Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs

Acta Numerica (2011), pp. 291–467
doi:10.1017/S0962492911000055

Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs*

Christoph Schwab and Claude Jeffrey Gittelson
Seminar for Applied Mathematics, ETH Zürich,
Rämistrasse 101, CH-8092 Zürich, Switzerland
E-mail: schwab@sam.math.ethz.ch, claude.gittelson@sam.math.ethz.ch

Partial differential equations (PDEs) with random input data, such as random loadings and coefficients, are reformulated as parametric, deterministic PDEs on parameter spaces of high, possibly infinite dimension. Tensorized operator equations for spatial and temporal k-point correlation functions of their random solutions are derived. Parametric, deterministic PDEs for the laws of the random solutions are derived. Representations of the random solutions' laws on infinite-dimensional parameter spaces in terms of 'generalized polynomial chaos' (GPC) series are established. Recent results on the regularity of solutions of these parametric PDEs are presented. Convergence rates of best N-term approximations, for adaptive stochastic Galerkin and collocation discretizations of the parametric, deterministic PDEs, are established. Sparse tensor products of hierarchical (multi-level) discretizations in physical space (and time), and GPC expansions in parameter space, are shown to converge at rates which are independent of the dimension of the parameter space. A convergence analysis of multi-level Monte Carlo (MLMC) discretizations of PDEs with random coefficients is presented. Sufficient conditions on the random inputs for superiority of sparse tensor discretizations over MLMC discretizations are established for linear elliptic, parabolic and hyperbolic PDEs with random coefficients.

* Work partially supported by the European Research Council under grant number ERC AdG 247277-STAHDPDE and by the Swiss National Science Foundation under grant number SNF 200021-120290/1.

CONTENTS

Introduction

The numerical solution of partial differential equation models in science and engineering has today reached a certain maturity, after several decades of progress in numerical analysis, mathematical modelling and scientific computing. While there certainly remain numerous mathematical and algorithmic challenges, for many 'routine' problems of engineering interest, today numerical solution methods exist which are mathematically understood and 'operational' in the sense that a number of implementations exist, both academic and commercial, which realize, in the best case, algorithms of provably optimal complexity in a wide range of applications. As a rule, the numerical analysis and the numerical solution methods behind such algorithms suppose that a model of the system of interest is described by a well-posed (in the sense of Hadamard) partial differential equation (PDE), and that the PDE is to be solved numerically to prescribed accuracy *for one given set of input data*.

With the availability of highly accurate numerical solution algorithms for a PDE of interest and *one* prescribed set of *exact* input data (such as source terms, constitutive laws and material parameters) there has been increasing awareness of the limited significance of such single, highly accurate 'forward' solves. Assuming, as we will throughout this article, that *the PDE model of the physical system of interest is correct*, this trend is due to two reasons: randomness and uncertainty of input data and the need for efficient prediction of system responses on high-dimensional parameter spaces.

First, the assumption of availability of exact input data is not realistic: often, the simulation's input parameters are obtained from measurements or from sampling a large, but finite number of specimens or system snapshots which are incomplete or stochastic. This is of increasing importance in classical engineering disciplines, but even more so in emerging models in

the life sciences and social sciences. Rather than producing efficiently accurate answers for single instances of exact input data, increasingly the goal of computation in numerical simulations is to *efficiently process statistical information on uncertain input data* for the PDE of interest. While mathematical formulations of PDEs with random inputs have been developed with an eye towards uncorrelated, or *white noise* inputs (see, *e.g.*, Holden, Oksendal, Uboe and Zhang (1996), Da Prato and Zabczyk (1992), Da Prato (2006), Lototsky and Rozovskii (2006), Prévôt and Röckner (2007), Dalang, Khoshnevisan, Mueller, Nualart and Xiao (2009) and the references therein), PDEs with random inputs in numerical simulation in science and engineering are of interest in particular in the case of so-called *correlated inputs* (or 'coloured noise').

Second, in the context of optimization, or of risk and sensitivity analysis for complex systems with random inputs, the interest is in computing the systems' responses efficiently given dependence on several, possibly countably many parameters, thereby leading to the challenge of *numerical simulation of deterministic PDEs on high-dimensional parameter spaces*.

Often, the only feasible approach in numerical simulation towards these two problems is to solve the forward problem for many instances, or samples, of the PDE's input parameters; for random inputs, this amounts to Monte Carlo-type sampling of the noisy inputs, and for parametric PDEs, responses of the system are interpolated from forward solves at judiciously chosen combinations of input parameters.

With the cost of one 'sample' being the numerical solution of a PDE, it is immediate that, in particular for transient problems in three spatial dimensions with solutions that exhibit multiple spatial and temporal length scales, the computational cost of uniformly sampling the PDE solution on the parameter space (resp. the probability space) is prohibitive. Responding to this by massive parallelism may alleviate this problem, but ultimately, the low convergence rate 1/2 of Monte Carlo (MC) sampling, respectively the so-called 'curse of dimensionality' of standard interpolation schemes in high-dimensional parameter spaces, requires advances at the mathematical core of the numerical PDE solution methods: the development of novel mathematical formulations of PDEs with random inputs, the study of the regularity of their solutions is of interest, both with respect to the physical variables and with respect to parameters, and the development of novel discretizations and solution methods of these formulations. Importantly, the parameters may take values in possibly infinite-dimensional parameter spaces: for example, in connection with Karhunen–Loève expansions of spatially inhomogeneous and correlated inputs.

The present article surveys recent contributions to the above questions. Our focus is on linear PDEs with random inputs; we present various formulations, new results on the regularity of their solutions and, based on these

regularity results, we design, formulate and analyse discretization schemes which allow one to 'sweep' the entire, possibly infinite-dimensional input parameter space approximately in a single computation. We also establish, for the algorithms proposed here, bounds on their efficiency (understood as accuracy versus the number of degrees of freedom) that do not deteriorate with respect to increasing dimension of the computational parameter domain, *i.e.*, that are free from the curse of dimensionality. The algorithms proposed here are variants and refinements of the recently proposed stochastic Galerkin and stochastic collocation discretizations (see, *e.g.*, Xiu (2009) and Matthies and Keese (2005) and the references therein for an account of these developments). We exhibit assumptions on the inputs' correlations which ensure an efficiency of these algorithms which is superior to that of MC sampling. One insight that emerges from the numerical analysis of recently proposed methods is that *the numerical resolution in physical space need not be high uniformly on the entire parameter space*. The use of 'polynomial chaos' type spectral representations (and their generalizations) of the laws of input and output random fields allows a theory of regularity of the random solutions and, based on this, the optimization of numerical methods for their resolution. Here, we have in mind discretizations in physical space and time as well as in stochastic or parameter space, aiming at achieving a prespecified accuracy with minimal computational work. From this broad view, the recently proposed *multi-level Monte Carlo methods* can also be interpreted as sparse tensor discretizations. Accordingly, we present in this article an error analysis of single- and multi-level MC methods for elliptic problems with random inputs.

As this article's title suggests, the notion of *sparse tensor products* of operators and hierarchical sequences of finite-dimensional subspaces pervades our view of numerical analysis of high-dimensional problems. Sparsity in connection with tensorization has become significant in several areas of scientific computing in recent years: in approximation theory as *hyperbolic cross* approximations (see, *e.g.*, Temlyakov (1993)) and, in finite element and finite difference discretizations, the so-called *sparse grids* (see Bungartz and Griebel (2004) and the references therein) are particular instances of this concept. We note in passing that the range of applicability of sparse tensor discretizations extends well beyond stochastic and parametric problems (see, *e.g.*, Schwab (2002), Hoang and Schwab (2004/05) and Schwab and Stevenson (2008) for applications to multiscale problems). On the level of numerical linear algebra, the currently emerging *hierarchical low-rank matrix formats*, which were inspired by developments in computational chemistry, are closely related to some of the techniques developed here.

The present article extends these concepts in several directions. First, on the level of mathematical formulation of PDEs with random inputs: we present deterministic tensorized operator equations for two- and k-point

correlation functions of the the random system responses. Such equations also arise in the context of moment closures of kinetic models in atomistic-to-continuum transitions. Discretizations for their efficient, deterministic numerical solution may therefore be of interest in their own right. For the spectral discretizations, we review the polynomial chaos representation of random fields and the Wiener–Itô chaos decomposition of probability spaces and of random fields into tensorized Hermite polynomials of a countable number of Gaussians. The spectral representation of random outputs of PDEs allows for a regularity theory of the laws of random fields which goes substantially beyond the mere existence of moments.

According to the particular application, in this article sparsity in tensor discretizations appears in roughly three forms. First, we use sparse tensor products of multi-level finite element spaces in the physical domain $D \subset \mathbb{R}^d$ to build efficient schemes for the Galerkin approximation of tensorized equations for k-point correlation functions. Second, we consider *heterogeneous* sparse tensor product discretizations of *multi-level finite element, finite volume and finite difference discretizations* in the physical domain with *hierarchical polynomial chaos bases* in the probability space. As we will show, the use of multi-level discretizations in physical space actually leads to substantial efficiency gains in MC methods; nevertheless, the resulting multi-level MC methods are of comparable efficiency as sparse tensor discretizations *for random outputs with finite second moments*. However, as soon as the outputs have additional summability properties (and the examples presented here suggest that this is so in many cases), adaptive sparse tensor discretizations outperform MLMC methods.

The outline of the article is as follows. We first derive tensorized operator equations for deterministic, linear equations with random data. We establish the well-posedness of these tensorized operator equations, and introduce sparse tensor Galerkin discretizations based on multi-level, wavelet-type finite element spaces in the physical domain. We prove, in particular, stability of sparse tensor discretizations in the case of indefinite operators such as those arising in acoustic or electromagnetic scattering. We also give an error analysis of MC discretizations which indicates the dependence of its convergence rate on the degree of summability of the random solution.

Section 2 is devoted to stochastic Galerkin formulations of PDEs with random coefficients. Using polynomial chaos representations of the random inputs, for example in a Karhunen–Loève expansion, we give a reformulation of the random PDEs of interest as deterministic PDEs which are posed on *infinite-dimensional parameter spaces*. While the numerical solution of these PDEs with standard tools from numerical analysis is foiled by the curse of dimensionality (the *raison d'être* for the use of sampling methods on the stochastic formulation), we review recent regularity results for these problems which indicate that sparse, adaptive tensorization of discretizations

in probability and physical space can indeed produce solutions whose accuracy, as a function of work, is independent of the dimension of the parameter space. We cover both affine dependence, as is typical in Karhunen–Loève representations of the random inputs, as well as log-normal dependence in inputs. We focus on Gaussian and on uniform measures, where 'polynomial chaos' representations use Hermite and Legendre polynomials, respectively (other probability measures give rise to other polynomial systems: see, *e.g.*, Schoutens (2000) and Xiu and Karniadakis (2002*b*)). Section 3 addresses the regularity of the random solutions in these polynomial chaos representations by an analysis of the associated parametric, deterministic PDE for their laws. The analysis allows us to deduce best N-term convergence rates of *polynomial chaos semidiscretizations* of the random solutions' laws.

Section 4 combines the results from the preceding sections with space and time discretizations in the physical domain. The error analysis of fully discrete algorithms reveals that it is crucial for efficiency that the level of spatial and temporal resolution be allowed to depend on the stochastic mode being discretized. Our analysis shows that, in fact, a highly non-uniform level of resolution in physical space should be adopted in order to achieve algorithms that scale favourably with respect to the dimension of the space of stochastic parameters.

As this article and the subject matter draw on tools from numerical analysis, from functional analysis and from probability theory, we provide some background reference material on the latter two items in the Appendix. This is done in order to fix the notation used in the main body of the text, and to serve as a reference for readers with a numerical analysis background. Naturally, the selection of the background material is biased towards the subject matter of the main text. It does not claim to be a reference on these subjects. For a more thorough introduction to tools from probability and stochastic analysis we refer the reader to Bauer (1996), Da Prato (2006), Da Prato and Zabczyk (1992), Prévôt and Röckner (2007) and the references therein.

1. Sparse tensor FEM for operator equations with stochastic data

For the variational setting of linear operator equations with deterministic, boundedly invertible operators, we assume that X, Y are separable Hilbert spaces over $\mathbb{R}$ with duals X' and Y', respectively, and $A \in L(X, Y')$ a linear, boundedly invertible deterministic operator. We denote its associated bilinear form by

$$a(u, v) := {}_{Y'}\langle Au, v\rangle_X : X \times Y \to \mathbb{R}. \tag{1.1}$$

Here, and throughout, for $w \in Y'$ and $v \in X$ the bilinear form ${}_{Y'}\langle w, v\rangle_X$ denotes the $Y' \times X$ duality pairing. As is well known (see, *e.g.*, Theorem C.20) the operator A from X onto Y' is boundedly invertible if and only if $a(\cdot,\cdot)$ satisfies the following conditions.

(i) $a(\cdot,\cdot)$ is continuous: there exists $C_1 < \infty$ such that

$$\forall w \in X,\ v \in Y: \quad |a(w,v)| \le C_1 \|w\|_X \|v\|_Y. \tag{1.2}$$

(ii) $a(\cdot,\cdot)$ is coercive: there exists $C_2 > 0$ such that

$$\inf_{0\neq w\in X} \sup_{0\neq v\in Y} \frac{a(w,v)}{\|w\|_X \|v\|_Y} \ge C_2 > 0. \tag{1.3}$$

(iii) $a(\cdot,\cdot)$ is injective:

$$\forall 0 \neq v \in Y: \quad \sup_{0\neq w\in X} a(w,v) > 0. \tag{1.4}$$

If (1.2)–(1.4) hold, then for every $f \in Y'$ the linear operator equation

$$u \in X: \quad a(u,v) = {}_{Y'}\langle f, v\rangle_X \quad \forall v \in Y \tag{1.5}$$

admits a unique solution $u \in X$ such that

$$\|u\|_X \le C_2^{-1} \|f\|_{Y'}. \tag{1.6}$$

We consider equation (1.5) with stochastic data: to this end, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $f : \Omega \to Y'$ be a random field, *i.e.*, a measurable map from $(\Omega, \mathcal{F}, \mathbb{P})$ into Y' which is *Gaussian* (see Appendix C for the definition of Gaussian random fields). Analogous to the characterization of Gaussian random variables by their mean and their (co)variance, a Gaussian random field $f \in L^2(\Omega, \mathcal{F}, \mathbb{P}; Y')$ is characterized by its mean $a_f \in Y'$ and its covariance operator $Q_f \in \mathcal{L}_1^+(Y')$.

We use the following *linear operator equation with Gaussian data*: given $f \in L^2(\Omega, \mathcal{F}, \mathbb{P}; Y')$, find $u \in L^2(\Omega, \mathcal{F}, \mathbb{P}; X)$ such that

$$Au = f \quad \text{in} \quad L^2(\Omega, \mathcal{F}, \mathbb{P}; Y') \tag{1.7}$$

admits a unique solution $u \in L^2(\Omega, \mathcal{F}, \mathbb{P}; X)$ if and only if A satisfies (1.2)–(1.4).

By Theorem C.31, the unique random solution $u \in L^2(\Omega, \mathcal{F}, \mathbb{P}; X)$ of (1.7) is Gaussian with associated Gaussian measure N_{a_u, Q_u} on X which, in turn, is *characterized* by the solution's mean,

$$a_u = \mathrm{mean}(u) = A^{-1} a_f, \tag{1.8}$$

and the solution's covariance operator $Q_u \in \mathcal{L}_1^+(X)$, which satisfies the (deterministic) equation

$$A Q_u A^* = Q_f \quad \text{in} \quad \mathcal{L}(Y', Y'). \tag{1.9}$$

In the Gaussian case, therefore, solving the stochastic problem (1.7) can be reduced to solving the two *deterministic problems* (1.8) and (1.9). Whereas the mean-field problem (1.8) is one instance of the operator equation (1.7), the covariance equation (1.8) is an equation for the *operator* $Q_u \in \mathcal{L}_1^+(X)$. As we show in Theorem C.31, this operator is characterized by the so-called *covariance kernel* C_u, which satisfies, in terms of the corresponding covariance kernel C_f of the data, the *covariance equation* (see (C.50))

$$(A \otimes A)C_u = C_f, \tag{1.10}$$

which is understood to hold in the sense of $(Y \otimes Y)' \simeq Y' \otimes Y'$. One approach to the numerical treatment of operator equations $Au = f$, where the data f are *random fields*, *i.e.*, measurable maps from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into the set Y' of admissible data for the operator A, is via tensorized equations such as (1.10) for their statistical moments.

The simplest approach to the numerical solution of the linear operator equation $Au = f$ with random input f is Monte Carlo (MC) simulation, *i.e.*, generating a large number M of i.i.d. data samples f_j and solving, possibly in parallel, for the corresponding solution ensemble $\{u_j = A^{-1}f_j\,;\ j = 1, \dots, M\}$. Statistical moments and probabilities of the random solution u are then estimated from $\{u_j\}$. As we will prove, convergence of the MC method as the number M of samples increases is ensured (for suitable sampling) by the central limit theorem. We shall see that the MC method allows in general only the convergence rate $\mathcal{O}(M^{-1/2})$.

If statistical moments, *i.e.*, mean-field and higher-order moments of the random solution u, are of interest, one can exploit the linearity of the equation $Au = f$ to derive a deterministic equation for the kth moment of the random solution, similar to the second-moment equation (1.10); this derivation is done in Section 1.1. For the Laplace equation with stochastic data, this approach is due to I. Babuška (1961). We then address the numerical computation of the moments of the solution by either Monte Carlo or by direct, deterministic finite element computation. If the physical problem is posed in a domain $D \subset \mathbb{R}^d$, the kth moment of the random solution is defined in the domain $D^k \subset \mathbb{R}^{kd}$; standard finite element (FE) approximations will therefore be inadequate for the efficient numerical approximation of the kth moments of the random solution.

The efficient deterministic equation and its FE approximation were investigated in Schwab and Todor (2003*a*, 2003*b*) in the case where A is an elliptic partial differential operator. It was shown that the kth moment of the solution could be computed in a complexity comparable to that of an FE solution for the mean-field problem by the use of sparse tensor products of standard FE spaces for which a hierarchical basis is available. The use of sparse tensor product approximations is a well-known device in high-dimensional numerical integration going back to Smolyak (1963), in

multivariate approximation (Temlyakov 1993), and in complexity theory; see Wasilkowski and Woźniakowski (1995) and the references therein.

In the present section, we address the case when A is a non-local operator, such as a strongly elliptic pseudodifferential operator, as arises in the boundary reduction of boundary value problems for strongly elliptic partial differential equations. In this case, efficient numerical solution methods require, in addition to Galerkin discretizations of the operator equation, some form of matrix compression (such as the fast multipole method or wavelet-based matrix compression) which introduces additional errors into the Galerkin solution that will also affect the accuracy of second and higher moments. We briefly present the numerical analysis of the impact of matrix compressions on the efficient computation of second and higher moments of the random solution. Therefore, the present section will also apply to strongly elliptic boundary integral equations obtained by reduction to the boundary manifold $D = \partial\mathcal{D}$ of elliptic boundary value problems in a bounded domain $\mathcal{D} \subset \mathbb{R}^{d+1}$, as is frequently done in acoustic and electromagnetic scattering. For such problems with stochastic data, the boundary integral formulation leads to an operator equation $Au = f$, where A is an integral operator or, more generally, a pseudodifferential operator acting on function spaces on $\partial\mathcal{D}$. The linearity of the operator equation allows, without any closure hypothesis, formulation of a *deterministic* tensor equation for the k-point correlation function of the random solution $u = A^{-1}f$. We show that, as in the case of differential operators, sparse tensor products of standard FE spaces allow deterministic approximation of the kth moment of the random solution u with relatively few degrees of freedom. To achieve computational complexity which scales log-linearly in the number of degrees of freedom in a Galerkin discretization of the mean-field problem, however, the Galerkin matrix for the operator A must be *compressed*.

Accordingly, one purpose of this section is the design and numerical analysis of deterministic and stochastic solution algorithms to obtain the kth moment of the random solution of possibly non-local operator equations with random data in log-linear complexity in the number N of degrees of freedom for the mean-field problem.

We illustrate the sparse tensor product Galerkin methods for the numerical solution of Dirichlet and Neumann problems for the Laplace or Helmholtz equation with stochastic data. Using a wavelet Galerkin finite element discretization allows straightforward construction of sparse tensor products of the trial spaces, and yields well-conditioned, sparse representations of stiffness matrices for the operator A as well as for its k-fold tensor product, which is the operator arising in the kth-moment problem.

We analyse the impact of the operator compression on the accuracy of functionals of the Galerkin solution, such as far-field evaluations of the random potential in a point. For example, means and variances of the potential

in a point can be computed with accuracy $\mathcal{O}(N^{-p})$ for any fixed order p, for random boundary data with known second moments in $\mathcal{O}(N)$ complexity, where N denotes the number of degrees of freedom on the boundary.

The outline of this section is as follows. In Section 1.1, we describe the operator equations considered here and derive the deterministic problems for the higher moments, generalizing Schwab and Todor (2003*b*). We establish the Fredholm property for the tensor product operator and regularity estimates for the statistical moments in anisotropic Sobolev spaces with mixed highest derivative. Section 1.2 addresses the numerical solution of the moment equations, in particular the impact of various matrix compressions on the accuracy of the approximated moments, the preconditioning of the product operator and the solution algorithm. In Section 1.4, we discuss the implementation of the sparse Galerkin and sparse MC methods and estimate their asymptotic complexity. Section 1.5 contains some examples from finite and boundary element methods.

1.1. Operator equations with stochastic data

Linear operator equations

We specialize the general setting (1.1) to the case $X = Y = V$, and consider the operator equation

$$Au = f, \tag{1.11}$$

where A is a bounded linear operator from the separable Hilbert space V into its dual V'.

The operator A is a differential or pseudodifferential operator of order ϱ on a bounded d-dimensional manifold D, which may be closed or have a boundary. Here, for a closed manifold and for $s \geq 0$, $\tilde{H}^s(D) := H^s(D)$ denotes the usual Sobolev space. For $s < 0$, we define the spaces $H^s(D)$ and $\tilde{H}^s(D)$ by duality. For a manifold D with boundary we assume that this manifold can be extended to a closed manifold $\tilde{D}$, and define

$$\tilde{H}^s(D) := \{\, u|_D \;;\; u \in H^s(\tilde{D}),\, u|_{\tilde{D}\setminus D} = 0\}$$

with the induced norm. If D is a bounded domain in $\mathbb{R}^d$ we use $\tilde{D} := \mathbb{R}^d$. We now assume that $V = \tilde{H}^{\varrho/2}(D)$. In the case when A is a second-order differential operator, this means that we have Dirichlet boundary conditions (other boundary conditions can be treated in an analogous way).

The manifold D may be smooth, but we also consider the case when D is a polyhedron in $\mathbb{R}^d$, or the boundary of a polyhedron in $\mathbb{R}^{d+1}$, or part of the boundary of a polyhedron.

For the deterministic operator A in (1.11), we assume strong ellipticity in the sense that there exists $\alpha > 0$ and a compact operator $T : V \to V'$ such

that the Gårding inequality

$$\forall v \in V: \quad \big\langle (A+T)\, v, v \big\rangle \geq \alpha \, \|v\|_V^2 \tag{1.12}$$

holds. For the deterministic algorithm in Section 1.4 we need the slightly stronger assumption that T' is smoothing with respect to a scale of smoothness spaces (see (1.63) below). Here and in what follows, $\langle \cdot, \cdot \rangle$ denotes the $V' \times V$ duality pairing. We assume also that A is injective, *i.e.*, that

$$\ker A = \{0\}, \tag{1.13}$$

which implies that for every $f \in V'$, (1.11) admits a unique solution $u \in V$ and, moreover, that $A^{-1} : V' \to V$ is continuous, *i.e.*, there exists $C_A > 0$ such that, for all $f \in V'$,

$$\|u\|_V = \|A^{-1} f\|_V \leq C_A \|f\|_{V'}. \tag{1.14}$$

Here $C_A = C_2^{-1}$ with the constant C_2 as in (1.3). We shall consider (1.11) in particular for data f, which are Gaussian random fields on the data space V'. By the linearity of the operator equation (1.11), then the solution $v \in V$ is a Gaussian random field as well. Throughout, we assume that V and V' are separable Hilbert spaces.

Random data

A Gaussian random field f with values in a separable Hilbert space X is a mapping $f\colon \Omega \to X$ which maps events $E \in \Sigma$ to Borel sets in X, and such that the image measure $f_{\#}\mathbb{P}$ on X is Gaussian. In the following, we allow more general random fields. Of particular interest will be their summability properties. We say that a random field $u\colon \Omega \to X$ is in the Bochner space $L^1(\Omega; X)$ if $\omega \mapsto \|u(\omega)\|_X$ is measurable and integrable so that $\|u\|_{L^1(\Omega;X)} := \int_\Omega \|u(\omega)\|_X \, \mathbb{P}(\mathrm{d}\omega)$ is finite. In particular, then the 'ensemble average'

$$\mathbb{E}u := \int_\Omega u(\omega) \mathbb{P}(\mathrm{d}\omega) \in X$$

exists as a Bochner integral of X-valued functions, and it satisfies

$$\|\mathbb{E}u\|_X \leq \|u\|_{L^1(\Omega;X)}. \tag{1.15}$$

Let $k \geq 1$. We say that a random field $u\colon \Omega \to X$ is in the Bochner space $L^k(\Omega; X)$ if $\|u\|_{L^k(\Omega;X)}^k = \int_\Omega \|u(\omega)\|_X^k \mathbb{P}(\mathrm{d}\omega)$ is finite. Note that $\omega \mapsto \|u(\omega)\|_X^k$ is measurable due to the measurability of u and the continuity of the norm $\|\cdot\|_X$ on X. Also, $L^k(\Omega; X) \supset L^l(\Omega; X)$ for $k < l$.

Let $B \in \mathcal{L}(X, Y)$ denote a continuous linear mapping from X to another separable Hilbert space Y. For a random field $u \in L^k(\Omega; X)$, this mapping defines a random variable $v(\omega) = Bu(\omega)$ taking values in Y. Moreover, $v \in L^k(\Omega; Y)$ and we have

$$\|Bu\|_{L^k(\Omega;Y)} \leq C \|u\|_{L^k(\Omega;X)}, \tag{1.16}$$

where the constant C is given by $C = \|B\|_{\mathcal{L}(X,Y)}$. In addition, we have

$$B \int_\Omega u \, \mathbb{P}(\mathrm{d}\omega) = \int_\Omega Bu \, \mathbb{P}(\mathrm{d}\omega). \tag{1.17}$$

MC estimation of statistical moments

We are interested in statistics of the random solution u of (1.11) and, in particular, in statistical moments. To define them, for a separable Hilbert space X and for any $k \in \mathbb{N}$ we define the k-fold tensor product space

$$X^{(k)} = \underbrace{X \otimes \cdots \otimes X}_{k \text{ times}},$$

and equip it with the natural cross-norm $\|\cdot\|_{X^{(k)}}$. The significance of a cross-norm was emphasized by Schatten. The cross-norm has the property that, for every $u_1, \ldots, u_k \in X$,

$$\|u_1 \otimes \cdots \otimes u_k\|_{X^{(k)}} = \|u_1\|_X \cdots \|u_k\|_X \tag{1.18}$$

(see Light and Cheney (1985) and the references therein for more on cross-norms on tensor product spaces). The k-fold tensor products of, for example, X' are denoted analogously by $(X')^{(k)}$. For $u \in L^k(\Omega; X)$ we now consider the random field $u^{(k)}$ defined by $u(\omega) \otimes \cdots \otimes u(\omega)$. By Lemma C.9, $u^{(k)} = u \otimes \cdots \otimes u \in L^1(\Omega; X^{(k)})$, and we have the *isometry*

$$\begin{aligned} \|u^{(k)}\|_{L^1(\Omega;X^{(k)})} &= \int_\Omega \|u(\omega) \otimes \cdots \otimes u(\omega)\|_{X^{(k)}} \mathbb{P}(\mathrm{d}\omega) \\ &= \int_\Omega \|u(\omega)\|_X \cdots \|u(\omega)\|_X \mathbb{P}(\mathrm{d}\omega) = \|u\|^k_{L^k(\Omega;X)}. \end{aligned} \tag{1.19}$$

We define the moment $\mathcal{M}^k u$ as the expectation of $u \otimes \cdots \otimes u$.

