geq 0)$ defined by

$$W^\beta(t) := \sum_{j\in\mathbb{N}} \lambda_j^{-\beta} B_j(t) e_j, \quad t \geq 0.$$

Here, $\{B_j\}_{j\in\mathbb{N}}$ is a sequence of independent real-valued Brownian motions. By construction, the covariance operator of W^β is the negative fractional power operator $A^{-\beta}$, which is of trace class if $2\alpha\beta > 1$. Furthermore, the process Y defined by $Y(t) := A^{-\beta}g + W^\beta(t)$ has the desired property, $Y(1) \overset{d}{=} u$. We then introduce the Kolmogorov backward equation

$$w_t(t,x) + \frac{1}{2}\operatorname{tr}\left(w_{xx}(t,x)A^{-2\beta}\right) = 0, \quad t \in [0,1],\ x \in H, \quad w(1,x) = \varphi(x),$$

with terminal value φ at time $t = 1$. Besides to the functional φ, this equation is also related to the process Y, as its solution $w\colon [0,1] \times H \to \mathbb{R}$ is given by the following expectation, see [11, Rem. 3.2.1 & Thm. 3.2.3],

$$w(t,x) = \mathbb{E}[\varphi(x + Y(1) - Y(t))].$$

Finally, the weak error is expressed as the difference between

$$\begin{aligned} &\mathbb{E}[w(1,Y(1))] = \mathbb{E}[\varphi(Y(1))] = \mathbb{E}[\varphi(u)] \\ \text{and} \quad &\mathbb{E}[w(1,\widetilde{Y}(1))] = \mathbb{E}[\varphi(\widetilde{Y}(1))] = \mathbb{E}[\varphi(u_{h,k}^Q)], \end{aligned}$$

which is bounded by applying the Itô formula from §4.4 to the processes

$$(w(t,Y(t)), t \in [0,1]) \quad \text{and} \quad (w(t,\widetilde{Y}(t)), t \in [0,1]).$$

Under similar assumptions as in Paper III, we thereby prove convergence of the weak error to zero with the rate

$$\min\{d(2\alpha\beta - 1), r, s\}.$$

Note that, compared to the strong convergence rate derived in Paper III, the component stemming from the stochasticity of the problem is doubled. We verify this outcome by numerical experiments for various functionals φ and the Matérn example (SDE_β) from Paper III on the unit square $D := (0,1)^2$.

The focus of Paper V is to use the quadrature approximation $u_{h,k}^Q$ introduced in Paper III for common tasks in statistics, such as inference and kriging, i.e., spatial prediction. Besides this method, we introduce the approximation $u_{h,m}^R$, referred to as the rational (SPDE) approximation of degree m, which, for any $\beta > 0$, exploits a rational approximation $\frac{p_\ell(x)}{p_r(x)}$ of the function x^β instead of the quadrature. Here, the functions p_ℓ and p_r are polynomials of degree $m_\ell := m + \max\{1, \lfloor \beta \rfloor\}$ and $m_r := m \in \mathbb{N}$, respectively. Specifically, the rational approximation of degree m is defined as the solution of the equation

$$P_{\ell,h} u_{h,m}^R = P_{r,h} \mathcal{W}_h^\Phi,$$

where the operators $P_{\ell,h}$, $P_{r,h}$ are defined by $P_{j,h} := p_j(A_h)$, $j \in \{\ell, r\}$, in terms of the polynomials p_ℓ, p_r, and the discretized operator A_h.