Definition 1.1. For $u \in L^k(\Omega; X)$, for some integer $k \geq 1$, the kth moment of $u(\omega)$ is defined by

$$\mathcal{M}^k u = \mathbb{E}\Big[\underbrace{u \otimes \cdots \otimes u}_{k \text{ times}}\Big] = \int_{\omega \in \Omega} \underbrace{u(\omega) \otimes \cdots \otimes u(\omega)}_{k \text{ times}} \, \mathbb{P}(\mathrm{d}\omega) \in X^{(k)}. \tag{1.20}$$

Note that (1.15) and (1.18) give, with Jensen's inequality and the convexity of the norm $\|\cdot\|_V \to \mathbb{R}$, the bound

$$\|\mathcal{M}^k u\|_{X^{(k)}} = \|\mathbb{E} u^{(k)}\|_{X^{(k)}} \leq \mathbb{E}\|u^{(k)}\|_{X^{(k)}} = \mathbb{E}\|u\|^k_X = \|u\|^k_{L^k(\Omega;X)}. \tag{1.21}$$

Deterministic equation for statistical moments

We now consider the operator equation $Au = f$, where $f \in L^k(\Omega; V')$ is given with $k \geq 1$. Since $A^{-1} \colon V' \to V$ is continuous, we obtain, using

(1.16), (1.14) and (1.21), that $u \in L^k(\Omega; V)$, and that we have the *a priori* estimate

$$\|\mathcal{M}^k u\|_{V^{(k)}} \leq \|u\|^k_{L^k(\Omega;V)} \leq C_A^k \|f\|^k_{L^k(\Omega;V')}. \tag{1.22}$$

Remark 1.2. One example of a probability measure $\mathbb{P}$ on X is a Gaussian measure; we refer to, *e.g.*, Vakhania, Tarieladze and Chobanyan (1987) and Ledoux and Talagrand (1991) for general probability measures over Banach spaces X and, in particular, to Bogachev (1998) and Janson (1997) for a general exposition of Gaussian measures on function spaces.

Since $A^{-1} : V' \to V$ in (1.11) is bijective, by (1.12) and (1.13), it induces a measure $\widetilde{\mathbb{P}} := A^{-1}_{\#}\mathbb{P}$ on the space V of solutions to (1.11). If $\mathbb{P}$ is Gaussian over V' and A in (1.11) is linear, then $\widetilde{\mathbb{P}}$ is Gaussian over V by Theorem C.18.

We recall that a Gaussian measure is completely determined by its mean and covariance, and hence only $\mathcal{M}^k u$ for $k = 1, 2$ are of interest in this case.

We now consider the tensor product operator $A^{(k)} = A \otimes \cdots \otimes A$ (k times). This operator maps $V^{(k)}$ to $(V')^{(k)}$. For $v \in V$ and $g := Av$, we obtain that $A^{(k)} v \otimes \cdots \otimes v = g \otimes \cdots \otimes g$. Consider a random field $u \in L^k(\Omega; V)$ and let $f := Au \in L^k(\Omega; V')$. Then the tensor product $u^{(k)} = u \otimes \cdots \otimes u$ (k times) belongs to the space $L^1(\Omega; V^{(k)})$, and we obtain from (1.17) with $B = A^{(k)}$ that the k-point correlations $u^{(k)}$ satisfy $\mathbb{P}$-a.s. the tensor equation

$$A^{(k)} u^{(k)} = f^{(k)},$$

where $f^{(k)} \in L^1(\Omega; (V')^{(k)})$. Now (1.17) implies *for linear and deterministic operators* A that the k-point correlation functions of the random solutions, *i.e.*, the expectations $\mathcal{M}^k u = \mathbb{E}[u^{(k)}]$, are solutions of the tensorized equations

$$A^{(k)} \mathcal{M}^k u = \mathcal{M}^k f. \tag{1.23}$$

In the case $k = 1$ this is just the equation $A\mathbb{E}u = \mathbb{E}f$ for the mean field. Note that this equation provides a way to compute the moments $\mathcal{M}^k u$ of the random solution in a deterministic fashion, for example by Galerkin discretization. As mentioned before, with the operator A acting on function spaces X, Y in the domain $D \subset \mathbb{R}^d$, the tensor equation (1.23) will require discretization in D^k, the k-fold Cartesian product of D with itself. Using tensor products of, for instance, finite element spaces in D, we find for $k > 1$ a reduction of efficiency in terms of accuracy versus number of degrees of freedom due to the 'curse of dimensionality'. This mandates sparse tensor product constructions.

We will investigate the numerical approximation of the tensor equation (1.23) in Section 1.4. The direct approximation of (1.23) by, for example, Galerkin discretization is an alternative to the Monte Carlo approximation of the moments which will be considered in Section 1.3.

In the deterministic approach, explicit knowledge of all joint probability densities of f (*i.e.*, the law of f) with respect to the probability measure $\mathbb{P}$ is not required to determine the order-k statistics of the random solution u from order-k statistics of f.

Remark 1.3. For nonlinear operator equations, associated systems of moment equations require a closure hypothesis, which must be additionally imposed and verified. For the linear operator equation (1.11), however, a closure hypothesis is not necessary, as (1.23) holds.

For solvability of (1.23), we consider the tensor product operator $A_1 \otimes A_2 \otimes \cdots \otimes A_k$ for operators $A_i \in \mathcal{L}(V_i, V_i')$, $i = 1, \ldots, k$.

Proposition 1.4. For integer $k > 1$, let V_i, $i = 1, \ldots, k$ be Hilbert spaces with duals V_i', and let $A_i \in \mathcal{L}(V_i, V_i')$ be injective and satisfy a Gårding inequality, *i.e.*, there are compact $T_i \in \mathcal{L}(V_i, V_i')$ and $\alpha_i > 0$ such that

$$\forall v \in V_i : \quad \big\langle (A_i + T_i)\, v, v \big\rangle \geq \alpha_i \, \|v\|_{V_i}^2, \tag{1.24}$$

where $\langle \cdot, \cdot \rangle$ denotes the $V_i' \times V_i$ duality pairing.

Then the product operator $\mathcal{A} = A_1 \otimes A_2 \otimes \cdots \otimes A_k \in \mathcal{L}(\mathcal{V}, \mathcal{V}')$, where $\mathcal{V} = V_1 \otimes V_2 \otimes \cdots \otimes V_k$ and $\mathcal{V}' = (V_1 \otimes V_2 \otimes \cdots \otimes V_k)' \cong V_1' \otimes V_2' \otimes \cdots \otimes V_k'$, is injective, and for every $f \in \mathcal{V}'$, the problem $\mathcal{A}u = f$ admits a unique solution u with

$$\|u\|_{\mathcal{V}} \leq C \, \|f\|_{\mathcal{V}'}.$$

Proof. The injectivity and the Gårding inequality (1.24) imply the bounded invertibility of A_i for each i. This implies the bounded invertibility of $\mathcal{A}$ on $\mathcal{V}' \to \mathcal{V}$ since we can write

$$\mathcal{A} = (A_1 \otimes I^{k-1}) \circ (I \otimes A_2 \otimes I^{(k-2)}) \circ \cdots \circ (I^{(k-1)} \otimes A_k),$$

where $I^{(j)}$ denotes the j-fold tensor product of the identity operator on the appropriate V_i. Note that each factor in the composition is invertible. □

To apply this result to (1.23), we require the special case

$$A^{(k)} := \underbrace{A \otimes A \otimes \cdots \otimes A}_{k \text{ times}} \in \mathcal{L}(V^{(k)}, (V')^{(k)}) = \mathcal{L}(V^{(k)}, (V^{(k)})'). \tag{1.25}$$

Theorem 1.5. If A in (1.11) satisfies (1.12) and (1.13), then for every $k > 1$ the operator $A^{(k)} \in \mathcal{L}(V^{(k)}, (V')^{(k)})$ is injective on $V^{(k)}$, and for every $f \in L^k(\Omega; V')$, the equation

$$A^{(k)} Z = \mathcal{M}^k f \tag{1.26}$$

has a unique solution $Z \in V^{(k)}$.

This solution coincides with the kth moment $\mathcal{M}^k u$ of the random field in (1.20):

$$Z = \mathcal{M}^k u.$$

Proof. By (1.21), the assumption $f \in L^k(\Omega; V')$ ensures that $\mathcal{M}^k f \in (V')^{(k)}$. The unique solvability of (1.26) follows immediately from Proposition 1.4 and the assumptions (1.12) and (1.13). The identity $Z = \mathcal{M}^k u$ follows from (1.23) and the uniqueness of the solution of (1.26). □

Regularity

The numerical analysis of approximation schemes for (1.26) will require a regularity theory for (1.26). To this end we introduce a smoothness scale $(Y_s)_{s\geq 0}$ for the data f with $Y_0 = V'$ and $Y_s \subset Y_t$ for $s > t$. We assume that we have a corresponding scale $(X_s)_{s\geq 0}$ of 'smoothness spaces' for the solutions with $X_0 = V$ and $X_s \subset X_t$ for $s > t$, so that $A^{-1}\colon Y_s \to X_s$ is continuous.

When D is a smooth closed manifold of dimension d embedded into Euclidean space $\mathbb{R}^{d+1}$, we choose $Y_s = H^{-\varrho/2+s}(D)$ and $X_s = H^{\varrho/2+s}(D)$. The case of differential operators with smooth coefficients in a manifold D with smooth boundary is also covered within this framework by the choices $Y_s = H^{-\varrho/2+s}(D)$ and $X_s = \widetilde{H}^{\varrho/2} \cap H^{\varrho/2+s}(D)$. Note that in other cases (a pseudodifferential operator on a manifold with boundary, or a differential operator on a domain with non-smooth boundary), the spaces X_s can be chosen as weighted Sobolev spaces which contain functions that are singular at the boundary.

Theorem 1.6. Assume (1.12) and (1.13), and that there is an $s^* > 0$ such that $A^{-1} : Y_s \to X_s$ is continuous for $0 \leq s \leq s^*$. Then we have for all $k \geq 1$ and for $0 \leq s \leq s^*$ some constant $C(k, s)$ such that

$$\|\mathcal{M}^k u\|_{X_s^{(k)}} \leq C\|\mathcal{M}^k f\|_{Y_s^{(k)}} = C\|f\|^k_{L^k(\Omega;Y_s)}. \tag{1.27}$$

Proof. If (1.12) and (1.13) hold, then the operator $A^{(k)}$ is invertible, and

$$\mathcal{M}^k u = (A^{(k)})^{-1}\mathcal{M}^k f = (A^{-1})^{(k)}\mathcal{M}^k f.$$

Since

$$\|A^{-1}f\|_{X_s} \leq C_s \|f\|_{Y_s}, \quad 0 \leq s \leq s^*,$$

it follows that

$$\|\mathcal{M}^k u\|_{X_s^{(k)}} = \|(A^{-1})^{(k)}\mathcal{M}^k f\|_{X_s^{(k)}} \leq C_s^k \|\mathcal{M}^k f\|_{Y_s^{(k)}}, \quad 0 \leq s \leq s^*. \qquad \square$$

1.2. Finite element discretization

In order to obtain a finite-dimensional problem, we need to discretize in both Ω and D. For D we will use a nested family of finite element spaces $V_\ell \subset V$, $\ell = 0, 1, \ldots$.

Nested finite element spaces

The Galerkin approximation of (1.11) is based on a sequence $\{V_\ell\}_{\ell=0}^\infty$ of subspaces of V of dimension $N_\ell = \dim V_\ell < \infty$ which are dense in V, *i.e.*, $V = \overline{\bigcup_{\ell\geq 0} V_\ell}$, and nested, *i.e.*,

$$V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_\ell \subset V_{\ell+1} \subset \cdots \subset V. \tag{1.28}$$

We assume that for functions u in the smoothness spaces X_s with $s \geq 0$ we have the asymptotic *approximation rate*

$$\inf_{v\in V_\ell} \|u - v\|_V \leq C N_\ell^{-s/d} \|u\|_{X_s}. \tag{1.29}$$

Finite elements with uniform mesh refinement

We will now describe examples for the subspaces V_ℓ which satisfy the assumptions of Section 1.2. We briefly sketch the construction of finite element spaces which are only continuous across element boundaries; see Braess (2007), Brenner and Scott (2002) and Ciarlet (1978) for presentations of the mathematical foundations of finite element methods. These elements are suitable for operators of order $\varrho < 3$. Throughout, we denote by $\mathcal{P}_p(K)$ the linear space of polynomials of total degree $\leq p$ on a set K.

Let us first consider the case of a bounded polyhedron $D \subset \mathbb{R}^d$. Let $\mathcal{T}_0$ be a regular partition of D into simplices K. Let $\{\mathcal{T}_\ell\}_{\ell=0}^\infty$ be the sequence of regular partitions of D obtained from $\mathcal{T}_0$ by uniform subdivision: for example, if $d = 2$, we bisect all edges of the triangulation $\mathcal{T}_\ell$ and obtain a new, regular partition of the domain D into possibly curved triangles which belong to finitely many congruency classes. We set

$$V_\ell = S^p(D, \mathcal{T}_\ell) = \{u \in C^0(\overline{D})\,;\; u|_K \in \mathcal{P}_p(K)\ \forall K \in \mathcal{T}_\ell\}$$

and let $h_\ell = \max\{\operatorname{diam}(K)\,;\; K \in \mathcal{T}_\ell\}$. Then $N_\ell = \dim V_\ell = \mathcal{O}(h_\ell^{-d})$ as $\ell \to \infty$. With $V = \tilde{H}^{\varrho/2}(D)$ and $X_s = H^{\varrho/2+s}(D)$, standard finite element approximation results imply that (1.29) holds for $s \in [0, p + 1 - \varrho/2]$, *i.e.*,

$$\inf_{v\in V_\ell} \|u - v\|_V \leq C N_\ell^{-s/d} \|u\|_{X_s}.$$

For the case when D is the boundary $D = \partial\mathcal{D}$ of a polyhedron $\mathcal{D} \subset \mathbb{R}^{d+1}$ we define finite element spaces on D in the same way as above, but now in local coordinates on D, and obtain the same convergence rates (see, *e.g.*, Sauter and Schwab (2010)): for a d-dimensional domain $D \subset \mathbb{R}^d$ with a smooth boundary we can first divide D into pieces D_J, which can be mapped to a simplex S by smooth mappings $\Phi_J : D_J \to S$ (which must be C^0-compatible where two pieces $D_J, D_{J'}$ touch). Then we can define on D finite element functions which on D_J are of the form $g \circ \Phi_J$, where g is a polynomial.

For a d-dimensional smooth surface $D \subset \mathbb{R}^{d+1}$ we can similarly divide D into pieces which can be mapped to simplices in $\mathbb{R}^d$, and again define finite elements using these mappings.

Finite element wavelet basis for V_ℓ

To facilitate the accurate numerical approximation of moments of order $k \geq 2$ of the random solution and for the efficient numerical solution of the partial differential equations, we use a hierarchical basis for the nested finite element (FE) spaces $V_0 \subset \cdots \subset V_L$.

To this end, we start with a basis $\{\psi_j^0\}_{j=1,\ldots,N_0}$ for the finite element space V_0 on the coarsest triangulation. We represent on the finer meshes $\mathcal{T}_\ell$ the corresponding FE spaces V_ℓ, with $\ell > 0$ as a direct sum $V_\ell = V_{\ell-1} \oplus W_\ell$. Since the subspaces are nested and finite-dimensional, this is possible with a suitable space W_ℓ for *any hierarchy* of FE spaces. We assume, in addition, that we are explicitly given basis functions $\{\psi_j^\ell\}_{j=1,\ldots,M_\ell}$ of W_ℓ. Iterating with respect to ℓ, we have that $V_L = V_0 \oplus W_1 \oplus \cdots \oplus W_L$, and $\{\psi_j^\ell\,;\ \ell = 0,\ldots,L,\ j = 1,\ldots,M_\ell\}$ is a hierarchical basis for V_L, where $M_0 := N_0$.

(W1) *Hierarchical basis.* $V_L = \operatorname{span}\{\psi_j^\ell\,;\ 1 \leq j \leq M_L,\ 0 \leq \ell \leq L\}$.

Let us define $N_\ell := \dim V_\ell$ and $N_{-1} := 0$; then we have $M_\ell := N_\ell - N_{\ell-1}$ for $\ell = 0, 1, 2, \ldots, L$.

The hierarchical basis property (W1) is in principle sufficient for the formulation and implementation of the sparse MC–Galerkin method and the deterministic sparse Galerkin method. In order to obtain algorithms of log-linear complexity for integrodifferential equations, impose on the hierarchical basis the additional properties (W2)–(W5) of a *wavelet basis*. This will allow us to perform matrix compression for non-local operators, and to obtain optimal preconditioning for the iterative linear system solver.

(W2) *Small support.* $\operatorname{diam} \operatorname{supp}(\psi_j^\ell) = \mathcal{O}(2^{-\ell})$.

(W3) *Energy norm stability.* There is a constant $C_B > 0$ independent of $L \in \mathbb{N} \cup \{\infty\}$, such that, for all $L \in \mathbb{N} \cup \{\infty\}$ and all

$$v^L = \sum_{\ell=0}^{L} \sum_{j=1}^{M_\ell} v_j^\ell\, \psi_j^\ell(x) \in V_L,$$

we have

$$C_B^{-1} \sum_{\ell=0}^{L} \sum_{j=1}^{M_\ell} |v_j^\ell|^2 \leq \|v^L\|_V^2 \leq C_B \sum_{\ell=0}^{L} \sum_{j=1}^{M_\ell} |v_j^\ell|^2. \tag{1.30}$$

Here, in the case $L = \infty$ it is understood that $V_L = V$.

(W4) Wavelets ψ_j^ℓ with $\ell \geq \ell_0$ have *vanishing moments* up to order $p_0 \geq p - \varrho$

$$\int \psi_j^\ell(x)\, x^\alpha \, \mathrm{d}x = 0, \quad 0 \leq |\alpha| \leq p_0, \tag{1.31}$$

except possibly for wavelets where the closure of the support intersects the boundary ∂D or the boundaries of the coarsest mesh. In the case of mapped finite elements we require the vanishing moments for the polynomial function $\psi_j^\ell \circ \Phi_J^{-1}$.

(W4) *Decay of coefficients for 'smooth' functions in X_s.* There exists $C > 0$ independent of L such that, for every $u \in X_s$ and every L,

$$\sum_{\ell=0}^{L} \sum_{j=1}^{M_\ell} |u_j^\ell|^2 \, 2^{2\ell s} \leq C L^\nu \|u\|_{X_s}^2, \quad \nu = \begin{cases} 0 & \text{for } 0 \leq s < p + 1 - \varrho/2, \\ 1 & \text{for } s = p + 1 - \varrho/2. \end{cases} \tag{1.32}$$

By property (W3), wavelets constitute Riesz bases: every function $u \in V$ has a unique wavelet expansion $u = \sum_{\ell=0}^{\infty} \sum_{j=1}^{M_\ell} u_j^\ell \, \psi_j^\ell$.

We define the projection $P_L : V \to V_L$ by truncating this wavelet expansion of u at level L, *i.e.*,

$$P_L u := \sum_{\ell=0}^{L} \sum_{j=1}^{M_\ell} u_j^\ell \, \psi_j^\ell. \tag{1.33}$$

Because of the stability (W3) and the approximation property (1.29), we obtain immediately that the wavelet projection P_L is quasi-optimal: with (1.29), for $0 \leq s \leq s^*$ and $u \in X_s$,

$$\|u - P_L u\|_V \lesssim N_L^{-s/d} \|u\|_{X_s}. \tag{1.34}$$

We remark in passing that the appearance of the factor $1/d$ in the convergence rate s/d in (1.34), when expressed in terms of N_L, the total number of degrees of freedom, indicates a reduction of the convergence rate as the dimension d of the computational domain increases. This reduction of the convergence rate with increasing dimension is commonly referred to as the 'curse of dimensionality'; as long as $d = 1, 2, 3$, this is not severe and, in fact, shared by almost all discretizations. If the dimension of the computational domain increases, however, this reduction becomes a severe obstacle to the construction of efficient discretizations. In the context of stochastic and parametric PDEs, the dimension of the computational domain can, in principle, become arbitrarily large, as we shall next explain.

Full and sparse tensor product spaces
To compute an approximation for

$$\mathcal{M}^k u \in V^{(k)} := \underbrace{V \otimes \cdots \otimes V}_{k \text{ times}}$$

we need a suitable finite-dimensional subspace of $V^{(k)}$. The simplest choice is the tensor product space $V_L \otimes \cdots \otimes V_L = V_L^{(k)}$. However, this full tensor product space has dimension

$$\dim(V_L^{(k)}) = N_L^k = (\dim(V_L))^k, \tag{1.35}$$

which is not practical for $k > 1$. A reduction in cost is possible by *sparse tensor products* of V_L. The k-fold *sparse tensor product space* $\widehat{V}_L^{(k)}$ is defined by

$$\widehat{V}_L^{(k)} = \sum_{\substack{\underline{\ell} \in \mathbb{N}_0^k \\ |\underline{\ell}| \le L}} V_{\ell_1} \otimes \cdots \otimes V_{\ell_k}, \tag{1.36}$$

where we denote by $\underline{\ell}$ the vector $(\ell_1, \ldots, \ell_k) \in \mathbb{N}_0^k$ and its length by $|\underline{\ell}| = \ell_1 + \cdots + \ell_k$. The sum in (1.36) is not direct in general. However, since the V_ℓ are finite-dimensional, we can write $\widehat{V}_L^{(k)}$ as a direct sum in terms of the complement spaces W_l:

$$\widehat{V}_L^{(k)} = \bigoplus_{\substack{\underline{\ell} \in \mathbb{N}_0^k \\ |\underline{\ell}| \le L}} W_{\ell_1} \otimes \cdots \otimes W_{\ell_k}. \tag{1.37}$$

If a hierarchical basis of the subspaces V_ℓ (*i.e.*, satisfying hypothesis (W1)) is available, we can define a sparse tensor quasi-interpolation operator $\widehat{P}_L^{(k)} : V^{(k)} \to \widehat{V}_L^{(k)}$ by a suitable truncation of the tensor product wavelet expansion: for every $x_1, \ldots, x_k \in D$,

$$(\widehat{P}_L^{(k)} v)(x) := \sum_{\substack{0 \le \ell_1 + \cdots + \ell_k \le L \\ 1 \le j_\nu \le M_{\ell_\nu}, \nu = 1, \ldots, k}} v_{j_1 \cdots j_k}^{\ell_1 \cdots \ell_k} \, \psi_{j_1}^{\ell_1}(x_1) \cdots \psi_{j_k}^{\ell_k}(x_k). \tag{1.38}$$

If a hierarchical basis is not explicitly available, we can still express $\widehat{P}_L^{(k)}$ in terms of the projections $Q_\ell := P_\ell - P_{\ell-1}$ for $\ell = 0, 1, \ldots$, and with the convention $P_{-1} := 0$ as

$$\widehat{P}_L^{(k)} = \sum_{0 \le \ell_1 + \cdots + \ell_k \le L} Q_{\ell_1} \otimes \cdots \otimes Q_{\ell_k}. \tag{1.39}$$

We also note that the dimension of $\widehat{V}_L^{(k)}$ is

$$\widehat{N}_L = \dim(\widehat{V}_L^{(k)}) = O(N_L (\log_2 N_L)^{k-1}), \tag{1.40}$$

that is, it is a log-linear function of the number N_L of the degrees of freedom used for approximation of the first moment. Given that the sparse tensor product space $\widehat{V}_L^{(k)}$ is substantially coarser, one wonders whether its approximation properties are substantially worse than that of the full tensor product space $V_L^{(k)}$. The basis for the use of the sparse tensor product spaces $\widehat{V}_L^{(k)}$ is the next result, which indicates that $\widehat{V}_L^{(k)}$ achieves, up to logarithmic terms, the same asymptotic rate of convergence, in terms of powers of the mesh width, as the full tensor product space. The approximation property of sparse grid spaces $\widehat{V}_L^{(k)}$ was established, for example, in Schwab and Todor (2003*b*, Proposition 4.2), Griebel, Oswald and Schiekofer (1999), von Petersdorff and Schwab (2004) and Todor (2009).

Proposition 1.7.

$$\inf_{v\in\widehat{V}_L^{(k)}} \|U-v\|_{V^{(k)}} \le C(k) \begin{cases} N_L^{-s/d}\|U\|_{X_s^{(k)}} & \text{if } 0 \le s < p+1-\varrho/2, \\ N_L^{-s/d} L^{\nu(k)} \|U\|_{X_s^{(k)}} & \text{if } s = p+1-\varrho/2. \end{cases} \tag{1.41}$$

Here, the exponent $\nu(k) = (k-1)/2$ is best possible on account of the V-orthogonality of the V best approximation.

Remark 1.8. The exponent $\nu(k)$ of the logarithmic terms in the sparse tensor approximation rates stated in Proposition 1.7 is best possible for the approximation in the sparse tensor product spaces $V^{(k)}$ given the regularity $U \in X_s^{(k)}$. In general, these logarithmic terms in the convergence estimate are unavoidable. Removal of all logarithmic terms in the convergence rate estimate as well as in the dimension estimate of $\widehat{V}_L^{(k)}$ is possible *only if* either (a) the norm $\|\circ\|_{V^{(k)}}$ on the left-hand side of (1.41) is weakened, or if (b) the norm $X_s^{(k)}$ on the right-hand side of (1.41) is strengthened. For example, in the context of sparse tensor FEM for the Laplacian in $(0,1)^d$, it was shown by von Petersdorff and Schwab (2004) and Bungartz and Griebel (2004) that all logarithmic terms can be removed; this is due to the observation that the $H^1((0,1)^d)$ norm is strictly weaker than the corresponding tensorized norm $H^1(0,1)^{(d)}$ which appears in the error bound (1.41) in the case of d-point correlations of a random field taking values in $H_0^1(0,1)$.

The same effect allows us to slightly coarsen the sparse tensor product space $\widehat{V}_L^{(k)}$. This was exploited, for example, by Bungartz and Griebel (2004) and Todor (2009).

The error bound (1.41) is for the best approximation of $U \in X_s^{(k)}$ from $\widehat{V}_L^{(k)}$. To achieve the exponent $\nu(k) = (k-1)/2$ in (1.41) for a sparse tensor quasi-interpolant such as (1.38), the multi-level basis ψ_j^ℓ of V must be V-orthogonal between successive levels ℓ. This V-orthogonality of the

multi-level basis can be achieved in $V \subset H^1(D)$, for example, by using so-called *spline prewavelets*.

Let us also remark that it is even possible to construct $L^2(D)$ *orthonormal* piecewise polynomial wavelet bases satisfying (W1)–(W5). We refer to Donovan, Geronimo and Hardin (1996) for details.

The stability property (W3) implies the following result (see, *e.g.*, von Petersdorff and Schwab (2004)).

Lemma 1.9. (on the sparse tensor quasi-interpolant $\widehat{P}_L^{(k)}$) Assume (W1)–(W5) and that the component spaces V_ℓ of $\widehat{V}_L^{(k)}$ are V-orthogonal between scales and have the approximation property (1.29). Then the sparse tensor projection $\widehat{P}_L^{(k)}$ is stable: there exists $C > 0$ (depending on k but independent of L) such that, for all for $U \in V^{(k)}$,

$$\|\widehat{P}_L^{(k)} U\|_{V^{(k)}} \leq C \, \|U\|_{V^{(k)}}. \tag{1.42}$$

For $U \in X_s^{(k)}$ and $0 \leq s \leq s^*$, if the basis functions ψ_j^ℓ satisfy (W1)–(W5) and *are V-orthogonal between different levels of mesh refinement*, we obtain quasi-optimal convergence of the sparse tensor quasi-interpolant $\widehat{P}_L^{(k)} U$ in (1.38):

$$\|U - \widehat{P}_L^{(k)} U\|_{V^{(k)}} \leq C(k) N_L^{-s/d} (\log N_L)^{(k-1)/2} \|U\|_{X_s^{(k)}}. \tag{1.43}$$

Remark 1.10. The convergence rate (1.43) of the approximation $\widehat{P}_L^{(k)} U$ from the sparse tensor subspace is, up to logarithmic terms, equal to the rate obtained for the best approximation of the mean field, *i.e.*, in the case $k = 1$. We observe, however, that the *regularity of U required to achieve this convergence rate* is quite high: the function U must belong to an anisotropic smoothness class $X_s^{(k)}$ which, in the context of ordinary Sobolev spaces, is a space of functions whose (*weak*) *mixed derivatives of order s* belong to V. Evidently, this *mixed smoothness regularity requirement becomes stronger as the number k of moments increases.* By Theorem 1.6, the k-point correlations $\mathcal{M}^k u$ of the random solution u *naturally* satisfy such regularity.

Galerkin discretization

We first consider the discretization of the problem $Au(\omega) = f(\omega)$ for a single realization ω, bearing in mind that in the Monte Carlo method this problem will have to be approximately solved for many realizations of $\omega \in \Omega$.