We investigate how accurately Gaussian random fields with Matérn covariance functions can be approximated by rational approximations of the solution u to (SDE_β), i.e., $A = \kappa^2 - \Delta$. For $\beta \in (0,1)$, we furthermore compare this performance as well as the computational cost with the quadrature method from Paper III. We find that the asymptotic behavior is the same, namely exponential convergence of the form $e^{-c\sqrt{N}}$, where, for the rational approximation, $N := m$ is the polynomial degree, and for the quadrature method $N := K$ is the total number of quadrature nodes $K := K^- + K^+ + 1 \propto k^{-2}$. However, when calibrating m and K with the spatial finite element discretization parameter h, we obtain $m = K/(4\beta)$. Due to $\beta > d/4$, this shows that, for the same accuracy, we can choose the degree m of the rational approximation smaller than the number of quadrature nodes. This observation is attested empirically by numerical experiments. Since the computational cost for inference increase rapidly with m and K, this is an important advantage of the rational approximation $u_{h,m}^R$ compared to the quadrature approximation $u_{h,k}^Q$.

Because of this computational benefit, we exploit the rational approximation for estimating the parameters $\beta, \kappa, \tau > 0$ in the model (SDE_β) from a generated data set of samples of a Gaussian Matérn field. We observe that the rational approximation facilitates likelihood-based (or Bayesian) inference of all model parameters, including the smoothness parameter β, which had to be fixed during inference until now.

Finally, we apply these approximations for spatial prediction and compare the accuracy of the method to that of covariance tapering, which is a common approach for reducing the computational cost in spatial statistics. The application shows that the approach based on the rational approximation is both faster and more accurate than the covariance tapering method.

To conclude, in the second part of thesis, we introduce a new numerical scheme for solving fractional order elliptic SPDEs with spatial white noise ap-

proximately. For the resulting approximations, we discuss

- strong convergence in Paper III,
- weak convergence for twice continuously Fréchet differentiable functionals, whose second derivatives are of polynomial growth, in Paper IV, and
- its usage for statistical inference in Paper V.

These outcomes should prove useful for further applications in spatial statistics, such as, to models based on non-Gaussian random fields (for the sake of brevity, we have only addressed Gaussian Matérn models in Paper V).

References

[1] D. Applebaum, *Lévy Processes and Stochastic Calculus*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2009.

[2] A. V. Balakrishnan, *Fractional powers of closed operators and the semigroups generated by them.*, Pacific J. Math., 10 (1960), pp. 419–437.

[3] ——, *Applied Functional Analysis*, vol. 3 of Applications of Mathematics, Springer-Verlag, New York-Berlin, second ed., 1981.

[4] A. Bonito and J. E. Pasciak, *Numerical approximation of fractional powers of elliptic operators*, Math. Comp., 84 (2015), pp. 2083–2110.

[5] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, vol. 15 of Texts in Applied Mathematics, Springer, New York, third ed., 2008.

[6] Z. Brzeźniak, *Some remarks on Itô and Stratonovich integration in 2-smooth Banach spaces*, in Probabilistic methods in fluids, World Sci. Publ., River Edge, NJ, 2003, pp. 48–69.

[7] Z. Brzeźniak, J. M. A. M. van Neerven, M. C. Veraar, and L. Weis, *Itô's formula in UMD Banach spaces and regularity of solutions of the Zakai equation*, J. Diff. Equations, 245 (2008), pp. 30–58.

[8] R. Courant and D. Hilbert, *Methods of Mathematical Physics. Vol. I*, Interscience Publishers, Inc., New York, N.Y., 1953.

[9] G. Da Prato, A. Jentzen, and M. Roeckner, *A mild Itô formula for SPDEs*. arXiv preprint, arXiv:1009.3526, 2012.

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

[11] ——, *Second Order Partial Differential Equations in Hilbert Spaces*, London Mathematical Society Lecture Note Series, Cambridge University Press, 2002.

[12] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology: Volume 5 Evolution Problems I*, Springer, 2000.

[13] N. Dunford and J. Schwartz, *Linear Operators. Part I: General Theory*, Pure and Applied Mathematics, Interscience Publishers, 1958.

[14] ——, *Linear Operators. Part II: Spectral theory. Self Adjoint Operators in Hilbert Space*, Pure and Applied Mathematics, Interscience Publishers, 1963.

[15] K.-J. Engel and R. Nagel, *One-Parameter Semigroups for Linear Evolution Equations*, vol. 194 of Graduate Texts in Mathematics, Springer, 2000.

[16] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, vol. 159 of Applied Mathematical Sciences, Springer-Verlag, New York, 2004.

[17] A. Grothendieck, *La théorie de Fredholm*, Bull. Soc. Math. France, 84 (1956), pp. 319–384.

[18] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*, Springer Series in Computational Mathematics, Springer, 2012.

[19] E. Hille and R. S. Phillips, *Functional Analysis and Semi-Groups*, American Mathematical Society Colloquium Publications, vol. 31, American Mathematical Society, Providence, R. I., 1957. rev. ed.

[20] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, vol. 24 of North-Holland Mathematical Library, North-Holland Publishing Co., Amsterdam; Kodansha, Ltd., Tokyo, second ed., 1989.

[21] W. Johnson and J. Lindenstrauss, *Handbook of the Geometry of Banach Spaces, Vol. 2*, Elsevier Science B.V., 2003.

[22] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, vol. 113 of Graduate Texts in Mathematics, Springer-Verlag, New York, second ed., 1991.

[23] R. Kruse, *Strong and Weak Approximation of Semilinear Stochastic Evolution Equations*, vol. 2093 of Lecture Notes in Mathematics, Springer, 2014.

[24] A. Lang, S. Larsson, and Ch. Schwab, *Covariance structure of parabolic stochastic partial differential equations*, Stoch. PDE: Anal. Comp., 1 (2013), pp. 351–364.

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

[26] A. Lunardi, *Analytic Semigroups and Optimal Regularity in Parabolic Problems*, Progress in Nonlinear Differential Equations and their Applications, Birkhäuser Verlag, 1995.

[27] B. Matérn, *Spatial variation*, Meddelanden från statens skogsforskningsinstitut, 49 (1960).

[28] J. Mercer, *Functions of positive and negative type and their connection with the theory of integral equations*, Philosophical Transactions of the Royal Society A, 209 (1909), pp. 415–446.

[29] D. Nualart, *The Malliavin Calculus and Related Topics*, Probability and its Applications, Springer-Verlag, Berlin, second ed., 2006.

[30] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, Universitext, Springer, 6th ed., 2013.

[31] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Applied Mathematical Sciences, Springer, 1983.

[32] S. Peszat and J. Zabczyk, *Stochastic Partial Differential Equations with Lévy Noise*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2007.

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

[34] R. Ryan, *Introduction to Tensor Products of Banach Spaces*, Springer Monographs in Mathematics, Springer, 2002.
[35] R. Schatten, *A Theory of Cross-Spaces*, Annals of Mathematics Studies, no. 26, Princeton University Press, Princeton, N. J., 1950.
[36] Ch. Schwab and R. Stevenson, *Space-time adaptive wavelet methods for parabolic evolution problems*, Math. Comp., 78 (2009), pp. 1293–1318.
[37] Ch. Schwab and E. Süli, *Adaptive Galerkin approximation algorithms for Kolmogorov equations in infinite dimensions*, Stoch. PDE: Anal. Comp., 1 (2013), pp. 204–239.
[38] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Series in Statistics, Springer New York, 1999.
[39] G. Strang and G. Fix, *An Analysis of the Finite Element Method*, Wellesley-Cambridge Press, 2008.
[40] V. Thomee, *Galerkin Finite Element Methods for Parabolic Problems*, Springer Series in Computational Mathematics, Springer Berlin Heidelberg, 2006.
[41] P. Whittle, *On stationary processes in the plane*, Biometrika, 41 (1954), pp. 434–449.
[42] ——, *Stochastic processes in several dimensions*, Bull. Internat. Statist. Inst., 40 (1963), pp. 974–994.
[43] K. Yosida, *Functional Analysis*, Classics in Mathematics, Springer, 6th ed., 1995.