The Galerkin discretization of (1.11) reads: find $u_L(\omega) \in V_L$ such that

$$\langle v_L, Au_L(\omega) \rangle = \langle v_L, f(\omega) \rangle \quad \forall v_L \in V_L, \quad \mathbb{P}\text{-a.e. } \omega \in \Omega, \tag{1.44}$$

where '$\mathbb{P}$-a.e.' stands for '$\mathbb{P}$ almost everywhere'. It is well known that the

injectivity (1.13) of A, the Gårding inequality (1.12) and the density in V of the subspace sequence $\{V_\ell\}_{\ell=0}^{\infty}$ imply that there exists $L_0 > 0$ such that, for $L \geq L_0$, problem (1.44) admits a unique solution $u_L(\omega)$. Furthermore, we have the *uniform inf-sup condition* (see, *e.g.*, Hildebrandt and Wienholtz (1964)): there exists a discretization level L_0 and a *stability constant* $\gamma > 0$ such that, for all $L \geq L_0$,

$$\inf_{0\neq u\in V_L} \sup_{0\neq v\in V_L} \frac{\langle Au, v\rangle}{\|u\|_V \|v\|_V} \geq \frac{1}{\gamma} > 0. \tag{1.45}$$

The inf-sup condition (1.45) implies quasi-optimality of the approximations $u_L(\omega)$ for $L \geq L_0$ (see, *e.g.*, Babuška (1970/71)): there exist $C > 0$ and $L_0 > 0$ such that

$$\forall L \geq L_0 : \quad \|u(\omega) - u_L(\omega)\|_V \leq C \inf_{v\in V_L} \|u(\omega) - v\|_V \quad \mathbb{P}\text{-a.e. } \omega \in \Omega. \tag{1.46}$$

From (1.46) and (1.29), we obtain the asymptotic error estimate: define $\sigma := \min\{s^*, p + 1 - \varrho/2\}$. Then there exists $C > 0$ such that for $0 < s \leq \sigma$

$$\forall L \geq L_0 : \quad \|u(\omega) - u_L(\omega)\|_V \leq C N_L^{-s/d} \|u\|_{X_s} \quad \mathbb{P}\text{-a.e. } \omega \in \Omega. \tag{1.47}$$

1.3. Sparse tensor Monte Carlo Galerkin FEM

We next review basic convergence results of the Monte Carlo method for the approximation of expectations of random variables taking values in a separable Hilbert space. As our exposition aims at the solution of operator equations with stochastic data, we shall first consider the MC method without discretization of the operator equation, and show convergence estimates of the *statistical error* incurred by the MC sampling. Subsequently, we turn to the Galerkin approximation of the operator equation and, in particular, the sparse tensor approximation of the two- and k-point correlation functions of the random solution.

Monte Carlo error for continuous problems

For a random variable Y, let $Y_1(\omega), \ldots, Y_M(\omega)$ denote $M \in \mathbb{N}$ *copies* of Y, *i.e.*, the Y_i are random variables which are mutually independent and identically distributed to $Y(\omega)$ on the same common probability space $(\Omega, \Sigma, \mathbb{P})$. Then the arithmetic average $\overline{Y}^M(\omega)$,

$$\overline{Y}^M(\omega) := \frac{1}{M}\big(Y_1(\omega) + \cdots + Y_M(\omega)\big),$$

is a random variable on $(\Omega, \Sigma, \mathbb{P})$ as well.

The simplest approach to the numerical solution of (1.11) for $f \in L^1(\Omega; V')$ is MC simulation. Let us first consider the situation without discretization of V. We generate M draws $f(\omega_j)$, $j = 1, 2, \ldots, M$, of $f(\omega)$ and find the

solutions $u(\omega_j) \in V$ of the problems

$$Au(\omega_j) = f(\omega_j), \quad j = 1, \dots, M. \tag{1.48}$$

We then approximate the kth moment $\mathcal{M}^k u$ with the sample mean $\bar{E}^M[u^{(k)}]$ of $u(\omega_j) \otimes \cdots \otimes u(\omega_j)$:

$$\bar{E}^M[u^{(k)}] := \overline{u \otimes \cdots \otimes u}^M = \frac{1}{M} \sum_{j=1}^{M} u(\omega_j) \otimes \cdots \otimes u(\omega_j). \tag{1.49}$$

It is well known that the Monte Carlo error decreases as $M^{-1/2}$ in a probabilistic sense *provided the variance of* $u^{(k)}$ *exists*. By (1.18), this is the case for $u \in L^{2k}(\Omega; V)$. We have the following convergence estimate.

Theorem 1.11. Let $k \geq 1$ and assume that in the operator equation (1.11) $f \in L^{2k}(\Omega; V')$. Then, for any $M \in \mathbb{N}$ of samples for the MC estimator (1.49), we have the error bound

$$\|\mathcal{M}^k u - \bar{E}^M[u^{(k)}]\|_{L^2(\Omega; V^{(k)})} \leq M^{-1/2} \big(C_A \|f\|_{L^{2k}(\Omega; V')}\big)^k. \tag{1.50}$$

Proof. We observe that $f \in L^{2k}(\Omega; V')$ implies with (1.22) that $u^{(k)} \in L^2(\Omega; V^{(k)})$. For $i = 1, \dots, M$ we denote by $\widehat{u}_i(\omega)$ the M i.i.d. copies of the random variable $u(\omega) = A^{-1} f(\omega)$, which corresponds to the M many MC samples $\widehat{u}_i = A^{-1} \widehat{f}_i$.

Using that the $\widehat{u}_i$ are independent and identically distributed, we infer that, for each value of i, $\widehat{u}_i(\omega) \in L^{2k}(\Omega; V)$. Therefore

$$\begin{aligned}
\|\mathbb{E}[u^{(k)}] - \bar{E}^M[u^{(k)}]\|^2_{L^2(\Omega; V^{(k)})} &= \mathbb{E}\big[\|\mathbb{E}[u^{(k)}] - \bar{E}^M[u^{(k)}]\|^2_{V^{(k)}}\big] \\
&= \mathbb{E}\left[\left\|\mathbb{E}[u^{(k)}] - \frac{1}{M}\sum_{i=1}^{M} \widehat{u}_i^{(k)}\right\|^2_{V^{(k)}}\right] \\
&= \mathbb{E}\left[\left\langle \mathbb{E}[u^{(k)}] - \frac{1}{M}\sum_{i=1}^{M} \widehat{u}_i^{(k)}, \mathbb{E}[u^{(k)}] - \frac{1}{M}\sum_{j=1}^{M} \widehat{u}_j^{(k)} \right\rangle\right] \\
&= \frac{1}{M^2} \sum_{i,j=1}^{M} \mathbb{E}\big[\big\langle \mathbb{E}[u^{(k)}] - \widehat{u}_i^{(k)}, \mathbb{E}[u^{(k)}] - \widehat{u}_j^{(k)} \big\rangle\big] \\
&= \frac{1}{M^2} \sum_{i=1}^{M} \mathbb{E}\big[\|\mathbb{E}[u^{(k)}] - \widehat{u}_i^{(k)}\|^2_{V^{(k)}}\big] \qquad (\widehat{u}_i(\omega) \text{ independent}) \\
&= \frac{1}{M} \mathbb{E}\big[\|u^{(k)} - \mathbb{E}[u^{(k)}]\|^2_{V^{(k)}}\big] \qquad (\widehat{u}_i(\omega) \text{ identically distributed}) \\
&= \frac{1}{M} \mathbb{E}\big[\big\langle u^{(k)} - \mathbb{E}[u^{(k)}], u^{(k)} - \mathbb{E}[u^{(k)}] \big\rangle\big]
\end{aligned}$$

$$= \frac{1}{M}\{\mathbb{E}[\langle u^{(k)} - \mathbb{E}[u^{(k)}], \mathbb{E}[u^{(k)}]\rangle] + \mathbb{E}[\langle u^{(k)} - \mathbb{E}[u^{(k)}], u^{(k)}\rangle]\}$$
$$= \frac{1}{M}\mathbb{E}[\|u^{(k)}\|^2_{V^{(k)}}] - \frac{1}{M}\|\mathbb{E}[u^{(k)}]\|^2_{V^{(k)}}$$
$$\leq M^{-1}\|u^{(k)}\|^2_{L^2(\Omega;V^{(k)})} = M^{-1}\|u\|^{2k}_{L^{2k}(\Omega;V)}.$$

Taking square roots on both sides completes the proof. □

The previous theorem required that $u^{(k)} \in L^2(\Omega; V^{(k)})$ or (equivalently by (1.18)) that $u \in L^{2k}(\Omega; V)$ (resp. $f \in L^{2k}(\Omega; V')$) in order to obtain the convergence rate $M^{-1/2}$ of the MC estimates (1.49), in $L^2(\Omega; v)$.

In the case of weaker summability of u, the next estimate shows that the MC method converges in $L^1(\Omega; V^{(k)})$ and at a rate that is possibly lower than $1/2$, as determined by the summability of u. We only state the result here and refer to von Petersdorff and Schwab (2006) for the proof.

Theorem 1.12. Let $k \geq 1$. Assume that $f \in L^{\alpha k}(\Omega; V')$ for some $\alpha \in (1, 2]$. For $M \geq 1$ samples we define the sample mean $\bar{E}^M[u^{(k)}]$ as in (1.49). Then there exists C such that, for every $M \geq 1$ and every $0 < \epsilon < 1$,

$$\mathbb{P}\left(\|\mathcal{M}^k u - \bar{E}^M[u^{(k)}]\|_{V^{(k)}} \leq C\frac{\|f\|^k_{L^{\alpha k}(\Omega;V')}}{\epsilon^{1/\alpha}M^{1-1/\alpha}}\right) \geq 1 - \epsilon. \tag{1.51}$$

The previous results show that one can obtain a rate of up to $M^{-1/2}$ in a probabilistic sense for the Monte Carlo method. Convergence rates beyond $1/2$ are not possible, in general, by the MC method, as is shown by the central limit theorem; in this sense, the rate $1/2$ is sharp.

So far, we have obtained the convergence rate $1/2$ of the MC method essentially in $L^1(\Omega, V^{(k)})$ and in $L^2(\Omega, V^{(k)})$. A $\mathbb{P}$-a.s convergence estimate of the MC method can be obtained using the separability of the Hilbert space of realizations and the law of the iterated logarithm; see, *e.g.*, Strassen (1964) and Ledoux and Talagrand (1991, Chapter 8) and the references therein for the vector-valued case.

Lemma 1.13. Assume that H is a separable Hilbert space and that $X \in L^2(\Omega; H)$. Then, with probability 1,

$$\limsup_{M\to\infty} \frac{\|\overline{X}^M - \mathbb{E}(X)\|_H}{(2M^{-1}\log\log M)^{1/2}} \leq \|X - \mathbb{E}(X)\|_{L^2(\Omega;H)}. \tag{1.52}$$

For the proof, we refer to von Petersdorff and Schwab (2006). Applying Lemma 1.13 to $X = u^{(k)} = u\otimes\cdots\otimes u$ and with $V^{(k)}$ in place of H gives (with C_A as in (1.14)) $\|u \otimes \cdots \otimes u\|_{L^2(\Omega;V^{(k)})} = \|u\|^k_{L^{2k}(\Omega;V)} \leq C_A^{2k}\|f\|^k_{L^{2k}(\Omega;V')}$, whence the following result.

Theorem 1.14. Let $f \in L^{2k}(\Omega; V')$. Then, with probability 1,

$$\limsup_{M\to\infty} \frac{\|\mathcal{M}^k u - \bar{E}^M[u^{(k)}]\|_{V^{(k)}}}{(2M^{-1}\log\log M)^{1/2}} \leq C(k)\|f\|^k_{L^{2k}(\Omega;V')}. \tag{1.53}$$

Sparse Monte Carlo Galerkin moment estimation

We now use Galerkin discretization with the subspaces $V_L \subset V$ to solve (1.48) approximately and to obtain, for each draw of the load function, Galerkin approximations $u_L(\omega_j)$. The resulting sample mean approximation of the kth moment $\mathcal{M}^k u$ equals

$$\bar{E}^M[u_L^{(k)}] := \frac{1}{M}\sum_{j=1}^{M} u_L(\omega_j) \otimes \cdots \otimes u_L(\omega_j). \tag{1.54}$$

This yields a first MC estimation for the k-point correlation function of $\mathcal{M}^k u$. The complexity of forming (1.54) is, however, prohibitive: to form in (1.54) the k-fold tensor product of the Galerkin approximations for the M data samples, one needs $\mathcal{O}(N_L^k)$ memory and $\mathcal{O}(MN_L^k)$ operations to compute this mean, which implies loss of linear complexity for $k > 1$. Therefore we propose using the sparse approximation

$$\widehat{E}^M[u^{(k)}] := \widehat{P}_L^{(k)}\bar{E}^M[u^{(k)}] = \bar{E}^M[\widehat{P}_L^{(k)}u^{(k)}], \tag{1.55}$$

which requires $\mathcal{O}(N_L(\log N_L)^{k-1})$ memory and operations.

To compute $\bar{E}^{M,L}[\widehat{P}_L^{(k)}\mathcal{M}^k u]$ we proceed as follows. First, we generate M data samples $f(\omega_j)$, $j = 1, \ldots, M$ and the corresponding Galerkin approximations $u_L(\omega_j) \in V_L$ as in (1.44).

Choosing a wavelet basis of V_L that satisfies (W1) ((W2)–(W5) are *not* required at this stage), by (1.33), $u_L(\omega_j)$ can then be represented as

$$u_L(\omega_j) = \sum_{\ell=0}^{L}\sum_{k=1}^{M_\ell} u_k^\ell(\omega_j)\psi_k^\ell, \tag{1.56}$$

with $u_k^\ell(\omega_j) = \langle u_L(\omega_j), \widetilde{\psi}_k^\ell\rangle$, where $\{\widetilde{\psi}_k^\ell\}_{\ell,k}$ denotes the dual wavelet basis to $\{\psi_k^\ell\}_{\ell,k}$ (Cohen 2003). Based on the representation (1.56), we can compute the *sparse tensor product MC estimate* of $\mathcal{M}^k u$ with the projection operators $\widehat{P}_L^{(k)}$ in (1.38) as follows:

$$\widehat{E}^M[u_L^{(k)}] = \frac{1}{M}\sum_{j=1}^{M} \widehat{P}_L^{(k)}\big[u_L(\omega_j) \otimes \cdots \otimes u_L(\omega_j)\big] \in \widehat{V}_L^{(k)}. \tag{1.57}$$

This quantity can be computed in $\mathcal{O}(MN_L(\log N_L)^{k-1})$ operations since, for each data sample $f(\omega_j)$, $j = 1, \ldots, M$, the projection $\widehat{P}_L$ onto the sparse tensor product space $\widehat{V}_L^{(k)}$ of the Galerkin approximation $u_L(\omega_j)$ is

given by $\widehat{P}_L^{(k)}[u_L(\omega_j)\otimes\cdots\otimes u_L(\omega_j)]$. This projection can be computed in $\mathcal{O}(N_L(\log_2 N_L)^{k-1})$ operations as follows. For each j we first compute $u_L(\omega_j)$ in the wavelet basis and then form $\widehat{P}_L^{(k)}[u_L(\omega_j)\otimes\cdots\otimes u_L(\omega_j)]$ using the formula

$$\widehat{P}_L^{(k)}(v\otimes\cdots\otimes v)=\sum_{\substack{0\le\ell_1+\cdots+\ell_k\le L\\ 1\le j_\nu\le M_{\ell_\nu},\nu=1,\dots,k}} v_{j_1}^{l_1}\cdots v_{j_k}^{l_k}\psi_{j_1}^{l_1}\cdots\psi_{j_k}^{l_k}. \tag{1.58}$$

The following result addresses the convergence of the sparse MC–Galerkin approximation of $\mathcal{M}^k u$. Recall $\sigma:=\min\{s^*,p+1-\varrho/2\}$ with s^* as in Theorem 1.6.

Theorem 1.15. Assume that $f\in L^k(\Omega;Y_s)\cap L^{\alpha k}(\Omega;V')$ for some $\alpha\in(1,2]$ and some $s\in(0,\sigma]$. Then there exists $C(k)>0$ such that, for all $M\ge 1$, $L\ge L_0$ and all $0<\epsilon<1$,

$$\mathbb{P}\big(\|\mathcal{M}^k u-\widehat{E}^{M,L}[\mathcal{M}^k u]\|_{V^{(k)}}<\lambda\big)\ge 1-\epsilon$$

with

$$\lambda=C(k)\big[N_L^{-s/d}(\log N_L)^{(k-1)/2}\|f\|^k_{L^k(\Omega;Y_s)}+\epsilon^{-1/\alpha}M^{-(1-\alpha^{-1})}\|f\|^k_{L^{\alpha k}(\Omega;V')}\big].$$

Proof. We estimate

$$\begin{aligned}
&\|\widehat{E}^{M,L}[\mathcal{M}^k u]-\mathcal{M}^k u\|_{V^{(k)}}\\
&\quad=\left\|\frac{1}{M}\sum_{j=1}^M \widehat{P}_L^{(k)}[u_L(\omega_j)\otimes\cdots\otimes u_L(\omega_j)]-\mathbb{E}(u\otimes\cdots\otimes u)\right\|_{V^{(k)}}\\
&\quad\le\|\widehat{P}_L^{(k)}[u_L(\omega_j)\otimes\cdots\otimes u_L(\omega_j)-u(\omega_j)\otimes\cdots\otimes u(\omega_j)]\|_{V^{(k)}}\\
&\qquad+\left\|\frac{1}{M}\sum_{j=1}^M \widehat{P}_L^{(k)}[u(\omega_j)\otimes\cdots\otimes u(\omega_j)]-\mathbb{E}(\widehat{P}_L^{(k)}[u\otimes\cdots\otimes u])\right\|_{V^{(k)}}\\
&\qquad+\|(I-\widehat{P}_L^{(k)})\mathcal{M}^k u\|_{V^{(k)}}.
\end{aligned}$$

The last term is estimated with (1.43), Theorem 1.6, for $0\le s\le s^*$ by

$$\|(I-\widehat{P}_L^{(k)})\mathcal{M}^k u\|_{V^{(k)}}\le C(k)N_L^{-s/d}(\log N_L)^{(k-1)/2}\|\mathcal{M}^k f\|_{Y_s^{(k)}}.$$

For the first term, we use (1.42) and (1.47) with a tensor product argument. For the second term, the statistical error, by (1.42) it suffices to bound

$$\left\|\mathbb{E}([u\otimes\cdots\otimes u])-\frac{1}{M}\sum_{j=1}^M[u(\omega_j)\otimes\cdots\otimes u(\omega_j)]\right\|_{V^{(k)}}=\|\mathcal{M}^k u-\bar{E}^M[u^{(k)}]\|_{V^{(k)}},$$

which was estimated in Theorem 1.12. □

Remark 1.16. All results in this section also hold in the case of a stochastic operator $A(\omega)$. Specifically, let X now denote the space of bounded linear mappings $V \to V'$. Assume that $A\colon \Omega \to X$ is measurable (with respect to Borel sets of X) and that there exists $C, \alpha > 0$ and a compact $T\colon V \to V'$ such that

$$\|A(\omega)\|_V \le C \qquad \text{almost everywhere,} \tag{1.59}$$

$$\langle (A(\omega)+T)u, u\rangle \ge \alpha \|u\|_V^2 \quad \text{almost everywhere.} \tag{1.60}$$

Let $k \ge 1$. Then $f \in L^k(\Omega; V')$ implies $u = A^{-1}f \in L^k(\Omega; V)$ and $\mathcal{M}^k u \in V^{(k)}$. Also $f \in L^k(\Omega; Y_s)$ implies $u = A^{-1}f \in L^k(\Omega; X_s)$ and $\mathcal{M}^k u \in X_s^{(k)}$. All proofs on the convergence of MC methods in this section still apply to that case. However, as we shall explain below, substantial computational efficiency for MCM can be gained by coupling the multi-level structure (1.28) of the Galerkin discretizations with a level-dependent sample size.

1.4. Deterministic Galerkin approximation of moments

Sparse Galerkin approximation of $\mathcal{M}^k u$

We now describe and analyse the deterministic computation of the k-point correlation function $\mathcal{M}^k u$ of the random solution u by Galerkin discretization (1.26). If we use in the Galerkin discretization the full tensor product space $V_L^{(k)}$, the inf-sup condition of the discrete operator on $V_L^{(k)}$ follows directly for $L \ge L_0$ from the discrete inf-sup condition (1.45) of the 'mean-field' operator A by a tensor product argument.

The (anisotropic) regularity estimate for the kth moment $\mathcal{M}^k u$,

$$\|\mathcal{M}^k u\|_{X_s^{(k)}} \le C_{k,s}\, \|\mathcal{M}^k f\|_{Y_s^{(k)}},\ 0 \le s \le s^*, \quad k \ge 1, \tag{1.61}$$

which was shown in Theorem 1.6, then allows us to obtain convergence rates. However, this 'full tensor product Galerkin' approach is prohibitively expensive: with N_L degrees of freedom in the physical domain D, it requires the set-up and solution of a linear system with N_L^k unknowns. We reduce this complexity by using in place of $V_L^{(k)}$ the sparse tensor product space $\widehat{V}_L^{(k)}$. The sparse Galerkin approximation $\widehat{Z}_L$ of $\mathcal{M}^k u$ is then obtained as follows:

$$\text{find}\ \ \widehat{Z}_L \in \widehat{V}_L^{(k)} \quad \text{such that}\ \ \langle A^{(k)}\, \widehat{Z}_L, v\rangle = \langle \mathcal{M}^k f, v\rangle \quad \forall v \in \widehat{V}_L^{(k)}. \tag{1.62}$$

We first consider the case where the operator A is coercive, *i.e.*, (1.12) holds with $T = 0$. Then $A^{(k)}\colon V^{(k)} \to (V')^{(k)}$ is also coercive, and the stability of the Galerkin method with $\widehat{V}_L^{(k)}$ follows directly from $\widehat{V}_L^{(k)} \subset V^{(k)}$.

In the case of $T \ne 0$ the stability of the Galerkin FEM on the sparse tensor product space $\widehat{V}_L^{(k)}$ is not obvious: we know that $(A+T)\otimes\cdots\otimes(A+T)$

is coercive for sufficiently fine meshes (*i.e.*, for sufficiently large L), but $(A+T)\otimes\cdots\otimes(A+T) - A\otimes\cdots\otimes A$ is not compact. Therefore we require some additional assumptions.

We assume that (1.12) holds with the additional requirement that $T'\colon V \to V'$ is smoothing with respect to the scale of spaces X_s, Y_s, and we also assume that the adjoint operator $A'\colon V \to V'$ satisfies a regularity property: we assume that there exists $\delta > 0$ such that

$$T'\colon V = X_0 \to Y_\delta \text{ is continuous,} \tag{1.63}$$

$$(A')^{-1}\colon Y_\delta \to X_\delta \text{ is continuous.} \tag{1.64}$$

Due to the indefiniteness of A we have to modify the sparse grid space: Let $L_0 \geq 0$ and $L \geq L_0$. We define a space $\widehat{V}^{(k)}_{L,L_0}$ with $\widehat{V}^{(k)}_{L} \subset \widehat{V}^{(k)}_{L,L_0} \subset \widehat{V}^{(k)}_{L+(k-1)L_0}$ as follows.

Definition 1.17. Let $S^1_{L,L_0} := \{0,\dots,L\}$. For $k \geq 2$, let S^k_{L,L_0} be the set of indices $l \in \mathbb{N}^k_0$ satisfying the following conditions:

$$l_1 + \cdots + l_k \leq L + (k-1)L_0, \tag{1.65}$$

$$(l_{i_1},\dots,l_{i_{k-1}}) \in S^{k-1}_{L,L_0} \text{ if } i_1,\dots,i_{k-1} \text{ are different indices in } \{1,\dots,k\}. \tag{1.66}$$

Then we define

$$\widehat{V}^{(k)}_{L,L_0} := \sum_{l\in S^k_{L,L_0}} W^{l_1}\otimes\cdots\otimes W^{l_k}. \tag{1.67}$$

Let $J_{L_0} := \{0,1,\dots,L_0\}$. Then the index set S^k_{L,L_0} has the following subsets:

$$J^k_{L_0},\quad J^{k-1}_{L_0}\times S^1_{L,L_0},\quad J^{k-2}_{L_0}\times S^2_{L,L_0},\quad \dots,\quad J_{L_0}\times S^{k-1}_{L,L_0}.$$

Therefore, $\widehat{V}^{(k)}_{L,L_0}$ contains the following subspaces:

$$V^{(k)}_{L_0},\quad V^{(k-1)}_{L_0}\otimes\widehat{V}^{(1)}_{L,L_0},\quad V^{(k-2)}_{L_0}\otimes\widehat{V}^{(2)}_{L,L_0},\quad \dots,\quad V_{L_0}\otimes\widehat{V}^{(k-1)}_{L,L_0}. \tag{1.68}$$

To achieve stability of sparse tensor discretizations in the presence of possible indefiniteness of the operator $A(\omega)$ in (1.12), we introduce a certain fixed $L_0 > 0$ of mesh refinement and consider the sequence of spaces $\widehat{V}^{(k)}_{L,L_0}$ with L tending to infinity. Since

$$\widehat{V}^{(k)}_{L} \subset \widehat{V}^{(k)}_{L,L_0} \subset \widehat{V}^{(k)}_{L+(k-1)L_0},$$

we see that $\dim \widehat{V}^{(k)}_{L,L_0}$ grows with the same rate as $\dim \widehat{V}^{(k)}_{L}$ as $L\to\infty$. We then have the following discrete stability property.

Theorem 1.18. Assume that A and T satisfy (1.12), (1.63) and (1.64). Then there exists $L_0 \in \mathbb{N}$ and $\gamma > 0$ such that, for all $L \geq L_0$,

$$\inf_{0\neq u\in \widehat{V}^{(k)}_{L,L_0}} \sup_{0\neq v\in \widehat{V}^{(k)}_{L,L_0}} \frac{\langle A^{(k)}u, v\rangle}{\|u\|_{V^{(k)}}\,\|v\|_{V^{(k)}}} \geq \frac{1}{\gamma} > 0. \tag{1.69}$$

In the positive definite case $T = 0$, this holds with $L_0 = 0$, whereas in the indefinite case, $L_0 > 0$ is necessary in general.

For the proof, we refer to von Petersdorff and Schwab (2006). As is by now classical (*e.g.*, Babuška (1970/71)), the discrete inf-sup condition (1.69) implies quasi-optimal convergence and therefore the convergence rate is given by the rate of best approximation. The following result from von Petersdorff and Schwab (2006) makes this precise.

Theorem 1.19. Assume (1.12) and (1.13).

(a) Let $f \in L^k(\Omega; V')$. Then with $L_0 \geq 0$ as in Theorem 1.18 (in particular, $L_0 = 0$ is admissible when $T = 0$ in (1.12)) such that, for all $L \geq L_0$, the sparse Galerkin approximation $\widehat{Z}_L \in \widehat{V}^{(k)}_{L,L_0}$ of $\mathcal{M}^k u$ is uniquely defined and converges quasi-optimally, *i.e.*, there exists $C > 0$ such that, for all $L \geq L_0$,

$$\|\mathcal{M}^k u - \widehat{Z}_L\|_{V^{(k)}} \leq C \inf_{v\in \widehat{V}^{(k)}_{L,L_0}} \|\mathcal{M}^k u - v\|_{V^{(k)}} \to 0 \quad \text{as } L \to \infty.$$

(b) Assume that $f \in L^k(\Omega; Y_s)$ and the approximation property (1.29). Then, for $0 \leq s \leq \sigma := \min\{s^*, p + 1 - \varrho/2\}$,

$$\|\mathcal{M}^k u - \widehat{Z}_L\|_{V^{(k)}} \leq C(k) N_L^{-s/d} (\log N_L)^{(k-1)/2} \|f\|^k_{Y_s}. \tag{1.70}$$

Matrix compression

When A is a differential operator, the number of non-zero entries in the stiffness matrix for the standard FEM basis is $\mathcal{O}(N)$ due to the local support assumption (W2) on the basis ψ^ℓ_j. This implies that we can compute a matrix–vector product arising typically in iterative solvers with $\mathcal{O}(N)$ operations. In the case of an integral or pseudodifferential equation, the operator A is non-local and all entries of the stiffness matrix are non-vanishing, in general. Then the cost of a matrix–vector product is $\mathcal{O}(N^2)$, which implies a loss of linear complexity of the algorithm. For boundary integral operators, it is well known that one can improve the complexity to $\mathcal{O}(N(\log N)^c)$ by using matrix compression techniques. Several approaches to this end are available: either fast multipole methods (*e.g.*, Beatson and Greengard (1997) and the references therein), multiresolution methods (such as wavelets; see, *e.g.*, Schneider (1998), Harbrecht (2001), Dahmen, Harbrecht and Schneider (2006), Dahmen (1997)), or low-rank matrix approximation techniques (*e.g.*, Bebendorf and Hackbusch (2003)).

These matrix compression methods have in common that they reduce complexity of the matrix–vector multiplication from $\mathcal{O}(N^2)$ to $\mathcal{O}(N(logN)^b)$ for some (small) non-negative number b. This complexity reduction comes, however, at the expense of being realized only *approximately*. We will elaborate on the effect of matrix compression on the accuracy of sparse tensor Galerkin approximations in this and in the following section.

In the *compression step*, we replace most of the entries $A_{JJ'}$ of the stiffness matrix $\boldsymbol{A}^L$ with zeros, yielding an approximate stiffness matrix $\widetilde{\boldsymbol{A}}^L$. The stiffness matrix $\boldsymbol{A}^L$ and its compressed variant $\widetilde{\boldsymbol{A}}^L$ induce mappings from V_L to $(V_L)'$, which we denote by $\mathcal{A}_L$ and $\widetilde{\mathcal{A}}_L$, respectively. We will require $\mathcal{A}_L$ and $\widetilde{\mathcal{A}}_L$ to be close in the following sense: for certain values $s, s' \in [0, \sigma]$ with $\sigma = p + 1 - \varrho/2$ and for $u \in X_s$, $v \in X_{s'}$, we have

$$\left|\left\langle (\mathcal{A}_L - \widetilde{\mathcal{A}}_L)\, P_L u, P_L v \right\rangle\right| \leq c(s,s') N_L^{-(s+s')/d} (\log N_L)^{q(s,s')}\, \|u\|_{X_s}\, \|v\|_{X_{s'}} \tag{1.71}$$

with $c(s,s') > 0$ and $q(s,s') \geq 0$ independent of L. The following result collects some properties of the corresponding approximate solutions.

Proposition 1.20. Assume (1.12) and (1.13).

(a) If (1.71) holds for $(s,s') = (0,0)$ with $q(0,0) = 0$ and $c(0,0)$ sufficiently small, then there is an $L_0 > 0$ such that, for every $L \geq L_0$, $(\widetilde{\mathcal{A}}^L)^{-1}$ exists and is uniformly bounded, *i.e.*

$$\forall L \geq L_0 : \quad \|(\widetilde{\mathcal{A}}_L)^{-1}\|_{(V_L)' \to V_L} \leq C \tag{1.72}$$

for some C independent of L.

(b) If, in addition to the assumptions in (a), (1.71) holds with $(s,s') = (\sigma, 0)$, then

$$\|(A^{-1} - (\widetilde{\mathcal{A}}_L)^{-1})\, f\|_V \leq C N_L^{-\sigma/d} (\log N_L)^{q(\sigma,0)} \|f\|_{Y_\sigma}. \tag{1.73}$$

(c) Let $g \in V'$ be such that the solution $\varphi \in V$ of $A'\varphi = g$ belongs to X_σ. If, in addition to the assumptions in (a) and (b), (1.71) also holds for $(s,s') = (0,\sigma)$ and for $(s,s') = (\sigma,\sigma)$, then

$$\left|\left\langle g, (A^{-1} - (\widetilde{\mathcal{A}}_L)^{-1})\, f \right\rangle\right| \leq C N_L^{-2\sigma/d} \cdot (\log N_L)^{\max\{q(0,\sigma)+q(\sigma,0), q(\sigma,\sigma)\}} \tag{1.74}$$

where $C = C(f,g)$.

Proof. **(a)** The Gårding inequality (1.12), the injectivity (1.13) and the density in V of the subspace sequence $\{V^L\}_L$ imply the discrete inf-sup condition (1.45).

Using (1.71) with $v_L \in V_L$ and $(s,s') = (0,0)$, we obtain with (1.45)

$$\|\widetilde{\mathcal{A}}_L\, v_L\|_{(V_L)'} \geq \|A v_L\|_{(V_L)'} - \|(A - \widetilde{\mathcal{A}}_L) v_L\|_{(V_L)'} \geq c_s^{-1} \|v_L\|_V - C c(0,0) \|v_L\|_V.$$

This implies that for $c(0,0) < 1/(2C\gamma)$ there is an $L_0 > 0$ such that, for all $L \geq L_0$,

$$\|v_L\|_V \leq \frac{c_s}{2} \, \|\widetilde{\mathcal{A}}_L \, v_L\|_{(V_L)'} \quad \forall v_L \in V_L, \tag{1.75}$$

whence we obtain (1.72).

(b) Let $f \in Y_\sigma$ and $u = A^{-1} f$, $\widetilde{u}_L = (\widetilde{\mathcal{A}}_L)^{-1} f$ for $L \geq L_0$. We have

$$\|u - \widetilde{u}_L\|_V \leq \|u - P_L u\|_V + \|P_L u - \widetilde{u}_L\|_V.$$

Using (1.45) and $\langle \widetilde{\mathcal{A}}_L \, u_L, v_L \rangle = \langle Au, v_L \rangle$ for all $v_L \in V_L$, we get

$$\|P_L u - \widetilde{u}_L\|_V \leq C \, \|\widetilde{\mathcal{A}}_L (P_L u - \widetilde{u}_L)\|_{(V_L)'} = C \|\widetilde{\mathcal{A}}_L \, P_L u - Au\|_{(V_L)'},$$

which yields the error estimate

$$\|u - \widetilde{u}_L\|_V \leq \|u - P_L u\|_V + C \, \|A(u - P_L u)\|_{(V_L)'} + C \, \|(A - \widetilde{\mathcal{A}}_L) \, P_L u\|_{(V_L)'}. \tag{1.76}$$

Here, the first two terms are estimated using the V-stability (W3) and (1.33) of the wavelet basis, which imply

$$\|u - P_L u\|_V \leq C \inf_{v \in V_L} \|u - v\|_V \leq C (N_L)^{-\sigma/d} \|u\|_{X_\sigma}, \tag{1.77}$$

and the continuity $A : V \to V'$. The third term in (1.76) is estimated with (1.71) for $(s, s') = (\sigma, 0)$ and $P_L v_L = v_L$ for all $v_L \in V_L$:

$$\left|\langle (A - \widetilde{\mathcal{A}}_L) \, P_L u, v_L \rangle\right| \lesssim c(\sigma,0) N_L^{-\sigma/d} (\log N_L)^{q(\sigma,0)} \|u\|_{X_\sigma} \, \|v\|_V. \tag{1.78}$$

(c) To show (1.74), we let $\varphi^L := P_L \varphi$ for $\varphi = (A')^{-1} g \in X_\sigma$ and $u = A^{-1} f$, $\widetilde{u}_L = (\widetilde{\mathcal{A}}_L)^{-1} f$ for $L \geq L_0$. Then

$$\left|\langle g, u - \widetilde{u}^L \rangle\right| = \left|\langle \varphi, A(u - \widetilde{u}^L) \rangle\right| \leq \left|\langle A(u - \widetilde{u}_L), \varphi - \varphi^L \rangle\right| + \left|\langle A(u - \widetilde{u}_L), \varphi^L \rangle\right|.$$

We estimate the first term by $C \|u - \widetilde{u}_L\|_V \, \|\varphi - P^L \varphi\|_V$, which implies the bound (1.74) using (1.73) and (1.77). The second term is bounded as follows:

$$\begin{aligned} \langle A(u - \widetilde{u}_L), \varphi^L \rangle &= \langle (\widetilde{\mathcal{A}}_L - A) \, \widetilde{u}_L, \varphi^L \rangle \\ &= \langle (\widetilde{\mathcal{A}}_L - A)(\widetilde{u}_L - P_L u), P_L \varphi \rangle + \langle (\widetilde{\mathcal{A}}_L - A) \, P_L u, P_L \varphi \rangle. \end{aligned}$$

Here we estimate the second term by (1.71) with $(s, s') = (\sigma, \sigma)$. For the first term, we use (1.71) with $(s, s') = (0, \sigma)$ to obtain

$$\begin{aligned} &\left|\langle (\widetilde{\mathcal{A}}_L - A) \, P_L (\widetilde{u}_L - P_L u), P_L \varphi \rangle\right| \\ &\qquad \lesssim N_L^{-\sigma/d} (\log N_L)^{q(0,\sigma)} \|\widetilde{u}_L - P_L u\|_V \, \|\varphi\|_{X_\sigma} \\ &\qquad \lesssim N_L^{-\sigma/d} (\log N_L)^{q(0,\sigma)} (\|\widetilde{u}^L - u\|_V + \|u - P_L u\|_V) \, \|\varphi\|_{X_\sigma}. \end{aligned}$$

Using (1.73) and (1.77), we complete the proof. □

Wavelet compression

We next describe how to obtain an approximate stiffness matrix $\widetilde{\boldsymbol{A}}^L$, which on the one hand has $\mathcal{O}(N_L(\log N_L)^a)$ non-zero entries (out of N_L^2), and on the other hand satisfies the consistency condition (1.71). Here we assume that the operator A is given in terms of its Schwartz kernel $k(x,y)$ by

$$(A\varphi)(x) = \int_{y\in\Gamma} k(x,y)\,\varphi(y)\,\mathrm{d}S(y) \tag{1.79}$$

for $\varphi \in C_0^\infty(\Gamma)$, where $\Gamma = \partial D$ and $k(x,z)$ satisfies the Calderón–Zygmund estimates

$$|D_x^\alpha\, D_y^\beta\, k(x,y)| \le C_{\alpha\beta}|x-y|^{-(d+\varrho+|\alpha|+|\beta|)}, \quad x \ne y \in \Gamma. \tag{1.80}$$

In the following, we combine the indices (ℓ, j) into a multi-index $J = (\ell, j)$ to simplify notation, and write $\psi_J, \psi_{J'}$, *etc.*

Due to the vanishing moment property (1.31) of the basis $\{\psi_J\}$, the entries $A^L_{JJ'} = \langle A\psi_J, \psi_{J'}\rangle$ of the moment matrix $\boldsymbol{A}^L$ with respect to the basis $\{\psi_J\}$ show fast decay (Schneider 1998, Dahmen *et al.* 2006). Let $S_J = \mathrm{supp}(\psi_J)$, $S_{J'} = \mathrm{supp}(\psi_{J'})$. Then we have the following decay estimate for the matrix entries $A_{JJ'}$ (see Schneider (1998, Lemma 8.2.1) and Dahmen *et al.* (2006)).

Proposition 1.21. If the wavelets $\psi_J, \psi_{J'}$ satisfy the moment condition (1.31) and A satisfies (1.79) and (1.80), then

$$\big|\langle A\psi_J, \psi_{J'}\rangle\big| \le C\,\mathrm{dist}(S_J, S_{J'})^{-\gamma}\, 2^{-\gamma(\ell+\ell')/2}, \tag{1.81}$$

where $\gamma := \varrho + d + 2 + 2(p^* + 1) > 0$.

This can be exploited to truncate $\boldsymbol{A}^L$ to obtain a sparse matrix $\widetilde{\boldsymbol{A}}^L$ with at most $\mathcal{O}(N_L(\log N_L)^2)$ non-zero entries and such that (1.71) is true with $c(0,0)$ as small as desired, independent of L, $q(0,0) = 0$, and $q(0,\sigma) = q(\sigma,0) \le \frac{3}{2}$, $q(\sigma,\sigma) \le 3$; see von Petersdorff and Schwab (1996), Schneider (1998), Harbrecht (2001) and Dahmen *et al.* (2006), for example. The number of non-zero entries, $\mathrm{nnz}((\widetilde{\boldsymbol{A}}^L_{\ell,\ell'})$, in the block $\widetilde{\boldsymbol{A}}^L_{\ell,\ell'}$ of the compressed Galerkin stiffness matrix $\widetilde{\boldsymbol{A}}^L$ is bounded by

$$\mathrm{nnz}(\widetilde{\boldsymbol{A}}^L_{\ell,\ell'}) \le C(\min(\ell,\ell')+1)^d\, 2^{d\max(\ell,\ell')}. \tag{1.82}$$

Remark 1.22. For integral operators A an alternative approach for the efficient computation of matrix–vector products with the stiffness matrix $\boldsymbol{A}^L$ is given by the cluster of fast multipole approximation. For these approximations, one additionally assumes for the operator (1.79) that the kernel $k(x,z)$ is analytic for $x \ne y$, and the size of its domain of analyticity is proportional to $|x-y|$. Then one can replace $k(x,y)$ in (1.79) for $|x-y|$ sufficiently large by a cluster of fast multipole approximation with degenerate kernels which are obtained by either truncated multipole expansions

or polynomial interpolants of order $\log N_L$, allowing us to apply the block $\widetilde{\boldsymbol{A}}^L_{\ell,\ell'}$ to a vector in at most

$$C(\log N_L)^d\, 2^{d\max(\ell,\ell')}, \quad 0 \le \ell, \ell' \le L, \tag{1.83}$$

operations. See Schmidlin, Lage and Schwab (2003) for details on this work estimate.

Error analysis for sparse Galerkin with matrix compression

Based on the compressed stiffness matrix $\widetilde{\boldsymbol{A}}^L$ and the corresponding operator $\widetilde{\mathcal{A}}_L : V_L \to (V_L)'$ induced by it, we define the sparse tensor product approximation of $\mathcal{M}^k u$ with matrix compression analogous to (1.62) as follows: find $\widetilde{Z}^k_L \in \widehat{V}^{(k)}_{L,L_0}$ such that, for all $v \in \widehat{V}^{(k)}_{L,L_0}$,

$$\langle \widetilde{\mathcal{A}}^{(k)}_L \widetilde{Z}^k_L, v\rangle = \langle \mathcal{M}^k f, v\rangle. \tag{1.84}$$

We prove bounds for the error $\widetilde{Z}^k_L - \mathcal{M}^k u$.

Lemma 1.23. Assume (1.12) and (1.13), and that the spaces V_L as in Example 1.2 admit a hierarchical basis $\{\psi^\ell_j\}_{\ell\ge 0}$ satisfying (W1)–(W5). Assume further that the operator $\widetilde{\mathcal{A}}_L$ in (1.84) satisfies the consistency estimate (1.71) for $s = s' = 0$, $q(0,0) = 0$, and with sufficiently small $c(0,0)$.

Then there exists $L_0 > 0$ such that, for all $L \ge L_0$, the kth-moment problem with matrix compression, (1.84), admits a unique solution and we have the error estimate

$$\begin{aligned} &\|\mathcal{M}^k u - \widetilde{Z}^k_L\|_{V^{(k)}} \\ &\quad \le C \inf_{v\in\widehat{V}^{(k)}_L} \left\{ \|\mathcal{M}^k u - v\|_{V^{(k)}} + \sup_{0\ne w\in\widehat{V}^{(k)}_L} \frac{|\langle (A^{(k)}_L - \widetilde{\mathcal{A}}^{(k)}_L)v, w\rangle|}{\|w\|_{V^{(k)}}} \right\}. \end{aligned} \tag{1.85}$$

Proof. We show unique solvability of (1.84) for sufficiently large L. By Theorem 1.18 we have that (1.69) holds. To show unique solvability of (1.84), we write for $k \ge 3$

$$\begin{aligned} A^{(k)} - \widetilde{\mathcal{A}}^{(k)}_L &= (A - \widetilde{\mathcal{A}}_L)\otimes A^{(k-1)} + \widetilde{\mathcal{A}}_L \otimes \big(A^{(k-1)} - \widetilde{\mathcal{A}}^{(k-1)}_L\big) \\ &= (A - \widetilde{\mathcal{A}}_L)\otimes A^{(k-1)} + \widetilde{\mathcal{A}}_L \otimes (A - \widetilde{\mathcal{A}}_L)\otimes A^{(k-2)} \\ &\quad + \widetilde{\mathcal{A}}^{(2)}_L \otimes \big(A^{(k-2)} - \widetilde{\mathcal{A}}^{(k-2)}_L\big), \end{aligned}$$

and obtain, after iteration,

$$\begin{aligned} A^{(k)} - \widetilde{\mathcal{A}}^{(k)}_L &= (A - \widetilde{\mathcal{A}}_L)\otimes A^{(k-1)} + \sum_{\nu=1}^{k-2} \widetilde{\mathcal{A}}^{(\nu)}_L \otimes (A - \widetilde{\mathcal{A}}_L)\otimes A^{(k-\nu-1)} \\ &\quad + \widetilde{\mathcal{A}}^{(k-1)}_L \otimes (A - \widetilde{\mathcal{A}}_L) \end{aligned} \tag{1.86}$$

(where the sum is omitted if $k = 2$). We get from (1.69) that for any $u \in \widehat{V}_L^{(k)}$ there exists $v \in \widehat{V}_L^{(k)}$ such that

$$\begin{aligned}\langle \widetilde{\mathcal{A}}_L^{(k)} u, v\rangle &= \langle A^{(k)} u, v\rangle + \langle (\widetilde{\mathcal{A}}_L^{(k)} - A^{(k)})u, v\rangle \qquad (1.87)\\ &\geq \left[\gamma^{-1} - \sup_{w\in\widehat{V}_L^{(k)}} \sup_{\widetilde{w}\in\widehat{V}_L^{(k)}} \frac{\langle (\widetilde{\mathcal{A}}_L^{(k)} - A^{(k)})w, \widetilde{w}\rangle}{\langle \|w\|_{V^{(k)}} \|\widetilde{w}\|_{V^{(k)}}}\right] \|u\|_{V^{(k)}}\, \|v\|_{V^{(k)}}.\end{aligned}$$

To obtain an upper bound for the supremum, we admit $w, \widetilde{w} \in V_{L,L_0}^{(k)} \supseteq \widehat{V}_L^{(k)}$, use (1.86) and (1.71) with $s = s' = 0$ and $q(0,0) = 0$ to get

$$\|\widetilde{\mathcal{A}}_L\|_{V_L\to(V_L)'} \leq \underbrace{\|A\|_{V\to V'}}_{c_A} + c(0,0),$$

and therefore estimate for any $w, \widetilde{w} \in V_{L,L_0}^{(k)}$

$$\begin{aligned}\big|\langle \widetilde{\mathcal{A}}_L^{(k)} - A^{(k)})w, \widetilde{w}\rangle\big| &\leq c(0,0)\left[c_A^{k-1} + \left(\sum_{\nu=1}^{k-2} \big(c_A + c(0,0)\big)^{\nu} c_A^{k-\nu-1}\right)\right.\\ &\qquad \left. + \big(c_A + c(0,0)\big)^{k-1}\right] \|w\|_{V^{(k)}} \|\widetilde{w}\|_{V^{(k)}}\\ &\lesssim c(0,0)\, \|w\|_{V^{(k)}}\, \|\widetilde{w}\|_{V^{(k)}}. \qquad (1.88)\end{aligned}$$

If $c(0,0)$ is sufficiently small, this implies with (1.85) the stability of $\widetilde{\mathcal{A}}_L^{(k)}$ on $\widehat{V}_{L,L_0}^{(k)}$: there exists $L_0 > 0$ such that

$$\inf_{0\neq u\in\widehat{V}_{L,L_0}^{(k)}} \sup_{0\neq v\in\widehat{V}_{L,L_0}^{(k)}} \frac{\langle \widetilde{\mathcal{A}}_L^{(k)}\, u, v\rangle}{\|u\|_{V^{(k)}}\, \|v\|_{V^{(k)}}} \geq \frac{1}{2\gamma} > 0, \qquad (1.89)$$

for all $L \geq L_0$, and hence the unique solvability of (1.84) for these L follows.

To prove (1.85), we follow the proof of the first lemma of Strang (*e.g.*, Ciarlet (1978)). □

We now use this lemma to obtain the following convergence result.

Theorem 1.24. Assume (1.12) and (1.13), $V = H^{\varrho/2}(\Gamma)$, and that the subspaces $\{V_\ell\}_{\ell=0}^{\infty}$ are as in Example 1.2, and that in the smoothness spaces $X_s = H^{\varrho/2+s}(\Gamma)$, $s \geq 0$, the operator $A : X_s \to Y_s$ is bijective for $0 \leq s \leq s^*$ with some $s^* > 0$. Assume further that a compression strategy for the matrix $\boldsymbol{A}^L$ in the hierarchical basis $\{\psi_j^\ell\}$ satisfying (W1)–(W5) is available with (1.71) for $s' = 0$, $0 \leq s \leq \sigma = p + 1 - \varrho/2$, $q(0,0) = 0$ and

with $c(0,0)$ sufficiently small, independent of L for $L \geq L_0$. Then, with $\delta = \min\{p+1-\varrho/2, s\}/d$, $0 \leq s \leq s^*$, we have the error estimate

$$\|\mathcal{M}^k u - \widetilde{Z}_L^k\|_{V^{(k)}} \leq C(\log N_L)^{\min\{(k-1)/2, q(s,0)\}} N_L^{-\delta} \|\mathcal{M}^k f\|_{Y_s^{(k)}}. \quad (1.90)$$

Proof. We use (1.85) with the choice $v = \widehat{P}_{L,L_0}^{(k)}$ and, for $\|\mathcal{M}^k u - v\|_{V^{(k)}}$, apply the approximation result (1.6). We express the difference $A_L^{(k)} - \tilde{\mathcal{A}}_L^{(k)}$ using (1.86). Then we obtain a sum of terms, each of which can be bounded using (1.71) and the continuity of $A_L^{(k)}$ and $\tilde{\mathcal{A}}_L^{(k)}$. This yields the following error bound:

$$\|\mathcal{M}^k u - \widetilde{Z}_L^k\|_{V^{(k)}} \leq C\big[(\log N_L)^{(k-1)/2} N_L^{-\delta} + c(s,0)(\log N_L)^{q(s,0)} N_L^{-s/d}\big] \|\mathcal{M}^k u\|_{X_s^{(k)}}. \qquad \square$$

Theorem 1.24 addressed only the convergence of $\widetilde{Z}_L^k$ in the 'energy' norm $V^{(k)}$. In the applications which we have in mind, however, functionals of the solution $\mathcal{M}^k u$ are also of interest, which we assume are given in the form $\langle G, \mathcal{M}^k u\rangle$ for some $G \in (V^{(k)})'$. We approximate such functionals by $\langle G, \widetilde{Z}_L^k\rangle$.

Theorem 1.25. With all assumptions as in Theorem 1.24, and in addition that the adjoint problem

$$(A^{(k)})' \Psi = G \quad (1.91)$$

admits a solution $\Psi \in X_{s'}^{(k)}$ for some $0 < s' \leq \sigma$, and that the compression $\widetilde{\boldsymbol{A}}^L$ of the stiffness matrix $\boldsymbol{A}^L$ satisfies (1.71) with $s = s' = \sigma$, we have

$$\big|\langle G, \mathcal{M}^k u\rangle - \langle G, \widetilde{Z}_L^k\rangle\big| \leq C(\log N_L)^{\min\{k-1, q(s,s')\}} N_L^{-(\delta+\delta')} \|\mathcal{M}^k f\|_{Y_s^{(k)}},$$

where $\delta = \min\{p+1-\varrho/2, s\}/d$, $\delta' = \min\{p+1-\varrho/2, s'\}/d$.

The proof is analogous to that of Proposition 1.20(c), using the sparse approximation property (1.41) in place of (1.29).

Iterative solution of the linear system

We solve the linear system (1.84) using iterative solvers and denote the matrix of this system by $\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)}$. We will consider three different methods.

(M1) If A is self-adjoint and (1.12) holds with $T = 0$, then the matrix $\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)}$ is Hermitian positive definite, and we use the conjugate gradient algorithm which requires one matrix–vector multiplication by the matrix $\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)}$ per iteration.

(M2) If A is not necessarily self-adjoint, but satisfies (1.12) with $T = 0$, then we can use the GMRES algorithm with restarts every μ iterations. In this case $\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)} + (\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)})^H$ is positive definite. This requires two matrix–vector multiplications per iteration, one with $\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)}$ and one with $(\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)})^H$.

(M3) In the general case where (1.12) is satisfied with some operator T, we multiply the linear system by the matrix $(\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)})^H$ and can then apply the conjugate gradient algorithm. This requires one matrix–vector multiplication with $\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)}$ and one matrix–vector multiplication with $(\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)})^H$ per iteration.

In order to achieve log-linear complexity, it is essential that we never explicitly form the matrix $\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)}$. Instead, we only store the matrix $\widetilde{\boldsymbol{A}}^L$ for the mean-field problem. We can then compute a matrix–vector product with $\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)}$ (or $(\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)})^H$) by an algorithm which multiplies parts of the coefficient vector with submatrices of $\widetilde{\boldsymbol{A}}^L$: see Algorithm 5.10 in Schwab and Todor (2003*b*). This requires $O\big((\log N_L)^{kd+2k-2} N_L\big)$ operations (Schwab and Todor 2003*b*, Theorem 5.12).

Let us explain the algorithm in the case $k = 2$ and $L_0 = 0$. In this case a coefficient vector $\underline{u}$ has components $u_{jj'}^{l\,l'}$, where l, l' are the levels used for $\widehat{V}_L^{(2)}$ (*i.e.*, $l, l' \in \{0, \dots, L\}$ such that $l + l' \le L + L_0$) and

$$j \in \{1, \dots, M_l\}, \quad j' \in \{1, \dots, M_{l'}\}.$$

Let $\widetilde{\boldsymbol{A}}^{L_1}$ denote the submatrix of $\widetilde{\boldsymbol{A}}^L$ corresponding to levels $l, l' \le L_1$. We can then compute the coefficients of the vector $\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)} \underline{u}$ as follows, where we overwrite at each step the current components with the result of a matrix–vector product.

- For $l = 0, \dots, L$, $j = 1, \dots, M_l$: multiply the column vector with components $(u_{jj'}^{l\,l'})_{\substack{l'=0\cdots L-l \\ j'=0\cdots M_{l'}}}$ by the matrix $\widetilde{\boldsymbol{A}}^{L-l}$.
- For $l' = 0, \dots, L$, $j' = 1, \dots, M_l$: multiply the column vector with components $(u_{jj'}^{l\,l'})_{\substack{l=0\cdots L-l \\ j=0\cdots M_l}}$ by the matrix $\widetilde{\boldsymbol{A}}^{L-l'}$.

We now analyse the convergence of the iterative solvers. The stability assumptions for the wavelet basis, the continuous and discrete operators imply the following results about the approximate stiffness matrix $\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)}$.

Proposition 1.26. Assume the basis $\{\psi_j^\ell\}$ satisfies (1.30) with c_B independent of L.

(a) Assume that $\widetilde{\mathcal{A}}_L$ satisfies (1.71) for $q(0,0) = 0$ with sufficiently small $c(0,0)$. Then there are constants C_1, C_2 such that, for all L, the matrix $\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)}$ of the problem (1.84) satisfies

$$\|\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)}\|_2 \le C_2 < \infty. \tag{1.92}$$

(b) Assume, in addition to the assumptions of (a), that (1.12) holds with $T = 0$. Then

$$\lambda_{\min}\Big(\Big(\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)} + \big(\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)}\big)^H\Big)/2\Big) \ge C_1 > 0. \tag{1.93}$$

(c) Assume the discrete inf-sup condition (1.69) holds, and that we have for some constant C independent of L

$$\|(\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)})^{-1}\|_2 \le C\gamma. \tag{1.94}$$

Proof. Because of (1.30) the norm $\|v_L\|_{V^{(k)}}$ of $v_L \in \widehat{V}_{L,L_0}^{(k)}$ is equivalent to the 2-vector norm $\|\underline{v}\|_2$ of the coefficient vector $\underline{v}$. For (a) we obtain an arbitrarily small upper bound for the bilinear form with the operator $\mathcal{A}-\tilde{\mathcal{A}}_L$ with respect to the norm $\|v_L\|_{V^{(k)}}$. Since $\mathcal{A}$ is continuous we get an upper bound for the norm of $\tilde{\mathcal{A}}$ and therefore for the corresponding 2-matrix-norm.

In (b), the bilinear form $\langle Av, v\rangle$ corresponds to the symmetric part of the matrix, and the lower bound corresponds to the smallest eigenvalue of the matrix. Since the norm of $\mathcal{A} - \tilde{\mathcal{A}}$ is arbitrarily small we also get the lower bound for the compressed matrix.

In (c), the inf-sup condition (1.69) states that for $L \ge L_0$, the solution operator mapping $(\widehat{V}_{L,L_0}^{(k)})'$ to $\widehat{V}_{L,L_0}^{(k)}$ is bounded by γ. Because of the norm equivalence (1.30), this implies

$$\|(\widehat{\widetilde{\boldsymbol{A}}}_L^{(k)})^{-1}\|_2 \le C\gamma. \qquad \square$$

For method (M1) with a self-adjoint positive definite operator A, we have that $\lambda_{\max}/\lambda_{\min} \le C_2/C_1 =: \kappa$ is bounded independently of L, and obtain for the conjugate gradient iterates error estimates

$$\|\underline{u}^{(m)} - \underline{u}\|_2 \le c\left(1 - \frac{2}{\kappa^{1/2}+1}\right)^m.$$

For method (M2) we obtain

$$\|\underline{u}^{(m)} - \underline{u}\|_2 \le c\left(1 - \frac{1}{\kappa}\right)^m$$

for the GMRES from Eisenstat, Elman and Schultz (1983) for the restarted GMRES method (*e.g.*, with restart $\mu = 1$).

For method (M3) we use the conjugate gradient method with the matrix

$$B := (\widehat{\widehat{\boldsymbol{A}}}_L^{(k)})^H \widehat{\widehat{\boldsymbol{A}}}_L^{(k)}$$

and need the largest and smallest eigenvalue of this matrix. Now (1.94) states that $\lambda_{\min}(B) \geq (C\gamma)^{-2} > 0$. Therefore we have with $\widetilde{\kappa} := C_2^2(C\gamma)^2$ that

$$\|\underline{u}^{(m)} - \underline{u}\|_2 \leq c\left(1 - \frac{2}{\widetilde{\kappa}^{1/2}+1}\right)^m.$$

Note that the 2-vector norm $\|\underline{u}\|_2$ of the coefficient vector is equivalent to the norm $\|u\|_{V^{(k)}}$ of the corresponding function on $D \times \cdots \times D$. If we start with initial guess zero we therefore need a number M of iterations proportional to L to have an iteration error which is less than the Galerkin error. However, if we start on the coarsest mesh with initial guess zero, perform M iterations, use this as the starting value on the next-finer mesh, use M iterations, *etc.*, we can avoid this additional factor L.

Therefore we have the following complexity result.

Proposition 1.27. We can compute an approximation Z_L^k for $\mathcal{M}^k u$ using a fixed number m_0 of iterations such that

$$\|Z_L^k - \mathcal{M}^k\|_{V^{(k)}} \leq C N_L^{-s/d} L^\beta$$

where $\beta = (k-1)/2$ for a differential operator, $\beta = \min\{(k-1)/2, q(s,0)\}$ with $q(s,0)$ from (1.90). The total number of operations is $\mathcal{O}(N(\log N)^{k-1})$ in the case of a differential operator. In the case of an integral operator we need at most $\mathcal{O}(N(\log N)^{k+1})$ operations.

1.5. Examples: FEM and BEM for the Helmholtz equation

To illustrate the above concepts for an indefinite elliptic operator equation, we now consider the Helmholtz equation in a domain $G \subset \mathbb{R}^n$ with $n \geq 2$ and Lipschitz boundary $\Gamma := \partial G$. We discuss two ways to solve this equation with stochastic data. First we use the finite element approximation of the differential equation and apply our results for $D = G$, which is of dimension $d = n$.

Secondly, we consider the boundary integral formulation, which is an integral equation on the boundary Γ. We discretize this equation and then apply our results for $D = \Gamma$, which is of dimension $d = n - 1$. In this case we can also allow exterior domains G as the computation is done on the bounded manifold Γ.

To keep the presentation simple we will just consider smooth boundaries and one type of boundary condition (Dirichlet condition for finite elements, Neumann condition for boundary elements). Other boundary conditions and operators can be treated in a similar way.

Finite element methods

Let $G \subset \mathbb{R}^n$ be a bounded domain with smooth boundary. We consider the boundary value problem

$$(-\Delta - \kappa^2)u(\omega) = f(\omega) \text{ in } G, \quad u|_\Gamma = 0.$$

Here we have $V = H^1(G)$, $V' = H^{-1}(G)$, and the operator $A\colon V \to V'$ is defined by

$$\langle Au, v\rangle = \int_G \big(\nabla u \cdot \nabla v - \kappa^2 uv\big)\,\mathrm{d}x,$$

and obviously satisfies the Gårding inequality

$$\langle Au, u\rangle \geq \|u\|_V^2 - (\kappa^2 + 1)\|u\|_{L^2(G)}^2.$$

The operator $-\Delta\colon V \to V'$ has eigenvalues $0 < \lambda_1 < \lambda_2 < \cdots$ which converge to ∞. We need to assume that κ^2 is not one of the eigenvalues λ_j, so that condition (1.13) is satisfied.

The spaces for smooth data for $s > 0$ are $Y_s = H^{-1+s}(G)$; the corresponding solution spaces are $X_s = H^{1+s}(G)$. We assume that the stochastic right-hand side function $f(\omega)$ satisfies $f \in L^k(\Omega; Y_s) = L^k(\Omega; H^{-1+s}(G))$ for some $s > 0$.

The space V_L has $N_L = \mathcal{O}(h_L^{-d}) = \mathcal{O}(2^{Ld})$ degrees of freedom and the sparse tensor product space $\widehat{V}_{L,L_0}^{(k)}$ has $\mathcal{O}(N_L(\log N_L)^{(k-1)})$ degrees of freedom. For $k \geq 1$ the sparse grid Galerkin approximation $Z_L^k \in \widehat{V}_{L,L_0}^{(k)}$ for $\mathcal{M}^k u$ using V-orthogonal wavelets, using a total of $\mathcal{O}(N_L(\log N_L)^{(k-1)})$ operations, satisfies the error estimate (see Remark 1.8 regarding the exponent of the logarithmic terms)

$$\|Z_L^k - \mathcal{M}^k u\|_{V^{(k)}} \leq c h_L^p |\log h_L|^{(k-1)/2} \|f\|_{L^k(\Omega;Y_p)}^k$$

provided that $f \in L^k(\Omega; Y_p)$.

Boundary element methods

We illustrate the preceding abstract results with the boundary reduction of the stochastic Neumann problem to a boundary integral equation of the first kind.

In a bounded domain $D \subset \mathbb{R}^d$ with Lipschitz boundary $\Gamma = \partial D$, we consider

$$(\Delta + \kappa^2)U = 0 \text{ in } D \tag{1.95a}$$

with wave number $\kappa \in \mathbb{C}$ subject to Neumann boundary conditions

$$\gamma_1 U = n \cdot (\nabla U)|_\Gamma = \sigma \text{ on } \Gamma, \tag{1.95b}$$

where $\sigma \in L^k(\Omega; H^{-\frac{1}{2}}(\Gamma))$ with integer $k \geq 1$ are given random boundary

data, n is the exterior unit normal to Γ, and $H^s(\Gamma)$, $|s| \leq 1$, denotes the usual Sobolev spaces on Γ: see, *e.g.*, McLean (2000). We assume in (1.95b) that $\mathbb{P}$-a.s.

$$\langle \sigma, 1 \rangle = 0 \tag{1.96}$$

and, if $d = 2$, in (1.95a) that

$$\operatorname{diam}(D) < 1. \tag{1.97}$$

Then problem (1.95) admits a unique solution $U \in L^k(\Omega; H^1(D))$ (Schwab and Todor 2003*a*, 2003*b*). For the boundary reduction, we define for $v \in H^{1/2}(\Gamma)$ the boundary integral operator

$$(Wv)(x) = -\frac{\partial}{\partial n_x} \int_\Gamma \frac{\partial}{\partial n_y} e(x,y)\, v(y)\, \mathrm{d}s_y \tag{1.98}$$

with $e(x,y)$ denoting the fundamental solution of $-\Delta - \kappa^2$. The integral operator W is continuous (*e.g.*, McLean (2000)):

$$W : H^{\frac{1}{2}}(\Gamma) \to H^{-\frac{1}{2}}(\Gamma). \tag{1.99}$$

To reduce the stochastic Neumann problem (1.95) to a boundary integral equation with $\sigma \in L^k(\Omega; H^{-\frac{1}{2}}(\Gamma))$ satisfying (1.96) a.s., we use a representation as double-layer potential R_2:

$$U(x,\omega) = (R_2\vartheta)(x,\omega) := -\int_{y\in\Gamma} \frac{\partial}{\partial n_y} e(x,y)\, \vartheta(y,\omega)\, \mathrm{d}s_y, \tag{1.100}$$

where E_ϑ satisfies the BIE

$$W_1\, E_\vartheta = E_\sigma, \tag{1.101}$$

with the hypersingular boundary integral operator $W_1 u := Wu + \langle u, 1\rangle$.

We see that the mean field $\mathcal{M}^1 U$ can be obtained by solving the deterministic boundary integral equation (1.101). Based on the compression error analysis in Section 1.4, we obtain an approximate solution $E_\vartheta^L \in V^L$ in $\mathcal{O}(N_L(\log N_L)^2)$ operations and memory with error bound

$$\|E_\vartheta - E_\vartheta^L\|_{H^{1/2}(\Gamma)} \lesssim N_L^{-(p+1/2)} (\log N_L)^{3/2} \|\sigma\|_{L^1(\Omega; H^{p+1}(\Gamma))}.$$

To determine the variance of the random solution U, second moments of ϑ are required. To derive boundary integral equations for them, we use that by Fubini's theorem, the operator $\mathcal{M}^2$ and the layer potential R_2 commute. For (1.95) with $\sigma \in L^2(\Omega; H^{-\frac{1}{2}}(\Gamma))$, we obtain that C_ϑ satisfies the BIE

$$(W_1 \otimes W_1)\, C_\vartheta = C_\sigma \text{ in } H^{\frac{1}{2},\frac{1}{2}}(\Gamma \times \Gamma). \tag{1.102}$$

Here, the 'energy' space V equals $H^{1/2}(\Gamma)$ and $A = W_1$.

The unique solvability of the BIE (1.102) is ensured by the following result.

Proposition 1.28. If $k = 0$, the integral operator $W_1 \otimes W_1$ is coercive, *i.e.*, there exists $\gamma > 0$ such that

$$\forall C_\vartheta \in H^{\frac{1}{2},\frac{1}{2}}(\Gamma\times\Gamma): \quad \langle (W_1\otimes W_1)C_\vartheta, C_\vartheta\rangle \geq \gamma \|C_\vartheta\|^2_{H^{\frac{1}{2},\frac{1}{2}}(\Gamma\times\Gamma)}. \quad (1.103)$$

Proof. We prove (1.103). The operator W_1 is self-adjoint and coercive in $H^{1/2}(\Gamma)$ (*e.g.*, Nédélec and Planchard (1973), Hsiao and Wendland (1977), McLean (2000)). Let $\{u_i\}_{i=1}^\infty$ denote an $H^{1/2}(\Gamma)$-orthonormal base in $H^{1/2}(\Gamma)$ consisting of eigenfunctions of W_1. Then, $\{u_i \otimes u_j\}_{i,j=1}^\infty$ is an orthonormal base in $H^{\frac{1}{2},\frac{1}{2}}(\Gamma\times\Gamma)$ and we may represent any $C_\vartheta \in H^{\frac{1}{2},\frac{1}{2}}(\Gamma\times\Gamma)$ in the form $C_\vartheta = \sum_{i,j=1}^\infty c_{ij} u_i \otimes u_j$. For any $M < \infty$, consider $C_\vartheta^M = \sum_{i,j=1}^M c_{ij} u_i \otimes u_j$. Then we calculate

$$\begin{aligned}\langle (W_1\otimes W_1)C_\vartheta^M, C_\vartheta^M\rangle &= \left\langle (W_1\otimes W_1)\sum_{i,j=1}^M c_{ij}u_i\otimes u_j, \sum_{i',j'=1}^M c_{i'j'}u_{i'}\otimes u_{j'}\right\rangle \\ &= \sum_{i,j=1}^M \lambda_i\lambda_j c_{ij}^2 \geq \lambda_1^2 \sum_{i,j=1}^M c_{ij}^2 = \lambda_1^2 \|C_\vartheta^M\|^2_{H^{1/2,1/2}(\Gamma\times\Gamma)}.\end{aligned}$$

Passing to the limit $M\to\infty$, we obtain (1.103) with $\gamma = \lambda_1^2$. □

We remark that the preceding proof shows that the continuity constant C_A^k in the *a priori* estimate (1.22) is sharp: in general the conditioning of the tensorized operator $A^{(k)}$ increases exponentially with k.

In the case $\kappa \neq 0$, we use that the integral operator W satisfies a Gårding inequality in $H^{1/2}(\Gamma)$ and obtain the unique solvability of the BIE (1.102) for C_ϑ from Theorem 1.4, provided that W is injective, *i.e.*, that κ is not a resonance frequency of the interior Dirichlet problem.

To compute the second moments of the random solution $U(x,\omega)$ at an interior point $x \in D$, we tensorize the representation formula (1.100) to yield

$$(\mathcal{M}^2U)(x,x) = \mathcal{M}^2(R_2\vartheta) = (R_2\otimes R_2)(\mathcal{M}^2\vartheta). \quad (1.104)$$

Then we obtain from Theorem 1.25 and from the sparse tensor Galerkin approximation (with spline wavelets which are V-orthogonal between levels) $\widetilde{Z}_L^2$ of $\mathcal{M}^2\vartheta$ in $\mathcal{O}(N_L(\log N_L)^3)$ operations and memory an approximation of $(\mathcal{M}^2U)(x,x)$ which satisfies, for smooth boundary Γ and data $\sigma \in L^2(\Omega; Y_{p+1/2}) = L^2(\Omega; H^{p+1}(\Gamma))$, at any interior point $x \in D$ the error bound

$$\left|(\mathcal{M}^2U)(x,x) - \langle R_2\otimes R_2, \widetilde{Z}_L^2\rangle\right| \leq c(x)(\log N_L)^3 N_L^{-2(p+1/2)} \|\sigma\|^2_{L^2(\Omega;H^{p+1}(\Gamma))}.$$

So far, we have considered the discretization of the boundary integral equation (1.102) by multi-level finite elements on the boundary surface Γ where convergence was achieved by mesh refinement.

We conclude this section with remarks on further developments in the analysis of sparse tensor discretizations of operator equations with random inputs. The results presented in this section are all based on hierarchies of subspaces $\{V_\ell\}_{\ell=0}^\infty$ of V consisting of piecewise polynomial functions of a fixed polynomial degree on a sequence $\{\mathcal{T}_\ell\}_{\ell=0}^\infty$ of triangulations of the physical domain. Alternatively, *spectral Galerkin discretizations* based on sequences of polynomial or trigonometric functions of increasing order can also be considered. For such spaces there are analogous sparse tensor constructions known as 'hyperbolic cross' spaces: see, *e.g.*, Temlyakov (1993) for the approximation theory of such spaces and Chernov and Schwab (2009) for an application to Galerkin approximations of boundary integral equations. Also, in the above presentation, we did not consider *adaptive refinements* in the sparse tensor discretizations. There is, however, a theory of sparse, adaptive tensor discretizations for subspace families satisfying axioms (W1)–(W5) of the present section available in Schwab and Stevenson (2008). As indicated at the beginning of this section, efficient discretization schemes for the tensorized equations (1.23) are also of interest in their own right, as such equations also arise in other models such as turbulence and transport equations. There is, however, one essential difference from (1.23): the equations (1.23) are *exact*, whereas in the two mentioned applications such equations can only be derived from additional *moment closure hypotheses*. For example, for PDEs with random operators, a suitable closure hypothesis could be smallness of fluctuations about the inputs' mean, and neglecting solution fluctuations beyond first order in the inputs' perturbation amplitude. This so-called *first-order second-moment* approach was first proposed in Dettinger and Wilson (1981) and was developed, in the context of random domains, in Harbrecht, Schneider and Schwab (2008).

The use of *quasi-Monte Carlo* (QMC) *methods* for the discretization of stochastic PDEs promises a rate of convergence higher than $M^{-1/2}$, which we proved here for MC methods. The numerical analysis of QMC for such problems is currently emerging. We refer to Graham, Kuo, Nuyens, Scheichl and Sloan (2010) and the references therein for algorithms and numerical experiments, as well as for references on QMC.

2. Stochastic Galerkin discretization

In Section 1 we considered Galerkin discretizations of k-point correlations of random fields in finite-dimensional spaces which were constructed from sparse tensor products of hierarchies of subspaces used for the approximations of single draws of u. The significance of this approach is twofold. First, for linear operator equations with stochastic data, we showed that k-point correlation functions of the random solutions are in fact solutions of high-dimensional *deterministic* equations for tensorized operators. We showed

that these tensorized operator equations naturally afford anistropic regularity results in scales of smoothness spaces of Sobolev or Besov–Triebel–Lizorkin type, so that the efficiency of sparse tensor approximations of kth moments will *not* incur the curse of dimensionality. In effect, with the formulation of deterministic equations for two- and k-point correlation functions of random solutions, we trade randomness for high-dimensionality.

The observation that k-point correlations of random fields satisfy deterministic tensorized operator equations is not new: it has been used frequently, for example in the derivation of moment closures in turbulence modelling or in the transition from atomistic-to-continuum models. For such nonlinear problems, however, there are generally no exact deterministic equations for the k-point correlation functions of the random solution: in general, a 'closure hypothesis' in some form is required. Due to the linear dependence of the random solution u on the random input f in (1.11), no closure hypothesis was required in the previous section.

While efficient computation of moments of order $k \geq 2$ of random solutions may be useful (*e.g.*, if the unknown random field is known *a priori* to be Gaussian), it is well known that even the knowledge of *all* k-point correlations will not characterize the law of the random solution if the moment problem is not solvable.

In the present section, we therefore address a second approach to the deterministic computation of random fields. It is based on parametrizing the random solution of a PDE with random data in terms of polynomials in a suitable coordinate representation of the random input. After having been pioneered by N. Wiener (1938), who established the representation of functionals of Wiener processes in terms of Hermite polynomials of a countable number of standard normal random variables, this 'spectral' view of random fields was shown to be generally applicable to any random field with finite second moments by Cameron and Martin (1947). Its use as a computational tool was pioneered in engineering applications in the 1990s. We mention only the book by Ghanem and Spanos (2007) and the series of papers by Xiu and Karniadakis (2002*a*), Xiu and Hesthaven (2005), Oden, Babuška, Nobile, Feng and Tempone (2005), Babuška, Nobile and Tempone (2005, 2007*a*, 2007*b*), Nobile, Tempone and Webster (2008*a*, 2008*b*), Nobile and Tempone (2009) and the references therein.

In these works, it was in particular observed that the expansion in Hermite polynomials of Gaussians originally proposed by Wiener (1938) and by Cameron and Martin (1947) is not always best suited for efficient computations. Often, other (bi)orthogonal function systems are better suited to the law of the random inputs, and are preferable in terms of computational efficiency. This has led to a formal generalization of Wiener's original polynomial chaos representations for computational purposes by G. E. Karniadakis and collaborators. We refer to Xiu and Karniadakis (2002*a*) and

to the survey by Xiu (2009) for further references on various applications of such 'generalized polynomial chaos' (GPC) methods.

The GPC approaches have received substantial attention in the past few years, both among numerical analysts and among computational scientists and engineers. From a computational point of view, the spectral representation once more renders the stochastic problem deterministic: rather than focusing on computation of spatiotemporal two- and k-point correlation functions, the *law of the unknown random solution* is approximated and computed in a parametric form. As we shall see shortly, in this context *questions of approximation and computation of deterministic quantities in infinite dimensions* arise naturally.

In this section, we present the mathematical formulation of generalized polynomial chaos representations of the laws of random solutions, starting with expansions into Hermite polynomials of Gaussians. Since the above-mentioned pioneering works of Wiener and of Cameron and Martin, these expansions have found numerous applications in stochastic analysis.

We then focus on elliptic problems with random diffusion coefficients, where expansions into polynomials of countably many non-Gaussian random variables are of interest.

We present several particular instances of such generalized polynomials chaos expansions, in particular the Wiener–Itô decomposition of random fields, and the Karhunen–Loève expansions, and review recent work on the *regularity of solutions of infinite-dimensional parametric, deterministic equations* in Section 3. Despite the formally infinite-dimensional setting, we review recent results that indicate that solutions of the infinite-dimensional parametric, deterministic equations for the laws of solutions of sPDEs exhibit regularity properties that allow finite-dimensional approximations free from the curse of dimensionality. We then address several adaptive strategies for the concrete construction of numerical approximations of such finite-dimensional approximations. We exhibit in particular sufficient conditions for convergence rates which are larger than the rate 1/2 proved in Theorem 1.11 above for MC methods.

2.1. Hermite chaos

Let H be a separable Hilbert space over $\mathbb{R}$ and let $\mu = N_Q$ be a non-degenerate centred Gaussian measure on H. For convenience, we assume that H is infinite-dimensional; all of the following also applies in the finite-dimensional setting.

Product structure of Gaussian measures

Let $(e_m)_{m\in\mathbb{N}}$ be an orthonormal basis of H such that

$$Qe_m = \lambda_m e_m, \quad m \in \mathbb{N}, \tag{2.1}$$

for a positive decreasing sequence $(\lambda_m)_{m\in\mathbb{N}}$. Note that (2.1) implies $e_m \in Q^{1/2}(H)$. We define a sequence of random variables on (H,μ) by

$$Y_m(x) := W_{e_m}(x) = \langle x, Q^{-1/2}e_m\rangle_H = \lambda_m^{-1/2}\langle x, e_m\rangle_H, \quad m\in\mathbb{N}. \tag{2.2}$$

Lemma 2.1. The random variables $(Y_m)_{m\in\mathbb{N}}$ on (H,μ) are independent and identically distributed. The distribution of each Y_m is the standard Gaussian measure N_1 on $\mathbb{R}$.

Proof. This is a consequence of Proposition C.36. We give, in addition, a direct proof. Let I be an arbitrary finite subset of $\mathbb{N}$ with $n := \#I$, and

$$Y_I : H \to \mathbb{R}^n, \quad x \mapsto (Y_m(x))_{m\in I}.$$

By the change of variables formula for Gaussian measures, the distribution of Y_I is the centred Gaussian measure

$$(Y_I)_{\#}\mu = N_{Y_I Q Y_I^*}$$

on $\mathbb{R}^n$, where $Y_I^* : \mathbb{R}^n \to H$ is the adjoint of Y_I given by

$$Y_I^*(\xi) = \sum_{m\in I} \xi_m Q^{-1/2} e_m, \quad \xi = (\xi_m)_{m\in I} \in \mathbb{R}^n.$$

Since $(e_m)_{m\in\mathbb{N}}$ is an orthonormal basis of H, for all $\xi = (\xi_m)_{m\in I} \in \mathbb{R}^n$,

$$\begin{aligned} Y_I Q Y_I^*(\xi) &= Y_I\left(Q\sum_{m\in I}\xi_m Q^{-1/2}e_m\right) = Y_I\left(\sum_{m\in I}\xi_m Q^{1/2}e_m\right) \\ &= \left(\sum_{m\in I}\langle \xi_m Q^{1/2}e_m, Q^{-1/2}e_{m'}\rangle_H\right)_{m'\in I} = \xi. \end{aligned}$$

Therefore,

$$(Y_I)_{\#}\mu = N_{\boldsymbol{I}_n} = \bigotimes_{m\in I} N_1,$$

where $\boldsymbol{I}_n$ is the identity matrix on $\mathbb{R}^n$. □

By Lemma 2.1, the distribution of the injective map

$$Y : H \to \mathbb{R}^\infty, \quad x \mapsto (Y_m(x))_{m\in\mathbb{N}} \tag{2.3}$$

is the countable product measure

$$\gamma := Y_{\#}\mu = \bigotimes_{m\in\mathbb{N}} N_1 \tag{2.4}$$

on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^\infty)$.

Proposition 2.2. The pullback

$$Y^* : L^2(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \gamma) \to L^2(H,\mu), \quad f \mapsto Y^* f = f\circ Y \tag{2.5}$$

is an isometric isomorphism of Hilbert spaces.

Proof. Since $\gamma = Y_{\#}\mu$, for all $f \in L^2(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \gamma)$,

$$\|Y^* f\|^2_{L^2(H,\mu)} = \int_H f(Y(x))^2 \mu(\mathrm{d}x) = \int_{\mathbb{R}^\infty} f(y)\gamma(\mathrm{d}y) = \|f\|^2_{L^2(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty),\gamma)},$$

so Y^* is an isometry. To show surjectivity, note that the Borel σ-algebra $\mathcal{B}(H)$ is generated by Y since H is separable and the topology of H is generated by Y. Therefore, the Doob–Dynkin lemma implies that any $g \in L^2(H,\mu)$ is of the form $g = f \circ Y$ for a $\mathcal{B}(\mathbb{R}^\infty)$-measurable function f on $\mathbb{R}^\infty$. The above computation implies $f \in L^2(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \gamma)$. □

The Hermite polynomial basis

Consider the analytic function

$$F(t,\xi) := \mathrm{e}^{-\frac{1}{2}t^2 + t\xi}, \quad t, \xi \in \mathbb{R}. \tag{2.6}$$

We define the *Hermite polynomials* $(H_n)_{n\in\mathbb{N}_0}$ through the power series representation of $F(\cdot,\xi)$ around $t = 0$,

$$F(t,\xi) = \sum_{n=0}^{\infty} \frac{t^n}{\sqrt{n!}} H_n(\xi), \quad t, \xi \in \mathbb{R}. \tag{2.7}$$

Lemma 2.3. For all $n \in \mathbb{N}_0$ and $\xi \in \mathbb{R}$, if $H_{-1}(\xi) := 0$,

$$H_n(\xi) = \frac{(-1)^n}{\sqrt{n!}} \mathrm{e}^{\frac{1}{2}\xi^2} D_\xi^n \mathrm{e}^{-\frac{1}{2}\xi^2}, \tag{2.8}$$

$$\xi H_n(\xi) = \sqrt{n+1} H_{n+1}(\xi) + \sqrt{n} H_{n-1}(\xi), \tag{2.9}$$

$$D_\xi H_n(\xi) = \sqrt{n} H_{n-1}(\xi), \tag{2.10}$$

$$-D_\xi^2 H_n(\xi) + \xi D_\xi H_n(\xi) = n H_n(\xi). \tag{2.11}$$

Proof. Equation (2.8) follows by Taylor expansion,

$$F(t,\xi) = \mathrm{e}^{\frac{1}{2}\xi^2} \mathrm{e}^{-\frac{1}{2}(t-\xi)^2} = \mathrm{e}^{\frac{1}{2}\xi^2} \sum_{n=0}^{\infty} \frac{t^n}{n!} D_t^n|_{t=0} \mathrm{e}^{-\frac{1}{2}(t-\xi)^2}$$

$$= \mathrm{e}^{\frac{1}{2}\xi^2} \sum_{n=0}^{\infty} \frac{t^n}{n!} (-1)^n D_\xi^n \mathrm{e}^{-\frac{1}{2}\xi^2},$$

and comparison with (2.7).

To show (2.9), we note that

$$D_t F(t,\xi) = \sum_{n=1}^{\infty} \frac{\sqrt{n}\, t^{n-1}}{\sqrt{(n-1)!}} H_n(\xi) = \sum_{n=0}^{\infty} \frac{t^n}{\sqrt{n!}} \sqrt{n+1} H_{n+1}(\xi).$$

Also, by (2.6),

$$D_t F(t,\xi) = (\xi - t)F(t,\xi) = \sum_{n=0}^{\infty} \frac{t^n}{\sqrt{n!}} \xi H_n(\xi) - \sum_{n=0}^{\infty} \frac{t^n}{\sqrt{n!}} \sqrt{n} H_{n-1}(\xi).$$

Similarly, (2.10) follows by comparing two representations of $D_\xi F(t,\xi)$,

$$\sum_{n=0}^{\infty} \frac{t^n}{\sqrt{n!}} D_\xi H_n(\xi) = D_\xi F(t,\xi) = tF(t,\xi) = \sum_{n=0}^{\infty} \frac{t^n}{\sqrt{n!}} \sqrt{n} H_{n-1}(\xi).$$

Finally, (2.11) is a consequence of (2.9) and (2.10),

$$D_\xi^2 H_n(\xi) - \xi D_\xi H_n(\xi) = \sqrt{n}(D_\xi H_{n-1}(\xi) - \xi H_{n-1}(\xi)) = -nH_n(\xi). \qquad \square$$

In particular, H_n is a polynomial of degree n. The first few Hermite polynomials are

$$H_0(\xi) = 1, \quad H_1(\xi) = \xi, \quad H_2(\xi) = \frac{1}{\sqrt{2}}(\xi^2 - 1). \tag{2.12}$$

Proposition 2.4. $(H_n)_{n\in\mathbb{N}_0}$ is an orthonormal basis of $L^2(\mathbb{R}, N_1)$.

Proof. We first show orthonormality. Note that for $\xi, s, t \in \mathbb{R}$,

$$\mathrm{e}^{-\frac{1}{2}(t^2+s^2)+\xi(t+s)} = F(t,\xi)F(s,\xi) = \sum_{n,m=0}^{\infty} \frac{t^n}{\sqrt{n!}} \frac{s^m}{\sqrt{m!}} H_n(\xi) H_m(\xi).$$

Integrating over $\mathbb{R}$ with respect to N_1, we have

$$\int_{\mathbb{R}} F(t,\xi)F(s,\xi) N_1(\mathrm{d}\xi) = \sum_{n,m=0}^{\infty} \frac{t^n}{\sqrt{n!}} \frac{s^m}{\sqrt{m!}} \int_{\mathbb{R}} H_n(\xi) H_m(\xi) N_1(\mathrm{d}\xi),$$

and also

$$\int_{\mathbb{R}} F(t,\xi)F(s,\xi) N_1(\mathrm{d}\xi) = \frac{\mathrm{e}^{ts}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}(\xi - t - s)^2} \,\mathrm{d}\xi = \mathrm{e}^{ts} = \sum_{n=0}^{\infty} \frac{t^n s^n}{n!}.$$

Therefore,

$$\int_{\mathbb{R}} H_n(\xi) H_m(\xi) N_1(\mathrm{d}\xi) = \delta_{nm}.$$

To show completeness, let $f \in L^2(\mathbb{R}, N_1)$ be orthogonal to H_n for all $n \in \mathbb{N}_0$. Then $g(\xi) := f(\xi)\mathrm{e}^{-\xi^2/4}$ is in $L^2(\mathbb{R})$, and for all $t \in \mathbb{R}$,

$$0 = \int_{\mathbb{R}} f(\xi) F(t,\xi) N_1(\mathrm{d}\xi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(\xi) \mathrm{e}^{-\frac{1}{2}t^2 + t\xi - \frac{1}{4}\xi^2} \,\mathrm{d}\xi.$$

Since $-\frac{1}{2}t^2 + t\xi - \frac{1}{4}\xi^2 = -\frac{1}{4}(2t - \xi)^2 + \frac{1}{2}t^2$, this implies that the convolution $g * \varphi = 0$ for $\varphi(\xi) := \mathrm{e}^{-\xi^2/4}$. Taking the Fourier transform, we have $\widehat{g}\widehat{\varphi} = 0$,

and since $\widehat{\varphi}$ is non-zero everywhere, $\widehat{g} = 0$ in $L^2(\mathbb{R})$. This implies $g = 0$ almost everywhere, and therefore $f = 0$ almost everywhere. □

Using Theorem 2.12 below and Proposition 2.2, we construct an orthonormal basis of $L^2(H, \mu)$. Define the index set of finitely supported sequences in $\mathbb{N}$,

$$\mathfrak{F} := \{\nu \in \mathbb{N}_0^{\mathbb{N}} \,;\, \#\operatorname{supp}\nu < \infty\}, \tag{2.13}$$

where

$$\operatorname{supp}\nu := \{m \in \mathbb{N} \,;\, \nu_m \neq 0\}, \quad \nu \in \mathbb{N}_0^{\mathbb{N}}. \tag{2.14}$$

For all $\nu \in \mathfrak{F}$, we define the tensor product Hermite polynomial

$$H_\nu := \bigotimes_{m\in\mathbb{N}} H_{\nu_m}, \tag{2.15}$$

i.e., for all $y \in \mathbb{R}^\infty$, since $H_0(\xi) = 1$,

$$H_\nu(y) = \prod_{m\in\mathbb{N}} H_{\nu_m}(y_m) = \prod_{m\in\operatorname{supp}\nu} H_{\nu_m}(y_m). \tag{2.16}$$

The degree of the polynomial H_ν for $\nu \in \mathfrak{F}$ is given by

$$|\nu| := \sum_{m\in\mathbb{N}} \nu_m = \sum_{m\in\operatorname{supp}\nu} \nu_m. \tag{2.17}$$

We use the pullback Y^* from (2.5) to define H_ν on H. For all $x \in H$ and $\nu \in \mathfrak{F}$,

$$H_\nu(x) := (Y^* H_\nu)(x) = H_\nu(Y(x)) = \prod_{m\in\mathbb{N}} H_{\nu_m}(W_{e_m}(x)). \tag{2.18}$$

As in (2.16), the product in (2.18) is finite, since all but finitely many factors are one by definition of $\mathfrak{F}$ and $H_0(\xi) = 1$.

Theorem 2.5. $(H_\nu)_{\nu\in\mathfrak{F}}$ is an orthonormal basis of $L^2(H, \mu)$.

Proof. By Proposition 2.4 and Theorem 2.12 below, $(H_\nu)_{\nu\in\mathfrak{F}}$ from (2.15) is an orthonormal basis of $L^2(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \gamma)$. Since the pullback Y^* is an isometric isomorphism from $L^2(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \gamma)$ to $L^2(H, \mu)$ by Proposition 2.2, $(H_\nu)_{\nu\in\mathfrak{F}}$ from (2.18) is an orthonormal basis of $L^2(H, \mu)$. □

We call $(H_\nu)_{\nu\in\mathfrak{F}}$ the *Hermite chaos* basis of $L^2(H, \mu)$.

Wiener–Itô decomposition

For all $n \in \mathbb{N}_0$, we define the Wiener chaos of order n as the closed subspace

$$L^2_n(H, \mu) := \operatorname{span}\{H_n(W_f(x)) \,;\, f \in H,\ \|f\|_H = 1\} \subset L^2(H, \mu), \tag{2.19}$$

where the white noise map W_f is defined by (C.52) and continuous extension to H. We recall the identity

$$\int_H \mathrm{e}^{W_f(x)}\mu(\mathrm{d}x) = \mathrm{e}^{\frac{1}{2}\|f\|_H^2} \quad \forall f \in H \tag{2.20}$$

from Proposition C.36.

Lemma 2.6. For all $f, g \in H$ with $\|f\|_H = \|g\|_H = 1$ and all $n, m \in \mathbb{N}_0$,

$$\int_H H_n(W_f(x))H_m(W_g(x))\mu(\mathrm{d}x) = \delta_{nm}\langle f, g\rangle_H^n. \tag{2.21}$$

Proof. As in the proof of Proposition 2.4, for $t, s \in \mathbb{R}$,

$$\int_H F(t, W_f)F(s, W_g)\,\mathrm{d}\mu = \sum_{n,m=0}^{\infty} \frac{t^n}{\sqrt{n!}}\frac{s^m}{\sqrt{m!}}\int_H H_n(W_f)H_m(W_g)\,\mathrm{d}\mu,$$

and also, using (2.20) and $F(t, W_f)F(s, W_g) = \mathrm{e}^{-\frac{1}{2}(t^2+s^2)+tW_f+sW_g}$,

$$\begin{aligned}\int_H F(t, W_f)F(s, W_g)\,\mathrm{d}\mu &= \mathrm{e}^{-\frac{1}{2}(t^2+s^2)}\int_H \mathrm{e}^{W_{tf+sg}}\,\mathrm{d}\mu = \mathrm{e}^{-\frac{1}{2}(t^2+s^2)}\mathrm{e}^{\frac{1}{2}\|tf+sg\|_H^2}\\ &= \mathrm{e}^{ts\langle f,g\rangle_H} = \sum_{n=0}^{\infty}\frac{t^n s^n}{n!}\langle f, g\rangle_H^n. \qquad \square\end{aligned}$$

Theorem 2.7. (Wiener–Itô decomposition)

$$L^2(H, \mu) = \bigoplus_{n\in\mathbb{N}_0} L_n^2(H, \mu). \tag{2.22}$$

Proof. Orthogonality of the spaces $L_n^2(H, \mu)$ follows from Lemma 2.6. It remains to be shown that these spaces span $L^2(H, \mu)$. Let $g \in L^2(H, \mu)$ be orthogonal to $H_n(W_f)$ for all $n \in \mathbb{N}_0$ and all $f \in H$ with $\|f\|_H = 1$. Then, for all $t \in \mathbb{R}$ and any $f \in H$ with $\|f\|_H = 1$,

$$0 = \int_H F(t, W_f)g\,\mathrm{d}\mu = \mathrm{e}^{-\frac{1}{2}t^2}\int_H \mathrm{e}^{tW_f}g\,\mathrm{d}\mu.$$

Consequently, the entire function

$$\varphi(t) := \int_H \mathrm{e}^{tW_f}g\,\mathrm{d}\mu$$

vanishes on $\mathbb{R}$, and thus is equal to 0 on $\mathbb{C}$. Let ϑ be the signed measure $\mathrm{d}\vartheta = g\,\mathrm{d}\mu$. An arbitrary element $h \in H$ is of the form $h = tQ^{-1/2}f$ for an $f \in Q^{1/2}(H)$ with $\|f\|_H = 1$ and some $t \in \mathbb{R}$. The Fourier transform of ϑ evaluated at h is

$$\widehat{\vartheta}(h) = \int_H \mathrm{e}^{\mathrm{i}\langle x, tQ^{-1/2}f\rangle_H}g(x)\mu(\mathrm{d}x) = \int_H \mathrm{e}^{\mathrm{i}tW_f}g\,\mathrm{d}\mu = \varphi(\mathrm{i}t) = 0.$$

Therefore, $\vartheta = 0$ and it follows that $g = 0$ almost everywhere. □

The following proposition describes the connection between the Wiener–Itô decomposition (2.22) and the Hermite chaos basis $(H_\nu)_{\nu\in\mathfrak{F}}$ of $L^2(H,\mu)$ from Theorem 2.5.

Proposition 2.8. For all $n \in \mathbb{N}_0$,

$$L^2_n(H,\mu) = \operatorname{span}\{H_\nu(x)\,;\ \nu\in\mathfrak{F},\ |\nu| = n\}. \tag{2.23}$$

Proof. It suffices to show that for $n \in \mathbb{N}_0$ and $\nu \in \mathfrak{F}$ with $|\nu| \neq n$,

$$\int_H H_\nu(x)H_n(W_f(x))\mu(\mathrm{d}x) = 0 \quad \forall f \in H : \|f\|_H = 1. \tag{2.24}$$

Then the inclusions '$\subset$' and '$\supset$' follow from Theorem 2.5 and Theorem 2.7, respectively.

Let $f \in H$ with $\|f\|_H = 1$. Since $\operatorname{supp}\nu$ is finite for $\nu \in \mathfrak{F}$, there is an $N \in \mathbb{N}_0$ with $\nu_i = 0$ for all $i \geq N+1$. In particular,

$$H_\nu(x) = H_{\nu_1}(W_{e_1}(x))H_{\nu_2}(W_{e_2}(x))\cdots H_{\nu_N}(W_{e_N}(x)), \quad x \in H.$$

For $t_1,\dots,t_{N+1} \in \mathbb{R}$, we compute

$$I := \int_H F(t_1,W_{e_1})\cdots F(t_N,W_{e_N})F(t_{N+1},W_f)\,\mathrm{d}\mu$$

twice. Using (2.6), linearity of W and (2.20), we have

$$\begin{aligned} I &= \mathrm{e}^{-\frac12(t_1^2+\cdots+t_{N+1}^2)}\int_H \mathrm{e}^{W_{t_1e_1+\cdots+t_Ne_N+t_{N+1}f}}\,\mathrm{d}\mu \\ &= \mathrm{e}^{-\frac12(t_1^2+\cdots+t_{N+1}^2)}\mathrm{e}^{\frac12\|t_1e_1+\cdots+t_Ne_N+t_{N+1}f\|_H^2}. \end{aligned}$$

Abbreviating $f_i := \langle f,e_i\rangle_H$, $i\in\mathbb{N}$, since $f_1^2+f_2^2+\cdots = \|f\|_H^2 = 1$,

$$\begin{aligned} &\|t_1e_1+\cdots+t_Ne_N+t_{N+1}f\|_H^2 \\ &\quad= (t_1+t_{N+1}f_1)^2+\cdots+(t_N+t_{N+1}f_N)^2+t_{N+1}^2(f_{N+1}^2+f_{N+2}^2+\cdots) \\ &\quad= t_1^2+\cdots+t_N^2+2t_{N+1}(t_1f_1+\cdots+t_Nf_N)+t_{N+1}^2. \end{aligned}$$

Since the quadratic terms cancel, we are left with

$$I = \mathrm{e}^{t_{N+1}(t_1f_1+\cdots+t_Nf_N)} = \sum_{n=0}^\infty \frac{t_{N+1}^n(t_1f_1+\cdots+t_Nf_N)^n}{n!}.$$

Also, (2.7) implies

$$I = \sum_{k_1,\dots,k_{N+1}=0}^\infty \frac{t_1^{k_1}\cdots t_{N+1}^{k_{N+1}}}{\sqrt{k_1!\cdots k_{N+1}!}}\int_H H_{k_1}(W_{e_1})\cdots H_{k_N}(W_{e_N})H_{k_{N+1}}(W_f)\,\mathrm{d}\mu.$$

Comparing the last two equations leads to (2.24). □

2.2. Generalized polynomial chaos

The construction of an orthonormal basis in Section 2.1 is not specific to Gaussian measures or Hermite polynomials. The important ingredient is the countable product structure of the probability space (H, μ), which we illustrated by the measure-preserving map Y into the product measure space $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \gamma)$. We generalize the construction of the chaos basis to countable products of arbitrary probability spaces. Again, all of the following also holds for finite products with the obvious modifications; to simplify notation, we consider only the countable case. We refer to Gittelson (2011*a*) for further details and a more general construction.

Countable products of probability spaces
For all $m \in \mathbb{N}$, let Γ_m be an arbitrary non-empty set, endowed with a σ-algebra Σ_m. Let $(Y_m)_{m\in\mathbb{N}}$ be independent random variables on a probability space $(\Omega, \Sigma, \mathbb{P})$, such that Y_m maps into (Γ_m, Σ_m). This sequence constitutes a map

$$Y : \Omega \to \Gamma := \prod_{m\in\mathbb{N}} \Gamma_m, \quad \omega \mapsto (Y_m(\omega))_{m\in\mathbb{N}}, \tag{2.25}$$

which is measurable with respect to the product σ-algebra $\Sigma := \bigotimes_{m\in\mathbb{N}} \Sigma_m$ on Γ. By the independence of $(Y_m)_{m\in\mathbb{N}}$, the distribution of Y is the countable product probability measure

$$\mu := \bigotimes_{m\in\mathbb{N}} \mu_m \tag{2.26}$$

on (Γ, Σ), where $\mu_m = (Y_m)_\#(\mathbb{P})$ is the distribution of Y_m on (Γ_m, Σ_m).

Countable product bases
For all $m \in \mathbb{N}$, let $(\varphi_{m,i})_{i\in\mathbb{N}_0}$ be an orthonormal basis of $L^2(\Gamma_m, \Sigma_m, \mu_m)$ such that $\varphi_{m,0} = 1$; the constant 1 is normalized in $L^2(\Gamma_m, \Sigma_m, \mu_m)$ since μ_m is a probability measure. As in Section 2.1, we define the index set

$$\mathfrak{F} := \{\nu \in \mathbb{N}_0^{\mathbb{N}} ; \ \#\operatorname{supp}\nu < \infty\}. \tag{2.27}$$

If $L^2(\Gamma_m, \Sigma_m, \mu_m)$ is finite-dimensional for some m, then of course its orthonormal basis $(\varphi_{m,i})_{i=0}^N$ is finite, and we restrict ν_m to the values $0, 1, \ldots, N$ in the definition of $\mathfrak{F}$.

For all $\nu \in \mathfrak{F}$, define the tensor product

$$\varphi_\nu := \bigotimes_{m\in\mathbb{N}} \varphi_{m,\nu_m}, \tag{2.28}$$

i.e., for all $y = (y_m)_{m\in\mathbb{N}} \in \Gamma$, since $\varphi_{m,0} = 1$ for all $m \in \mathbb{N}$,

$$\varphi_\nu(y) = \prod_{m\in\mathbb{N}} \varphi_{m,\nu_m}(y_m) = \prod_{m\in\operatorname{supp}\nu} \varphi_{m,\nu_m}(y_m). \tag{2.29}$$

Let $\mathcal{F}(\mathbb{N})$ denote the set of all finite subsets of $\mathbb{N}$. For $I \in \mathcal{F}(\mathbb{N})$, define the finite product σ-algebra

$$\Sigma_I := \bigotimes_{m\in I} \Sigma_m = \sigma(y_m\,;\ m \in I) \subset \Sigma. \tag{2.30}$$

A function is Σ_I-measurable if it is Σ-measurable and only depends on $(y_m)_{m\in I}$. Also, let $\mathfrak{F}_I := \{\nu \in \mathfrak{F}\,;\ \operatorname{supp}\nu \subset I\}$.

Lemma 2.9. For all $I \in \mathcal{F}(\mathbb{N})$, the set $(\varphi_\nu)_{\nu\in\mathfrak{F}_I}$ is an orthonormal basis of $L^2(\Gamma, \Sigma_I, \mu)$.

Proof. Since $\varphi_{m,0} = 1$ for all $m \in \mathbb{N}$, if $\operatorname{supp}\nu \subset I$, then

$$\varphi_\nu(y) = \prod_{m\in\mathbb{N}} \varphi_{m,\nu_m}(y_m) = \prod_{m\in I} \varphi_{m,\nu_m}(y_m), \quad y \in \Gamma.$$

Due to the assumption that I is finite,

$$L^2(\Gamma, \Sigma_I, \mu) \cong \bigotimes_{m\in I} L^2(\Gamma_m, \Sigma_m, \mu_m).$$

As $(\varphi_{m,i})_{i\in\mathbb{N}}$ is an orthonormal basis of $L^2(\Gamma_m, \Sigma_m, \mu_m)$ for each $m \in I$ by assumption, the claim follows since finite tensor products of orthonormal bases form an orthonormal basis in the product space. □

The monotone class theorem implies that any function in $L^2(\Gamma, \Sigma, \mu)$ can be approximated by Σ_I-measurable functions with $I \in \mathcal{F}(\mathbb{N})$. We recall that a set $\mathfrak{M}$ of real-valued functions on Γ is *multiplicative* if $vw \in \mathfrak{M}$ whenever $v, w \in \mathfrak{M}$. A *monotone vector space* over Γ is a real vector space $\mathfrak{H}$ of bounded, real-valued functions on Γ such that all constants are in $\mathfrak{H}$, and if $(v_n)_{n\in\mathbb{N}}$ is a sequence in $\mathfrak{H}$ with $0 \le v_n \le v_{n+1}$ for all $n \in \mathbb{N}$ and $v := \sup_n v_n$ is a bounded function on Γ, then $v \in \mathfrak{H}$.

Theorem 2.10. (monotone class theorem) Let $\mathfrak{M}$ be a multiplicative class of bounded, real-valued functions on Γ and let $\mathfrak{H}$ be a monotone vector space containing $\mathfrak{M}$. Then $\mathfrak{H}$ contains all bounded $\sigma(\mathfrak{M})$-measurable functions.

We refer to Protter (2005, Theorem I.8) for a proof of Theorem 2.10.

Proposition 2.11. $\bigcup_{I\in\mathcal{F}(\mathbb{N})} L^2(\Gamma, \Sigma_I, \mu)$ is dense in $L^2(\Gamma, \Sigma, \mu)$.

Proof. Let $\mathfrak{V} := \overline{\bigcup_{I\in\mathcal{F}(\mathbb{N})} L^2(\Gamma, \Sigma_I, \mu)} \subset L^2(\Gamma, \Sigma, \mu)$ and define $\mathfrak{H} := \mathfrak{V} \cap L^\infty(\Gamma, \Sigma, \mu)$ as the vector space of bounded functions in $\mathfrak{V}$. Let $\mathfrak{M}$ be the set of indicator functions in $L^2(\Gamma, \Sigma_I, \mu)$ for any $I \in \mathcal{F}(\mathbb{N})$. Then $\mathfrak{M} \subset \mathfrak{H}$, $1 \in \mathfrak{H}$, and $\mathfrak{M}$ is closed under multiplication. Let $0 \le v_1 \le v_2 \le \cdots$ be a pointwise monotonic sequence in $\mathfrak{H}$ and $v := \sup_n v_n$ its pointwise supremum. If $v \in L^\infty(\Gamma, \Sigma, \mu) \subset L^2(\Gamma, \Sigma, \mu)$, then $(v_n)_n$ converges to v

in $L^2(\Gamma,\Sigma,\mu)$ by dominated convergence. Since $\mathfrak{V}$ is closed in $L^2(\Gamma,\Sigma,\mu)$, it follows that $v\in\mathfrak{V}$ and therefore $v\in\mathfrak{H}$. Thus $\mathfrak{H}$ is a monotone vector space and, using that $\Sigma=\sigma(\mathfrak{M})$ is the σ-algebra generated by $\mathfrak{M}$, the monotone class theorem implies $\mathfrak{H}=L^\infty(\Gamma,\Sigma,\mu)$.

If $v\in L^2(\Gamma,\Sigma,\mu)$, then for any $N\in\mathbb{N}$, $v1_{\{|v|\le N\}}\in L^\infty(\Gamma,\Sigma,\mu)=\mathfrak{H}\subset\mathfrak{V}$ and $v\in\mathfrak{V}$ by dominated convergence. □

Theorem 2.12. $(\varphi_\nu)_{\nu\in\mathfrak{F}}$ is an orthonormal basis of $L^2(\Gamma,\Sigma,\mu)$.

Proof. Orthonormality follows from Lemma 2.9 since, for any $\nu,\widetilde{\nu}\in\mathfrak{F}$,

$$I:=\operatorname{supp}\nu\cup\operatorname{supp}\widetilde{\nu}\in\mathcal{F}(\mathbb{N}).$$

Density follows from Proposition 2.11 since $(\varphi_\nu)_{\nu\in\mathfrak{F}_I}$ spans $L^2(\Gamma,\Sigma_I,\mu)$ for all $I\in\mathcal{F}(\mathbb{N})$. □

Let $(\varphi_{m,i})_{i\in\mathbb{N}_0}$ be a graded basis of $L^2(\Gamma_m,\Sigma_m,\mu_m)$ for each $m\in\mathbb{N}$, *i.e.*, there is a map $\ell_m:\mathbb{N}_0\to\mathbb{N}_0$ assigning to each index $i\in\mathbb{N}_0$ a level $\ell(i)$. This might be the degree of a polynomial, or the level of a wavelet, depending on $(\varphi_{m,i})_{i\in\mathbb{N}_0}$. We assume that $\ell_m(0)=0$ for all $m\in\mathbb{N}$. This allows us to define a grading function for the orthonormal basis $(\varphi_\nu)_{\nu\in\mathfrak{F}}$ of $L^2(\Gamma,\Sigma,\mu)$ by

$$\ell(\nu):=\sum_{m\in\mathbb{N}}\ell_m(\nu_m)=\sum_{m\in\operatorname{supp}_\nu}\ell_m(\nu_m),\quad \nu\in\mathfrak{F}. \tag{2.31}$$

This function induces a decomposition of $L^2(\Gamma,\Sigma,\mu)$ into the closed subspaces

$$L^2_n(\Gamma,\Sigma,\mu):=\operatorname{span}\{\varphi_\nu\,;\,\nu\in\mathfrak{F},\,\ell(\nu)=n\}\subset L^2(\Gamma,\Sigma,\mu),\quad n\in\mathbb{N}_0. \tag{2.32}$$

Corollary 2.13.

$$L^2(\Gamma,\Sigma,\mu)=\bigoplus_{n\in\mathbb{N}_0}L^2_n(\Gamma,\Sigma,\mu). \tag{2.33}$$

Proof. By Theorem 2.12 and (2.32),

$$L^2(\Gamma,\Sigma,\mu)=\bigoplus_{\nu\in\mathfrak{F}}\operatorname{span}\varphi_\nu=\bigoplus_{n\in\mathbb{N}_0}\bigoplus_{\ell(\nu)=n}\operatorname{span}\varphi_\nu=\bigoplus_{n\in\mathbb{N}_0}L^2_n(\Gamma,\Sigma,\mu).\qquad\square$$

We note that other choices of ℓ are possible. For example, the dimensions $m\in\mathbb{N}$ can be weighted differently, leading to an anisotropic decomposition. Also, the ℓ^1-norm in (2.32) can be generalized to an arbitrary ℓ^p-quasi-norm for any $p>0$. Rounding the final value ensures that ℓ maps into $\mathbb{N}_0$.

Orthogonal polynomials

We assume that Γ_m is a Borel subset of $\mathbb{R}$ and that μ_m has finite moments

$$M_n:=\int_{\Gamma_m}\xi^n\mu_m(\mathrm{d}\xi),\quad n\in\mathbb{N}_0. \tag{2.34}$$

Orthonormal polynomials with respect to μ_m can be constructed by the well-known three-term recursion

$$\beta_{n+1}P_{n+1}(\xi) = (\xi - \alpha_n)P_n(\xi) - \beta_n P_{n-1}(\xi), \quad n \in \mathbb{N}_0, \tag{2.35}$$

with the initialization $P_{-1}(\xi) = 0$ and $P_0(\xi) = 1$. The coefficients are

$$\alpha_n := \int_{\Gamma_m} \xi P_n(\xi)^2 \mu_m(\mathrm{d}\xi) \quad \text{and} \quad \beta_n := \frac{c_{n-1}}{c_n}, \tag{2.36}$$

where c_n is the leading coefficient of P_n, and $\beta_0 := 1$. The values of $(\alpha_n)_{n\in\mathbb{N}_0}$ and $(\beta_n)_{n\in\mathbb{N}_0}$ are tabulated for many common distributions μ_m (Gautschi 2004). Formula (2.35) can be derived by Gram–Schmidt orthogonalization of the monomials $(\xi^n)_{n\in\mathbb{N}_0}$. Note that β_{n+1} depends on P_{n+1}, and can be computed by normalizing the right-hand side of (2.35) in $L^2(\Gamma_m, \Sigma_m, \mu_m)$.

Lemma 2.14. For all $n \in \mathbb{N}_0$, P_n is a polynomial of degree n if $n < N := \dim L^2(\Gamma_m, \Sigma_m, \mu_m)$ and zero otherwise. The sequence $(P_n)_{n\in\mathbb{N}_0}$ (resp. $(P_n)_{n=0}^{N-1}$ if N is finite) is orthonormal in $L^2(\Gamma_m, \Sigma_m, \mu_m)$.

Proof. By the Gram–Schmidt orthogonalization process applied to the monomials $(\xi^n)_{n\in\mathbb{N}_0}$, μ_m-orthonormal polynomials $(P_n)_{n\in\mathbb{N}}$ exist. If N is finite, then $(\xi^n)_{n=0}^{N-1}$ is a basis of $L^2(\Gamma_m, \Sigma_m, \mu_m)$, and thus $P_n = 0$ for all $n \geq N$. We show that the orthonormal polynomials constructed by Gram–Schmidt orthogonalization satisfy (2.35).

Note that $\beta_{n+1}P_{n+1}(\xi) - \xi P_n(\xi)$ is a polynomial of degree at most n. Therefore, and since $(P_k)_{k=0}^{n+1}$ are orthonormal,

$$\beta_{n+1}P_{n+1}(\xi) - \xi P_n(\xi) = \gamma_n P_n(\xi) + \gamma_{n-1}P_{n-1}(\xi) + \cdots + \gamma_0$$

with

$$\gamma_k = \int_{\Gamma_m} (\beta_{n+1}P_{n+1}(\xi) - \xi P_n(\xi))P_k(\xi)\mu_m(\mathrm{d}\xi) = -\int_{\Gamma_m} \xi P_n(\xi)P_k(\xi)\mu_m(\mathrm{d}\xi)$$

for $k = 0, 1, \ldots, n$. In particular, $\gamma_n = -\alpha_n$, and $\gamma_k = 0$ for $k \leq n-2$ since $\xi P_k(\xi)$ is a polynomial of degree at most $n-1$. We note that $\xi P_{n-1}(\xi) = \beta_n P_n(\xi) + q(\xi)$ for a polynomial q of degree at most $n-1$. This implies $\gamma_{n-1} = -\beta_n$. □

If $N = \dim L^2(\Gamma_m, \Sigma_m, \mu_m)$ is finite, then it follows from Lemma 2.14 that $(P_n)_{n=0}^{N-1}$ is an orthonormal basis of $L^2(\Gamma_m, \Sigma_m, \mu_m)$. In general, this requires an additional assumption. We consider the case $N = \infty$ in the following.

The measure μ_m is called *determinate* if it is uniquely characterized by its moments $(M_n)_{n\in\mathbb{N}_0} \subset \mathbb{R}$. We note that μ_m is always determinate if $\Gamma_m \subset \mathbb{R}$ is bounded (Gautschi 2004, Theorem 1.41). The following result was shown by F. Riesz in 1923 (see, *e.g.*, Szegő (1975) for a proof).

Proposition 2.15. If μ_m is determinate, then $(P_n)_{n\in\mathbb{N}_0}$ is an orthonormal basis of $L^2(\Gamma_m, \Sigma_m, \mu_m)$.

Examples of generalized polynomial chaos bases

We combine Theorem 2.12 with Proposition 2.15 to construct a countable tensor product basis of $L^2(\Gamma, \Sigma, \mu)$. Again, we assume for simplicity that $L^2(\Gamma_m, \Sigma_m, \mu_m)$ is infinite-dimensional for all $m \in \mathbb{N}$. Analogous results hold in the general setting.

For all $m \in \mathbb{N}$, let $(P_n^m)_{n\in\mathbb{N}_0}$ be the orthonormal polynomial basis of $L^2(\Gamma_m, \Sigma_m, \mu_m)$ from Proposition 2.15. Then, by Theorem 2.12, the tensor product polynomials

$$P_\nu := \bigotimes_{m\in\mathbb{N}} P_{\nu_m}^m, \quad \nu \in \mathfrak{F}, \tag{2.37}$$

form an orthonormal basis of $L^2(\Gamma, \Sigma, \mu)$, which we call the *generalized polynomial chaos basis*.

If $\mu_m = N_1$ for all $m \in \mathbb{N}$, then $(P_n^m)_{n\in\mathbb{N}_0}$ are the Hermite polynomials (2.8) and $(P_\nu)_{\nu\in\mathfrak{F}}$ is the Hermite chaos basis, interpreted as a basis of $L^2(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \gamma)$ instead of $L^2(H, \mu)$. In this case, Corollary 2.13 reduces to the Wiener–Itô decomposition, Theorem 2.7, due to Proposition 2.8.

We consider as another example the case when μ_m is the uniform distribution on $\Gamma_m := [-1, 1]$ for all $m \in \mathbb{N}$, *i.e.*, $\mu_m(\mathrm{d}\xi) = \frac{1}{2}\,\mathrm{d}\xi$. The corresponding orthonormal polynomials are the Legendre polynomials, which are defined by the three-term recursion

$$\frac{n+1}{\sqrt{2n+3}\sqrt{2n+1}}\,L_{n+1}(\xi) = \xi L_n(\xi) - \frac{n}{\sqrt{2n+1}\sqrt{2n-1}}\,L_{n-1}(\xi), \quad n \in \mathbb{N}_0, \tag{2.38}$$

with $L_{-1}(\xi) = 0$ and $L_0(\xi) = 1$. The Legendre polynomials satisfy Rodrigues' formula

$$L_n(\xi) = \frac{\sqrt{2n+1}}{2^n n!}\,\frac{\mathrm{d}^n}{\mathrm{d}\xi^n}(\xi^2 - 1)^n, \quad n \in \mathbb{N}_0. \tag{2.39}$$

The first few Legendre polynomials are

$$L_0(\xi) = 1, \quad L_1(\xi) = \sqrt{3}\,\xi, \quad L_2(\xi) = \frac{\sqrt{5}}{2}(3\xi^2 - 1). \tag{2.40}$$

The tensor product Legendre polynomials L_ν are defined as in (2.37) for $\nu \in \mathfrak{F}$.

In this case, the measure space (Γ, Σ, μ) is a countable product of identical factors $([-1, 1], \mathcal{B}([-1, 1]), \frac{1}{2}\,\mathrm{d}\xi)$,

$$\Gamma = [-1, 1]^\infty, \quad \Sigma = \mathcal{B}([-1, 1])^\infty = \mathcal{B}([-1, 1]^\infty), \quad \mu = \bigotimes_{m\in\mathbb{N}} \mu_m, \tag{2.41}$$

with $\mu_m(\mathrm{d}\xi) = \frac{1}{2}\,\mathrm{d}\xi$ for all $m \in \mathbb{N}$. Note that even though each μ_m is absolutely continuous with respect to the Lebesgue measure, *i.e.*, it has a density, the product of these densities is zero. Also, the countable product of the Lebesgue measure on $[-1,1]$ is not well-defined since the factors are not normalized. Thus μ cannot be defined via a density function.

Corollary 2.16. The tensor product Legendre polynomials $(L_\nu)_{\nu\in\mathfrak{F}}$ form an orthonormal basis of $L^2([-1,1]^\infty, \mathcal{B}([-1,1]^\infty), \mu)$.

Proof. The claim follows from Theorem 2.12 and Proposition 2.15. □

We shall refer to $(L_\nu)_{\nu\in\mathfrak{F}}$ as the *Legendre chaos basis*.

2.3. PDEs with uniform stochastic parameters

Parametric and stochastic operators

Let V be a separable real Hilbert space with dual V', and let $\langle\cdot,\cdot\rangle$ denote the (V', V)-duality pairing. We consider operator equations of the form

$$Au = f, \tag{2.42}$$

with $f \in V'$ and $A \in \mathcal{L}(V, V')$ a bounded linear operator from V to V'. If A is boundedly invertible, then (2.42) has the unique solution $u = A^{-1}f$.

Let Γ be a topological space. A parametric operator from V to V' is given by a continuous map

$$A : \Gamma \to \mathcal{L}(V, V'). \tag{2.43}$$

We assume that $A(y)$ is boundedly invertible for all $y \in \Gamma$, and consider the parametric operator equation

$$A(y)u(y) = f(y) \quad \forall y \in \Gamma \tag{2.44}$$

for a map $f : \Gamma \to V'$.

Proposition 2.17. Equation (2.44) has a unique solution $u : \Gamma \to V$. It is continuous if and only if $f : \Gamma \to V'$ is continuous.

Proof. Since $A(y)$ is boundedly invertible for all $y \in \Gamma$, (2.44) has the unique solution $u(y) = A(y)^{-1}f(y)$. If u is continuous in y, then $f(y) = A(y)u(y)$ is also continuous in y, since A is continuous by definition and application of an operator to a vector is continuous on $\mathcal{L}(V, V') \times V$. Furthermore, the map $y \mapsto A(y)^{-1}$ is continuous as a consequence of the abstract property (Kadison and Ringrose 1997, Proposition 3.1.6) of Banach algebras, so continuity of u follows from continuity of f by the same argument as above. □

We derive a weak formulation of (2.44) in the parameter y under the additional assumptions that $A(y)$ is symmetric positive definite for all $y \in \Gamma$,

and there exist constants $\widehat{c}$ and $\check{c}$ such that

$$\|A(y)\|_{V\to V'} \le \widehat{c}, \quad \|A(y)^{-1}\|_{V'\to V} \le \check{c} \quad \forall y \in \Gamma, \tag{2.45}$$

i.e., the bilinear form ${}_{V'}\langle A(y)\cdot,\cdot\rangle_V$ is a scalar product on V that induces a norm equivalent to $\|\cdot\|_V$. The estimates (2.45) always hold if Γ is compact.

Let μ be a probability measure on the Borel-measurable space $(\Gamma, \mathcal{B}(\Gamma))$. Then the operator $A(y)$ becomes stochastic in the sense that it depends on a parameter y in a probability space $(\Gamma, \mathcal{B}(\Gamma), \mu)$. Similarly, f is a random variable on Γ with values in V'. We assume

$$f \in L^2(\Gamma, \mu; V'). \tag{2.46}$$

Multiplying (2.44) by a test function $v : \Gamma \to V$ and integrating over Γ, we formally derive the linear variational problem

$$\int_\Gamma \langle A(y)u(y), v(y)\rangle \mu(\mathrm{d}y) = \int_\Gamma \langle f(y), v(y)\rangle \mu(\mathrm{d}y) \tag{2.47}$$

as the weak formulation of (2.44). By (2.45) and (2.46), both integrals are well-defined if $v \in L^2(\Gamma, \mu; V)$.

Theorem 2.18. Under the conditions (2.45) and (2.46), the solution u of (2.44) is the unique element of $L^2(\Gamma, \mu; V)$ satisfying (2.47) for all $v \in L^2(\Gamma, \mu; V)$. Furthermore,

$$\|u\|_{L^2(\Gamma,\mu;V)} \le \check{c}\|f\|_{L^2(\Gamma,\mu;V')}. \tag{2.48}$$

Proof. We first show that there is a unique $\widetilde{u} \in L^2(\Gamma, \mu; V)$ such that

$$\int_\Gamma \langle A(y)\widetilde{u}(y), v(y)\rangle \mu(\mathrm{d}y) = \int_\Gamma \langle f(y), v(y)\rangle \mu(\mathrm{d}y) \quad \forall v \in L^2(\Gamma, \mu; V). \tag{$*$}$$

By Cauchy–Schwarz and (2.45), for all $v, w \in L^2(\Gamma, \mu; V)$,

$$\begin{aligned}\left|\int_\Gamma \langle A(y)w(y), v(y)\rangle \mu(\mathrm{d}y)\right| &\le \int_\Gamma \widehat{c}\|w(y)\|_V\|v(y)\|_V \mu(\mathrm{d}y)\\ &\le \widehat{c}\|v\|_{L^2(\Gamma,\mu;V)}\|w\|_{L^2(\Gamma,\mu;V)}.\end{aligned}$$

Let $R : V \to V'$ denote the Riesz isomorphism. By positivity of $A(y)$, there is a unique positive $S(y) \in \mathcal{L}(V)$ such that $A(y) = RS(y)S(y)$ for all $y \in \Gamma$, and $\|S(y)\| \le \sqrt{\widehat{c}}$. Furthermore, $S(y)$ is invertible for all $y \in \Gamma$ and $\|S(y)^{-1}\| \le \sqrt{\check{c}}$. Consequently, for all $v \in L^2(\Gamma, \mu; V)$,

$$\int_\Gamma \langle A(y)v(y), v(y)\rangle \mu(\mathrm{d}y) = \int_\Gamma \|S(y)v(y)\|_V^2 \mu(\mathrm{d}y) \ge \check{c}^{-1}\|v\|^2_{L^2(\Gamma,\mu;V)}.$$

Similarly, by Cauchy–Schwarz,

$$\int_\Gamma \langle f(y), v(y)\rangle \mu(\mathrm{d}y) \le \|f\|_{L^2(\Gamma,\mu;V')}\|v\|_{L^2(\Gamma,\mu;V)}$$

for all $v \in L^2(\Gamma,\mu;V)$. The Lax–Milgram lemma implies existence and uniqueness of the solution $\widetilde{u}$ of $(*)$, and (2.48) for $\widetilde{u}$. By $(*)$ with $v(y) = v_0 1_E(y)$ for $v_0 \in V$ and $E \in \mathcal{B}(\Gamma)$,

$$\int_E \langle A(y)\widetilde{u}(y) - f(y), v_0\rangle \mu(\mathrm{d}y) = 0.$$

Since this holds for all measurable sets E, the integrand is 0 a.e. in Γ for any $v_0 \in V$, and therefore $\widetilde{u}$ satisfies (2.44) for μ-a.e. $y \in \Gamma$. This implies $\widetilde{u} = u$ in $L^2(\Gamma,\mu;V)$. □

Remark 2.19. (tensor product structure) For any separable Hilbert space X, the Lebesgue–Bochner space $L^2(\Gamma,\mu;X)$ is isometrically isomorphic to the Hilbert tensor product $L^2(\Gamma,\mu)\otimes X$. In particular, the solution u of (2.45) can be interpreted as an element of $L^2(\Gamma,\mu)\otimes V$, and f can be seen as an element of $L^2(\Gamma,\mu)\otimes V'$. Theorem 2.18 implies that the stochastic operator A induces an isomorphism between $L^2(\Gamma,\mu)\otimes V$ and $L^2(\Gamma,\mu)\otimes V'$, whose inverse maps f onto u.

The diffusion equation with a stochastic diffusion coefficient

Let D be a bounded Lipschitz domain in $\mathbb{R}^d$, and $(\Omega,\Sigma,\mathbb{P})$ a probability space. We consider as a model problem the isotropic diffusion equation on D with a stochastic diffusion coefficient and, for simplicity, homogeneous Dirichlet boundary conditions,

$$\begin{aligned} -\nabla\cdot\big(a(\omega,x)\nabla U(\omega,x)\big) = f(x), \quad x\in D, \quad \omega\in\Omega,\\ U(\omega,x) = 0, \quad x\in\partial D, \quad \omega\in\Omega. \end{aligned} \tag{2.49}$$

The differential operators in (2.49) are understood with respect to the physical variable $x \in D$. We assume there are constants a_- and a_+ such that

$$0 < a_- \le a(\omega,x) \le a_+ < \infty \quad \forall x\in D, \quad \forall\omega\in\Omega. \tag{2.50}$$

Furthermore, we select some deterministic approximation $\bar{a} \in L^\infty(D)$ to the stochastic diffusion coefficient $a(\cdot,\cdot)$. For example, $\bar{a}$ could be the mean field,

$$\bar{a}(x) := \int_\Omega a(\omega,x)\,\mathbb{P}(\mathrm{d}\omega), \quad x\in D, \tag{2.51}$$

or simply a constant such as $\bar{a} := (a_+ + a_-)/2$, $\bar{a} := \sqrt{a_+ a_-}$ or $\bar{a} := 1$.

We consider a series expansion of the difference $a(\omega,x) - \bar{a}(x)$. Let $(\varphi_m)_{m\in\mathbb{N}} \subset L^\infty(D)$ be a biorthogonal basis of $L^2(D)$ with associated dual basis $(\widetilde{\varphi}_m)_{m\in\mathbb{N}} \subset L^2(D)$, *i.e.*,

$$\langle \varphi_m, \widetilde{\varphi}_n\rangle_{L^2(D)} = \delta_{mn} \quad \forall m,n\in\mathbb{N}, \tag{2.52}$$

and

$$v = \sum_{m=1}^{\infty} \langle v, \widetilde{\varphi}_m \rangle_{L^2(D)} \varphi_m \quad \forall v \in L^2(D), \tag{2.53}$$

with unconditional convergence in $L^2(D)$. For a positive sequence $(\alpha_m)_{m\in\mathbb{N}}$, to be determined below, we define the random variables

$$Y_m(\omega) := \frac{1}{\alpha_m} \int_D (a(\omega, x) - \bar{a}(x)) \widetilde{\varphi}_m(x)\, \mathrm{d}x, \quad m \in \mathbb{N}. \tag{2.54}$$

By (2.53), for all $\omega \in \Omega$,

$$a(\omega, x) = \bar{a}(x) + \sum_{m=1}^{\infty} Y_m(\omega) \alpha_m \varphi_m(x), \tag{2.55}$$

with unconditional convergence in $L^2(D)$.

Lemma 2.20. There is a positive sequence $(\alpha_m)_{m\in\mathbb{N}}$ such that $Y_m(\omega) \in [-1, 1]$ for all $\omega \in \Omega$ and all $m \in \mathbb{N}$.

Proof. By Hölder's inequality,

$$\left| \int_D (a(\omega, x) - \bar{a}(x)) \widetilde{\varphi}_m(x)\, \mathrm{d}x \right| \le \|a(\omega, \cdot) - \bar{a}\|_{L^\infty(D)} \|\widetilde{\varphi}_m\|_{L^1(D)}.$$

Due to (2.50), the first term is bounded independently of ω, and we can choose

$$\alpha_m := \sup_{\omega\in\Omega} \|a(\omega, \cdot) - \bar{a}\|_{L^\infty(D)} \|\widetilde{\varphi}_m\|_{L^1(D)}. \qquad \square$$

Motivated by Lemma 2.20, we define as a parameter domain the compact topological space

$$\Gamma := [-1, 1]^\infty = \prod_{m=1}^{\infty} [-1, 1]. \tag{2.56}$$

Let $(\alpha_m)_{m\in\mathbb{N}}$ be a sequence as in Lemma 2.20. We assume that the series

$$\sum_{m=1}^{\infty} \alpha_m |\varphi_m(x)| \tag{2.57}$$

converges in $L^\infty(D)$, *i.e.*,

$$\lim_{M\to\infty} \operatorname*{ess\,sup}_{x\in D} \sum_{m=M}^{\infty} \alpha_m |\varphi_m(x)| = 0. \tag{2.58}$$

Then

$$a_\varphi(y, x) := \bar{a}(x) + \sum_{m=1}^{\infty} y_m \alpha_m \varphi_m(x), \quad y = (y_m)_{m\in\mathbb{N}} \in \Gamma, \quad x \in D, \tag{2.59}$$

converges uniformly in $L^\infty(D)$, and the stochastic diffusion coefficient satisfies

$$a(\omega, x) = a_\varphi(Y(\omega), x) \quad \forall x \in D, \quad \forall \omega \in \Omega, \tag{2.60}$$

where $Y(\omega) := (Y_m(\omega))_{m\in\mathbb{N}} \in \Gamma$.

We define the operators $A(y), \bar{A}, A_m : H_0^1(D) \to H^{-1}(D)$ by

$$ {}_{H^{-1}}\langle A(y)v, w\rangle_{H_0^1} := \int_D a_\varphi(y, x)\nabla v(x) \cdot \nabla w(x)\,\mathrm{d}x, \quad y \in \Gamma, \tag{2.61}$$

$$ {}_{H^{-1}}\langle \bar{A}v, w\rangle_{H_0^1} := \int_D \bar{a}(x)\nabla v(x) \cdot \nabla w(x)\,\mathrm{d}x, \tag{2.62}$$

$$ {}_{H^{-1}}\langle A_m v, w\rangle_{H_0^1} := \int_D \alpha_m\varphi_m(x)\nabla v(x) \cdot \nabla w(x)\,\mathrm{d}x, \quad m \in \mathbb{N}, \tag{2.63}$$

for all $v, w \in H_0^1(D)$. By (2.60), $A(y)$ is the operator associated to (2.49) for all $\omega \in \Omega$. Therefore,

$$U(\omega) = u(Y(\omega)) \quad \forall \omega \in \Omega, \tag{2.64}$$

for the solution u of (2.44) for (2.61), provided it exists.

Lemma 2.21. Under condition (2.58),

$$A(y) = \bar{A} + \sum_{m=1}^{\infty} y_m A_m, \quad y \in \Gamma, \tag{2.65}$$

with convergence in $\mathcal{L}(H_0^1(D), H^{-1}(D))$ uniformly in y. Furthermore, $A(y)$ depends continuously on $y \in \Gamma$.

Proof. Let $y \in \Gamma$ and $v, w \in H_0^1(D)$. By (2.59) and Fubini's theorem, using (2.58),

$$\langle A(y)v, w\rangle = \langle \bar{A}v, w\rangle + \sum_{m=1}^{\infty} y_m\langle A_m v, w\rangle.$$

Similarly, for all $M \in \mathbb{N}$, using $|y_m| \le 1$ for all $m \in \mathbb{N}$,

$$\begin{aligned}\left|\sum_{m=M}^{\infty} y_m\langle A_m v, w\rangle\right| &= \left|\int_D \left(\sum_{m=M}^{\infty} y_m\alpha_m\varphi_m(x)\right)\nabla v(x) \cdot \nabla w(x)\,\mathrm{d}x\right| \\ &\le \operatorname*{ess\,sup}_{x\in D}\left(\sum_{m=M}^{\infty} \alpha_m|\varphi_m(x)|\right)\|v\|_{H_0^1}\|w\|_{H_0^1}.\end{aligned}$$

Convergence of the series in $\mathcal{L}(H_0^1(D), H^{-1}(D))$ follows with (2.58).

A sequence $(y^n)_{n\in\mathbb{N}} \subset \Gamma$ converges to $y \in \Gamma$ if $y_m^n \to y_m$ for all $m \in \mathbb{N}$. In this case, $A(y^n) \to A(y)$ in $\mathcal{L}(H_0^1(D), H^{-1}(D))$ since, as above, using $|y_m^n - y_m| \le 2$,

$$\left\| \sum_{m=M}^{\infty} (y_m^n - y_m) A_m \right\|_{H_0^1(D)\to H^{-1}(D)} \le 2 \operatorname*{ess\,sup}_{x\in D} \left(\sum_{m=M}^{\infty} \alpha_m |\varphi_m(x)| \right).$$

The right-hand side is independent of n, and can be made smaller than ϵ for sufficiently large $M \in \mathbb{N}$. Then

$$\|A(y^n) - A(y)\|_{H_0^1(D)\to H^{-1}(D)} \le \epsilon + \sum_{m=1}^{M-1} |y_m^n - y_m| \|A_m\|_{H_0^1(D)\to H^{-1}(D)},$$

which is less than 2ϵ for sufficiently large $n \in \mathbb{N}$. □

We assume that the bilinear form associated to the operator $\bar{A}$ is coercive on $H_0^1(D)$, or equivalently, that

$$\exists \bar{a}_- : \quad \operatorname*{ess\,inf}_{x\in D} \bar{a}(x) \ge \bar{a}_- > 0. \tag{2.66}$$

Proposition 2.22. If

$$\gamma := \frac{1}{\bar{a}_-} \operatorname*{ess\,sup}_{x\in D} \sum_{m=1}^{\infty} \alpha_m |\varphi_m(x)| < 1, \tag{2.67}$$

then $A(y) : H_0^1(D) \to H^{-1}(D)$ is boundedly invertible for all $y \in \Gamma$, and

$$\sup_{y\in\Gamma} \|A(y)^{-1}\|_{H^{-1}(D)\to H_0^1(D)} \le \frac{\bar{a}_-^{-1}}{1-\gamma}. \tag{2.68}$$

Furthermore, $A(y)$ is bounded with

$$\sup_{y\in\Gamma} \|A(y)\|_{H_0^1(D)\to H^{-1}(D)} \le \|\bar{a}\|_{L^\infty(D)}(1+\gamma). \tag{2.69}$$

Proof. The operator $\bar{A} : H_0^1(D) \to H^{-1}(D)$ is invertible due to (2.66) and the Lax–Milgram lemma. The norm of its inverse is bounded by $1/\bar{a}_-$. By (2.67), as in the proof of Lemma 2.21,

$$\|\bar{A}^{-1}(\bar{A} - A(y))\|_{H_0^1(D)\to H_0^1(D)} \le \frac{1}{\bar{a}_-} \operatorname*{ess\,sup}_{x\in D} \sum_{m=1}^{\infty} \alpha_m |\varphi_m(x)| = \gamma < 1.$$

Therefore,

$$I - \bar{A}^{-1}(\bar{A} - A(y)) = \bar{A}^{-1} A(y)$$

is invertible by a Neumann series and has norm less than $(1-\gamma)^{-1}$. The claim follows by multiplying from the left by $\bar{A}$. □

Discretization by the Legendre chaos basis

For a separable Hilbert space V, we consider a parametric operator in $\mathcal{L}(V, V')$ of the form

$$A(y) = \bar{A} + \sum_{m=1}^{\infty} y_m A_m, \quad y \in \Gamma = [-1,1]^{\infty}, \tag{2.70}$$

with $\bar{A}, A_m \in \mathcal{L}(V, V')$ and convergence in $\mathcal{L}(V, V')$ uniformly in $y \in \Gamma$. As in Section 2.3, we assume that $A(y)$ is positive and boundedly invertible for all y, depends continuously on $y \in \Gamma$, and satisfies (2.45). By Proposition 2.22, this holds for (2.61) under the assumptions of Section 2.3. We make the additional assumption

$$\sum_{m=1}^{\infty} \|A_m\|_{V \to V'} < \infty, \tag{2.71}$$

which is stronger than (2.58) in the setting of Section 2.3.

Let the measure μ on $(\Gamma, \mathcal{B}(\Gamma))$ be the countable product of uniform measures on $[-1,1]$ as in (2.41). Then, by Corollary 2.16, the tensor product Legendre polynomials $(L_\nu)_{\nu \in \mathfrak{F}}$ form an orthonormal basis of $L^2(\Gamma, \mu)$, called the Legendre chaos basis. We use it to discretize the parameter domain Γ, *i.e.*, to reformulate (2.44) and (2.47) as an equation on a space of sequences in V.

By Remark 2.19, the parametric operator $A(y)$ induces a boundedly invertible operator between the Hilbert tensor product spaces $L^2(\Gamma, \mu) \otimes V$ and $L^2(\Gamma, \mu) \otimes V'$. The structure of (2.70) carries over to this operator. For all $m \in \mathbb{N}$, we define the multiplication operator

$$M_{y_m} : L^2(\Gamma, \mu) \to L^2(\Gamma, \mu), \quad g(y) \mapsto y_m g(y). \tag{2.72}$$

It follows from $y_m \in [-1,1]$ that M_{y_m} is self-adjoint and

$$\|M_{y_m}\|_{L^2(\Gamma,\mu) \to L^2(\Gamma,\mu)} = 1, \quad m \in \mathbb{N}. \tag{2.73}$$

Proposition 2.23. The operator in $\mathcal{L}(L^2(\Gamma, \mu) \otimes V, L^2(\Gamma, \mu) \otimes V')$ induced by $A(y)$ via (2.47), as in Remark 2.19, is

$$\mathcal{A} = I \otimes \bar{A} + \sum_{m=1}^{\infty} M_{y_m} \otimes A_m, \tag{2.74}$$

where I is the identity on $L^2(\Gamma, \mu)$. The sum in (2.74) converges unconditionally in $\mathcal{L}(L^2(\Gamma, \mu) \otimes V, L^2(\Gamma, \mu) \otimes V')$.

Proof. The operator $\mathcal{A}$ is well-defined by (2.74) since, by (2.73),

$$\left\| \sum_{m=M}^{N} M_{y_m} \otimes A_m \right\| \leq \sum_{m=M}^{N} \|M_{y_m}\| \|A_m\| = \sum_{m=M}^{N} \|A_m\|,$$

which can be made arbitrarily small for sufficiently large M by (2.71). The convergence is unconditional since the convergence of (2.71) is unconditional.

Let $g \in L^2(\Gamma,\mu)$ and $v \in V$. Then using (2.72),

$$\mathcal{A}(g \otimes v)(y) = g(y)\bar{A}v + \sum_{m=1}^{\infty} y_m g(y) A_m v = A(y)(g(y)v), \quad y \in \Gamma.$$

Therefore, $\mathcal{A}$ is the operator induced by $A(y)$. □

Since the tensor product Legendre polynomials $(L_\nu)_{\nu\in\mathfrak{F}}$ form an orthonormal basis of $L^2(\Gamma,\mu)$, the map

$$T_L : \ell^2(\mathfrak{F}) \to L^2(\Gamma,\mu), \quad (c_\nu)_{\nu\in\mathfrak{F}} \mapsto \sum_{\nu\in\mathfrak{F}} c_\nu L_\nu, \tag{2.75}$$

is a unitary isomorphism by Parseval's identity. Tensorizing with the identity I_V on V, we get the isometric isomorphism

$$T_L \otimes I_V : \ell^2(\mathfrak{F}) \otimes V \to L^2(\Gamma,\mu) \otimes V \tag{2.76}$$

with adjoint

$$(T_L \otimes I_V)' = T_L' \otimes I_{V'} : L^2(\Gamma,\mu) \otimes V' \to \ell^2(\mathfrak{F}) \otimes V'. \tag{2.77}$$

We define the semidiscrete operator

$$\mathfrak{A} := (T_L \otimes I_V)' \mathcal{A} (T_L \otimes I_V) : \ell^2(\mathfrak{F}) \otimes V \to \ell^2(\mathfrak{F}) \otimes V'. \tag{2.78}$$

Similarly, interpreting $f \in L^2(\Gamma,\mu;V')$ as an element of $L^2(\Gamma,\mu) \otimes V'$, we define

$$\mathfrak{f} := (T_L \otimes I_V)' f = \left(\int_\Gamma f(y) L_\nu(y) \mu(\mathrm{d}y) \right)_{\nu\in\mathfrak{F}}, \tag{2.79}$$

which is simply the sequence of Legendre coefficients of f. This leads to the semidiscrete operator equation

$$\mathfrak{A}\mathfrak{u} = \mathfrak{f}. \tag{2.80}$$

Theorem 2.24. The operator $\mathfrak{A}$ from (2.78) has the form

$$\mathfrak{A} = \boldsymbol{I} \otimes \bar{A} + \sum_{m=1}^{\infty} \boldsymbol{K}_m \otimes A_m, \tag{2.81}$$

with convergence in $\mathcal{L}(\ell^2(\mathfrak{F}) \otimes V, \ell^2(\mathfrak{F}) \otimes V')$, where $\boldsymbol{I}$ is the identity on $\ell^2(\mathfrak{F})$ and $\boldsymbol{K}_m := T_L' M_{y_m} T_L$. Furthermore, $\mathfrak{A}$ is boundedly invertible, and the solutions of (2.44) and (2.80) are related by

$$u = (T_L \otimes I_V)\mathfrak{u}. \tag{2.82}$$

Proof. Equation (2.81) follows from (2.74) since $T_L' = T_L^{-1}$. The operator $\mathfrak{A}$ is boundedly invertible since $\mathcal{A}$ is boundedly invertible by Remark 2.19 and $(T_L \otimes I_V)'$ and $(T_L \otimes I_V)$ are isomorphisms by definition. Applying the inverse of $(T_L \otimes I_V)'$ to (2.80) and inserting (2.78) and (2.79), it follows that

$$\mathcal{A}(T_L \otimes I_V)\mathfrak{u} = f,$$

which characterizes u by Theorem 2.18. □

Lemma 2.25. For all $m \in \mathbb{N}$, the operator

$$\boldsymbol{K}_m := T_L' M_{y_m} T_L : \ell^2(\mathfrak{F}) \to \ell^2(\mathfrak{F}) \tag{2.83}$$

has the form

$$\boldsymbol{K}_m(c_\nu)_{\nu\in\mathfrak{F}} = (\beta_{\nu_m+1} c_{\nu+\epsilon_m} + \beta_{\nu_m} c_{\nu-\epsilon_m})_{\nu\in\mathfrak{F}}, \tag{2.84}$$

where $\beta_0 := 0$, and

$$\beta_n := \frac{n}{\sqrt{2n+1}\sqrt{2n-1}} = \frac{1}{\sqrt{4-n^{-2}}} \in \left(\frac{1}{2}, \frac{1}{\sqrt{3}}\right], \quad n \in \mathbb{N}. \tag{2.85}$$

Here, ϵ_m is the Kronecker sequence $(\epsilon_m)_n := \delta_{mn}$, and if $\nu_m = 0$, the term $c_{\nu-\epsilon_m}$ is irrelevant in (2.84) since it is multiplied by $\beta_0 = 0$. Furthermore, $\boldsymbol{K}_m$ is self-adjoint and

$$\|\boldsymbol{K}_m\|_{\ell^2(\mathfrak{F})\to\ell^2(\mathfrak{F})} = 1, \quad m \in \mathbb{N}. \tag{2.86}$$

Proof. By definition, since $T_L^{-1} = T_L'$,

$$T_L \boldsymbol{K}_m (c_\nu)_{\nu\in\mathfrak{F}} = M_{y_m} T_L (c_\nu)_{\nu\in\mathfrak{F}} = \sum_{\nu\in\mathfrak{F}} c_\nu y_m L_\nu(y).$$

Therefore, (2.84) is equivalent to

$$y_m L_\nu(y) = \beta_{\nu_m+1} L_{\nu+\epsilon_m}(y) + \beta_{\nu_m} L_{\nu-\epsilon_m}(y).$$

By (2.38),

$$\xi L_n(\xi) = \beta_{n+1} L_{n+1}(\xi) + \beta_n L_{n-1}(\xi), \quad \xi \in [-1,1], \quad n \in \mathbb{N}_0.$$

Then the claim follows from (2.29). Note that $(\beta_n)_{n\in\mathbb{N}}$ is decreasing in n, $\beta_1 = 1/\sqrt{3}$, and $\beta_n \to 1/2$. □

Corollary 2.26. The solution u of (2.44) is

$$u(y) = \sum_{\nu\in\mathfrak{F}} u_\nu L_\nu(y) \in V, \quad y \in \Gamma, \tag{2.87}$$

with convergence in $L^2(\Gamma, \mu; V)$, where the coefficients $(u_\nu)_{\nu\in\mathfrak{F}} \in V$ are determined uniquely by the equations

$$\bar{A} u_\nu + \sum_{m=1}^{\infty} A_m(\beta_{\nu_m+1} u_{\nu+\epsilon_m} + \beta_{\nu_m} u_{\nu-\epsilon_m}) = f_\nu, \quad \nu \in \mathfrak{F}, \tag{2.88}$$

for $(\beta_n)_{n\in\mathbb{N}_0}$ and $(\epsilon_m)_{m\in\mathbb{N}}$ as in Lemma 2.25, and

$$f_\nu := \int_\Gamma f(y) L_\nu(y)\,\mu(\mathrm{d}y) \in V', \quad \nu \in \mathfrak{F}. \tag{2.89}$$

Proof. The claim follows from Theorem 2.24 using the definitions (2.75), (2.76), (2.78), (2.79), Lemma 2.25, and the identification of Lebesgue–Bochner spaces with Hilbert tensor product spaces as in Remark 2.19. □

Finite element approximation

The discretization from Section 2.3 does not include any approximations. The infinite system of equations in Corollary 2.26 determines the Legendre coefficients $(u_\nu)_{\nu\in\mathfrak{F}} \in V$ of the exact solution u of (2.44). However, this system of equations lends itself to discretization by standard finite elements.

For all $\nu \in \mathfrak{F}$, let $V_{N,\nu} \subset V$ be a finite-dimensional space. We assume that $V_{N,\nu} = \{0\}$ for all but finitely many $\nu \in \mathfrak{F}$ and define the finite-dimensional space

$$\mathcal{V}_N := \{v \in L^2(\Gamma; V)\,;\ v_\nu \in V_{N,\nu}\ \forall \nu \in \mathfrak{F}\}, \tag{2.90}$$

where $v_\nu \in V$ is the νth coefficient in the expansion of $v \in L^2(\Gamma; V)$ with respect to the tensor product Legendre polynomials $(L_\nu)_{\nu\in\mathfrak{F}}$. This space can be interpreted as a subspace of $L^2(\Gamma; V)$, as in (2.90), or as the space of sequences $(v_\nu)_{\nu\in\mathfrak{F}}$ in V with $v_\nu \in V_{N,\nu}$ for all $\nu \in \mathfrak{F}$, which is a subspace of $\ell^2(\mathfrak{F}; V)$. By Parseval's identity, the norms induced by these two spaces coincide.

Accordingly, the Galerkin projection of u onto $\mathcal{V}_N$ can be characterized in two equivalent ways. We define the Galerkin approximation u_N of u on $\mathcal{V}_N$ as the unique element of $\mathcal{V}_N$ satisfying

$$\int_\Gamma \langle A(y)u_N(y), v_N(y)\rangle \mu(\mathrm{d}y) = \int_\Gamma \langle f(y), v_N(y)\rangle \mu(\mathrm{d}y) \quad \forall v_N \in \mathcal{V}_N. \tag{2.91}$$

As in Corollary 2.26, the Legendre coefficients of u_N satisfy a system of equations

$$\langle \bar{A} u_{N,\nu}, v_N\rangle + \sum_{m=1}^{\infty} \langle A_m(\beta_{\nu_m+1} u_{N,\nu+\epsilon_m} + \beta_{\nu_m} u_{N,\nu-\epsilon_m}), v_N\rangle = \langle f_\nu, v_N\rangle \tag{2.92}$$

for all $v_N \in V_{N,\nu}$ and all $\nu \in \mathfrak{F}$. Since $V_{N,\nu} = \{0\}$ for all but finitely many $\nu \in \mathfrak{F}$, there are only finitely many non-trivial equations (2.92). Also, for the same reason, the sum in each equation is finite. Therefore, without any explicit truncation, the infinite system of equations (2.88) becomes a finite system when considered on a finite-dimensional space.

Proposition 2.27. The Galerkin projection u_N of u onto $\mathcal{V}_N$ is well-defined by (2.91), and satisfies

$$\|u_N\|_{L^2(\Gamma,\mu;V)} \le \check{c} \|f\|_{L^2(\Gamma,\mu;V')}. \tag{2.93}$$

Its Legendre coefficients $(u_{N,\mu})_{\mu\in\mathfrak{F}}$ are uniquely characterized by (2.92) for $\nu \in \mathfrak{F}$. Furthermore,

$$\|u - u_N\|_{L^2(\Gamma,\mu;V)} \le \sqrt{\widehat{c}\check{c}} \inf_{v_N\in\mathcal{V}_N} \|u - v_N\|_{L^2(\Gamma,\mu;V)}. \tag{2.94}$$

Proof. As shown in the proof of Theorem 2.18, the bilinear form in (2.91) is continuous with constant $\widehat{c}$ and coercive with constant $\check{c}^{-1}$. Existence and uniqueness of u_N as well as (2.93) follow from the Lax–Milgram lemma applied to the space $\mathcal{V}_N$. The equivalence of (2.91) and (2.92) is a consequence of Theorem 2.24, using Lemma 2.25. The quasi-optimality property (2.94) holds since the bilinear form in (2.91) is a scalar product, and the norm induced by it on $L^2(\Gamma,\mu;V)$ is equivalent to the standard norm with constants $\sqrt{\widehat{c}}$ and $\sqrt{\check{c}}$. □

Given a space $\mathcal{V}_N$, the Galerkin projection u_N of u onto $\mathcal{V}_N$ can be computed iteratively, for example by a conjugate gradient iteration; see Gittelson (2011*b*). The inverse of the deterministic operator $\bar{A}$ can be used as a preconditioner.

The sparse tensor product construction of $\mathcal{V}_N$, which amounts to a problem-adapted selection of finite element spaces $V_{N,\nu}$, is discussed in Section 4.1. Approximation results are presented in Section 3.1.

2.4. PDEs with Gaussian parameters

The log-normal diffusion equation

We consider again the diffusion equation (2.49) with a stochastic diffusion coefficient $a(\cdot,\cdot)$ on a bounded Lipschitz domain $D \subset \mathbb{R}^d$. However, instead of expanding $a(\cdot,\cdot)$ in a series, we expand its logarithm. More precisely, we take a series expansion of $\log(a - a_*)$, where a_* is a bounded function on D with $a_*(x) \ge 0$ for all $x \in D$. Then, instead of (2.59), we have a diffusion coefficient of the form

$$a(y,x) = a_*(x) + a_0(x)\exp\left(\sum_{m=1}^{\infty} y_m a_m(x)\right), \quad x \in D, \tag{2.95}$$

for $y = (y_m)_{m\in\mathbb{N}} \in \mathbb{R}^\infty$. We assume that the coefficients $(y_m)_{m\in\mathbb{N}}$ are independent standard Gaussian random variables. This is the case, for instance, if $\log(a - a_*)$ is Gaussian and we expand it in its Karhunen–Loève series, or more generally if $(a_m)_{m\in\mathbb{N}}$ are orthonormal in the Cameron–Martin space of the distribution of $\log(a-a_*)$: see Section 2.1 and Gittelson (2010*b*).

The diffusion equation with the stochastic coefficient (2.95) and, for simplicity, homogeneous Dirichlet boundary conditions, is

$$\begin{aligned} -\nabla\cdot\big(a(y,x)\nabla u(y,x)\big) &= f(y,x), \quad x \in D, \\ u(y,x) &= 0, \quad x \in \partial D. \end{aligned} \tag{2.96}$$

By the above assumptions, the parameter y is in the probability space $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty), \gamma)$, where

$$\gamma = \bigotimes_{m=1}^{\infty} N_1, \tag{2.97}$$

as in (2.4). For the sake of generality, we allow the forcing term f in (2.96) to depend on $y \in \mathbb{R}^\infty$.

The series in (2.95) may not converge for all $y \in \mathbb{R}^\infty$. We assume that $a_m \in L^\infty(D)$ for all $m \in \mathbb{N}_0$, $a_0(x) \geq \check{a}_0 > 0$ for all $x \in D$, and

$$\sum_{m=1}^{\infty} \|a_m\|_{L^\infty(D)} < \infty, \tag{2.98}$$

i.e., the sequence $\alpha_m := \|a_m\|_{L^\infty(D)}$, $m \in \mathbb{N}$, is in $\ell^1(\mathbb{N})$. Then the series in (2.95) converges in $L^\infty(D)$, at least for all y in the set

$$\Gamma := \left\{ y \in \mathbb{R}^\infty \,;\, \sum_{m=1}^{\infty} \alpha_m |y_m| < \infty \right\}. \tag{2.99}$$

Lemma 2.28. $\Gamma \in \mathcal{B}(\mathbb{R}^\infty)$ and $\gamma(\Gamma) = 1$.

Proof. Borel-measurability of Γ follows from

$$\Gamma = \bigcup_{N=1}^{\infty} \bigcap_{M=1}^{\infty} \left\{ y \in \mathbb{R}^\infty \,;\, \sum_{m=1}^{M} \alpha_m |y_m| \leq N \right\}.$$

By the monotone convergence theorem, using

$$\int_{\mathbb{R}^\infty} |y_m| \gamma(\mathrm{d}y) = \frac{2}{\sqrt{2\pi}} \int_0^\infty \xi \exp\left(-\frac{\xi^2}{2}\right) \mathrm{d}\xi = \sqrt{\frac{2}{\pi}},$$

it follows that

$$\int_{\mathbb{R}^\infty} \sum_{m=1}^{\infty} \alpha_m |y_m| \gamma(\mathrm{d}y) = \sum_{m=1}^{\infty} \alpha_m \int_{\mathbb{R}^\infty} |y_m| \gamma(\mathrm{d}y) = \sqrt{\frac{2}{\pi}} \sum_{m=1}^{\infty} \alpha_m < \infty,$$

which implies that the sum converges γ-a.e. on $\mathbb{R}^\infty$, and thus $\gamma(\Gamma) = 1$ by (2.99). □

Lemma 2.29. For all $y \in \Gamma$, the diffusion coefficient (2.95) is well-defined and satisfies

$$0 < \check{a}(y) := \operatorname*{ess\,inf}_{x \in D} a(y,x) \leq \operatorname*{ess\,sup}_{x \in D} a(y,x) =: \widehat{a}(y) < \infty \tag{2.100}$$

with

$$\widehat{a}(y) \le \|a_*\|_{L^\infty(D)} + \|a_0\|_{L^\infty(D)} \exp\left(\sum_{m=1}^{\infty} \alpha_m |y_m|\right),$$

$$\check{a}(y) \ge \operatorname*{ess\,inf}_{x\in D} a_*(x) + \check{a}_0 \exp\left(-\sum_{m=1}^{\infty} \alpha_m |y_m|\right).$$

Proof. Let $y \in \Gamma$ and $x \in D$ with $|a(x)| \le \alpha_m$ for all $m \in \mathbb{N}$. Then

$$\sum_{m=1}^{\infty} |a_m(x)||y_m| \le \sum_{m=1}^{\infty} \alpha_m |y_m| < \infty.$$

By continuity and positivity of $\exp(\cdot)$,

$$\exp\left(\sum_{m=1}^{\infty} a_m(x) y_m\right) = \prod_{m=1^\infty} \exp(a_m(x) y_m) \in (0, \infty). \tag{2.101}$$

Then the claim follows from (2.95). □

Due to Lemmas 2.28 and 2.29, we consider Γ as the parameter space of (2.96) instead of $\mathbb{R}^\infty$. Even though Γ is not a product domain, we can define product measures such as γ on Γ by restriction.

For each $y \in \Gamma$, we consider the weak formulation of (2.96) on $V := H_0^1(D)$ with norm

$$\|v\|_V := \left(\int_D |\nabla v(x)|^2 \,\mathrm{d}x\right)^{1/2}. \tag{2.102}$$

We define the bilinear form

$$b(y; w, v) := \int_D a(y, x) \nabla w(x) \cdot \nabla v(x) \,\mathrm{d}x, \quad w, v \in V, \tag{2.103}$$

and reinterpret the forcing term f as a map into the dual space V' by

$$f(y; v) := \int_D f(y, x) v(x) \,\mathrm{d}x, \quad v \in V, \tag{2.104}$$

for all $y \in \Gamma$. Then the weak formulation on V of the diffusion equation (2.96) is given by the linear variational problem of determining $u(y) \in V$ such that

$$b(y; u(y), v) = f(y; v) \quad \forall v \in V. \tag{2.105}$$

Theorem 2.30. For all $y \in \Gamma$, (2.105) has a unique solution $u(y) \in V$. It satisfies

$$\|u(y)\|_V \le \frac{1}{\check{a}(y)} \|f(y; \cdot)\|_{V'} \quad \forall y \in \Gamma. \tag{2.106}$$

Proof. By Lemma 2.29 and (2.102), the bilinear form $b(y; \cdot, \cdot)$ is continuous and coercive on V with coercivity constant $\check{a}(y)$ for all $y \in \Gamma$. Therefore, the claim follows by the Lax–Milgram lemma. □

Auxiliary Gaussian measures

For any sequence $\boldsymbol{\sigma} = (\sigma_m)_{m\in\mathbb{N}} \in \exp(\ell^1(\mathbb{N}))$, *i.e.*, $\sigma_m = \exp(s_m)$ with $(s_m)_m \in \ell^1(\mathbb{N})$, we define the product measure

$$\gamma_{\boldsymbol{\sigma}} := \bigotimes_{m=1}^{\infty} N_{\sigma_m^2} \tag{2.107}$$

on $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$, where $N_{\sigma_m^2}$ is the centred Gaussian measure on $\mathbb{R}$ with standard deviation σ_m. In particular, the standard Gaussian measure on $\mathbb{R}^\infty$ is $\gamma = \gamma_{\mathbf{1}}$.

Proposition 2.31. For all $\boldsymbol{\sigma} = (\sigma_m)_{m\in\mathbb{N}} \in \exp(\ell^1(\mathbb{N}))$, the measure $\gamma_{\boldsymbol{\sigma}}$ is equivalent to γ. The density of $\gamma_{\boldsymbol{\sigma}}$ with respect to γ is

$$\zeta_{\boldsymbol{\sigma}}(y) = \left(\prod_{m=1}^{\infty} \frac{1}{\sigma_m}\right) \exp\left(-\frac{1}{2}\sum_{m=1}^{\infty} (\sigma_m^{-2} - 1) y_m^2\right). \tag{2.108}$$

Proof. Note that $\mathrm{d}N_{\sigma_m^2} = \zeta_{\boldsymbol{\sigma},m}\,\mathrm{d}N_1$ for

$$\zeta_{\boldsymbol{\sigma},m}(y_m) = \frac{1}{\sigma_m}\,\exp\left(-\frac{1}{2}(\sigma_m^{-2} - 1) y_m^2\right).$$

We compute

$$\begin{aligned}\int_{\mathbb{R}} \sqrt{\zeta_{\boldsymbol{\sigma},m}(y_m)} N_1(\mathrm{d}y_m) &= \frac{1}{\sqrt{2\pi\sigma_m}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{4}(\sigma_m^{-2} + 1) y_m^2\right) \mathrm{d}y_m \\ &= \sqrt{\frac{2}{\sigma_m + \sigma^{-1}}} = \exp\left(\frac{1}{2}\,\beta_m\right)\end{aligned}$$

for some β_m with $|\beta_m| \le \log \sigma_m$. Therefore,

$$\prod_{m=1}^{\infty} \int_{\mathbb{R}} \sqrt{\zeta_{\boldsymbol{\sigma},m}(y_m)} N_1(\mathrm{d}y_m) = \exp\left(\frac{1}{2}\sum_{m=1}^{\infty} \beta_m\right),$$

which converges since $(\log \sigma_m)_m \in \ell^1(\mathbb{N})$. Then the claim follows by Theorem C.44. □

In particular, Proposition 2.31 implies that $\gamma_{\boldsymbol{\sigma}}(\Gamma) = 1$ for any $\boldsymbol{\sigma} \in \exp(\ell^1(\mathbb{N}))$. Therefore, the restriction of $\gamma_{\boldsymbol{\sigma}}$ to Γ is a probability measure.

We consider sequences $\boldsymbol{\sigma}$ that depend exponentially on $\boldsymbol{\alpha} = (\alpha_m)_{m\in\mathbb{N}}$,

$$\sigma_m(\chi) := \exp(\chi\alpha_m), \quad m \in \mathbb{N}, \quad \chi \in \mathbb{R}. \tag{2.109}$$

We abbreviate $\gamma_\chi := \gamma_{\boldsymbol{\sigma}(\chi)}$ and $\zeta_\chi := \zeta_{\boldsymbol{\sigma}(\chi)}$. In particular, $\gamma = \gamma_{\mathbf{1}} = \gamma_0$.

Lemma 2.32. Let $\eta < \chi$ and $k \ge 0$. Then for all $y \in \Gamma$,

$$\frac{\zeta_\eta(y)}{\zeta_\chi(y)} \exp\left(k \sum_{m=1}^{\infty} \alpha_m |y_m|\right) \le \exp\left(\left(\frac{k^2 \mathrm{e}^{2\chi\|\boldsymbol{\alpha}\|_{\ell^\infty}}}{4(\chi - \eta)} + \chi - \eta\right)\|\boldsymbol{\alpha}\|_{\ell^1}\right). \tag{2.110}$$

Proof. Let $y \in \Gamma$ and abbreviate $\sigma_m := \mathrm{e}^{\alpha_m}$. By (2.108),

$$\frac{\zeta_\eta(y)}{\zeta_\chi(y)} = \left(\prod_{m=1}^{\infty} \sigma_m^{\chi-\eta}\right) \exp\left(\frac{1}{2}\sum_{m=1}^{\infty}(\sigma_m^{-2(\chi-\eta)} - 1)\sigma_m^{-2\eta} y_m^2\right).$$

Using the estimate

$$\begin{aligned}(\sigma_m^{-2(\chi-\eta)} - 1)\sigma_m^{-2\eta} &= (\mathrm{e}^{-2(\chi-\eta)\alpha_m} - 1)\mathrm{e}^{-2\eta\alpha_m} \\ &= \mathrm{e}^{-2\chi\alpha_m}(1 - \mathrm{e}^{2(\chi-\eta)\alpha_m}) \\ &\le \mathrm{e}^{-2\chi\alpha_m}(-(\chi-\eta)\alpha_m),\end{aligned}$$

we have

$$\begin{aligned}&\log\left(\frac{\zeta_\eta(y)}{\zeta_\chi(y)} \exp\left(k\sum_{m=1}^{\infty}\alpha_m|y_m|\right)\right) \\ &\quad= k\sum_{m=1}^{\infty}\alpha_m|y_m| + \frac{1}{2}\sum_{m=1}^{\infty}(\sigma_m^{-2(\chi-\eta)} - 1)\sigma_m^{-2\eta}y_m^2 + (\chi-\eta)\sum_{m=1}^{\infty}\log\sigma_m \\ &\quad\le k\sum_{m=1}^{\infty}\alpha_m|y_m| - (\chi-\eta)\sum_{m=1}^{\infty}\alpha_m\mathrm{e}^{-2\chi\alpha_m}y_m^2 + (\chi-\eta)\sum_{m=1}^{\infty}\alpha_m \\ &\quad= -\sum_{m=1}^{\infty}\alpha_m\left(\sqrt{\chi-\eta}\,\mathrm{e}^{-\chi\alpha_m}|y_m| - \frac{k\,\mathrm{e}^{\chi\alpha_m}}{2\sqrt{\chi-\eta}}\right)^2 \\ &\qquad+ \sum_{m=1}^{\infty}\frac{\alpha_m k^2\mathrm{e}^{2\chi\alpha_m}}{4(\chi-\eta)} + (\chi-\eta)\sum_{m=1}^{\infty}\alpha_m \\ &\quad\le \sum_{m=1}^{\infty}\left(\frac{k^2\mathrm{e}^{2\chi\alpha_m}}{4(\chi-\eta)} + (\chi-\eta)\right)\alpha_m.\end{aligned}$$

□

In particular, if $k = 0$, then (2.110) reads

$$\frac{\zeta_\eta(y)}{\zeta_\chi(y)} \le \exp\big((\chi-\eta)\|\boldsymbol{\alpha}\|_{\ell^1}\big). \tag{2.111}$$

Proposition 2.33. Let $0 < p < \infty$ and $\eta < \chi$. Then

$$L^p(\Gamma, \gamma_\chi) \subset L^p(\Gamma, \gamma_\eta) \tag{2.112}$$

and

$$\|v\|_{L^p(\Gamma,\gamma_\eta)} \le \exp\left(\frac{\chi-\eta}{p}\|\boldsymbol{\alpha}\|_{\ell^1}\right)\|v\|_{L^p(\Gamma,\gamma_\chi)} \quad \forall v \in L^p(\Gamma, \gamma_\chi). \tag{2.113}$$

Proof. Let $v \in L^p(\Gamma, \gamma_\chi)$. Then

$$\|v\|^p_{L^p(\Gamma,\gamma_\eta)} = \int_\Gamma v^p \,\mathrm{d}\gamma_\eta = \int_\Gamma v^p \frac{\zeta_\eta}{\zeta_\chi}\,\mathrm{d}\gamma_\chi \le \sup_{y\in\Gamma} \frac{\zeta_\eta(y)}{\zeta_\chi(y)} \|v\|^p_{L^p(\Gamma,\gamma_\chi)}.$$

The claim follows from Lemma 2.32 with $k = 0$: see (2.111). □

Of course, Proposition 2.33 also applies to Lebesgue–Bochner spaces of functions mapping, for example, into V or V'. We will use it with $\eta = 0$, such that $\gamma_\eta = \gamma$.

Integrability of the solution

We consider integrability properties of the solution u of (2.105). Borel-measurability of the map $\mathbb{R}^\infty \ni y \mapsto u(y) \in V$ is shown in Gittelson (2010*a*, Lemma 3.4) under the assumption that f is Borel-measurable as a map from $\mathbb{R}^\infty$ to V'. Under stronger assumptions, measurability of u also follows from Theorem 2.44 below.

Proposition 2.34. Let $0 < p < \infty$ and $\varrho > 0$. If $f \in L^p(\Gamma, \gamma_\varrho; V')$, then the solution u of (2.105) is in $L^p(\Gamma, \gamma; V)$ and satisfies

$$\|u\|_{L^p(\Gamma,\gamma;V)} \le \bar{c}_{\varrho,p} \|f\|_{L^p(\Gamma,\gamma_\varrho;V')}$$

with

$$\bar{c}_{\varrho,p} = \min\left\{ \frac{\exp\left(\frac{\varrho}{p}\|\boldsymbol{\alpha}\|_{\ell^1}\right)}{\operatorname{ess\,inf}_{y\in\Gamma} a_*(y)}\,,\ \frac{1}{\check{a}_0} \exp\left(\left(\frac{p\,\mathrm{e}^{2\varrho\|\boldsymbol{\alpha}\|_{\ell^\infty}}}{4\varrho} + \frac{\varrho}{p}\right)\|\boldsymbol{\alpha}\|_{\ell^1}\right)\right\}.$$

Proof. By (2.106),

$$\begin{aligned}\int_\Gamma \|u(y)\|^p_V \gamma(\mathrm{d}y) &\le \int_\Gamma \zeta_\varrho(y)^{-1} \check{a}(y)^{-p} \|f(y;\cdot)\|^p_{V'} \gamma\varrho(\mathrm{d}y) \\ &\le \left(\operatorname*{ess\,inf}_{y\in\Gamma} \zeta_\varrho(y)^{-1}\check{a}(y)^{-p}\right) \int_\Gamma \|f(y;\cdot)\|^p_{V'} \gamma\varrho(\mathrm{d}y).\end{aligned}$$

The claim follows from Lemmas 2.29 and 2.32 with $\eta = 0$, $\chi = \varrho$ and $k = p$. □

However, we also need integrability of u with respect to the measure γ_ϱ.

Lemma 2.35. For all $\varrho \ge 0$ and all $0 < r < \infty$,

$$\exp\left(\sum_{m=1}^\infty \alpha_m |y_m|\right) \in L^r(\Gamma, \gamma_\varrho)$$

with

$$\left\|\exp\left(\sum_{m=1}^\infty \alpha_m|y_m|\right)\right\|_{L^r(\Gamma,\gamma_\varrho)} \le \exp\left(\frac{r}{2}\mathrm{e}^{2\varrho\|\boldsymbol{\alpha}\|_{\ell^\infty}}\|\boldsymbol{\alpha}\|^2_{\ell^2} + \sqrt{\frac{2}{\pi}}\mathrm{e}^{\varrho\|\boldsymbol{\alpha}\|_{\ell^\infty}}\|\boldsymbol{\alpha}\|_{\ell^1}\right).$$

Proof. The claim follows from Gittelson (2010*a*, Lemma 3.10) with the change of variables $z_m := \mathrm{e}^{-\varrho\alpha_m} y_m$. □

Theorem 2.36. Let $0 < q < p < \infty$ and $\varrho \geq 0$. If $f \in L^p(\Gamma, \gamma_\varrho; V')$, then the solution u of (2.105) is in $L^q(\Gamma, \gamma_\varrho; V)$ and satisfies

$$\|u\|_{L^q(\Gamma,\gamma_\varrho;V)} \leq \widetilde{c}_{\varrho,q,p} \|f\|_{L^p(\Gamma,\gamma_\varrho;V')}$$

with

$$\widetilde{c}_{\varrho,q,p} = \frac{1}{\check{a}_0} \exp\left(\frac{qp\,\mathrm{e}^{2\varrho\|\boldsymbol{\alpha}\|_{\ell^\infty}}}{2(p-q)} \|\boldsymbol{\alpha}\|_{\ell^2}^2 + \sqrt{\frac{2}{\pi}} \mathrm{e}^{\varrho\|\boldsymbol{\alpha}\|_{\ell^\infty}} \|\boldsymbol{\alpha}\|_{\ell^1} \right),$$

or, if $\operatorname{ess\,inf}_{y\in\Gamma} a_*(y) > 0$ and $q \leq p$, also with

$$\widetilde{c}_{\varrho,q,p} = \frac{1}{\operatorname{ess\,inf}_{y\in\Gamma} a_*(y)}.$$

Proof. Let $r = \frac{qp}{p-q}$. By (2.106) and Hölder's inequality,

$$\begin{aligned} \int_\Gamma \|u(y)\|_V^q \gamma_\varrho(\mathrm{d}y) &\leq \int_\Gamma \check{a}(y)^{-q} \|f(y;\cdot)\|_{V'}^q \gamma_\varrho(\mathrm{d}y) \\ &\leq \|\check{a}(\cdot)^{-1}\|_{L^r(\Gamma,\gamma_\varrho)}^q \|f\|_{L^p(\Gamma,\gamma_\varrho;V')}^q. \end{aligned}$$

Then the claim follows from Lemmas 2.29 and 2.35. □

In particular, if $f \in L^p(\Gamma, \gamma_\varrho; V')$ with $p > 2$, then $u \in L^2(\Gamma, \gamma_\varrho; V)$ and

$$\|u\|_{L^2(\Gamma,\gamma_\varrho;V)} \leq \widetilde{c}_{\varrho,p} \|f\|_{L^p(\Gamma,\gamma_\varrho;V')} \tag{2.114}$$

with

$$\widetilde{c}_{\varrho,p} = \frac{1}{\check{a}_0} \exp\left(\frac{p\,\mathrm{e}^{2\varrho\|\boldsymbol{\alpha}\|_{\ell^\infty}}}{p-2} \|\boldsymbol{\alpha}\|_{\ell^2}^2 + \sqrt{\frac{2}{\pi}} \mathrm{e}^{\varrho\|\boldsymbol{\alpha}\|_{\ell^\infty}} \|\boldsymbol{\alpha}\|_{\ell^1} \right). \tag{2.115}$$

By Proposition 2.4 and Theorem 2.12 for (2.15), the tensorized Hermite polynomials $(H_\nu)_{\nu\in\mathfrak{F}}$ form an orthonormal basis of $L^2(\Gamma, \gamma)$. We transform these to an orthonormal basis of $L^2(\Gamma, \gamma_\varrho)$ using the map

$$\tau_\varrho : \mathbb{R}^\infty \to \mathbb{R}^\infty, \quad (y_m)_{m\in\mathbb{N}} \mapsto (\mathrm{e}^{-\varrho\alpha_m} y_m)_{m\in\mathbb{N}}. \tag{2.116}$$

Note that τ_ϱ maps Γ bijectively onto Γ.

Lemma 2.37. For all $\varrho \in \mathbb{R}$, the map

$$L^2(\Gamma, \gamma) \to L^2(\Gamma, \gamma_\varrho), \quad v \mapsto v \circ \tau_\varrho \tag{2.117}$$

is a unitary isomorphism of Hilbert spaces. Furthermore,

$$\int_\Gamma v(y)\gamma(\mathrm{d}y) = \int_\Gamma v(\tau_\varrho(y))\gamma_\varrho(\mathrm{d}y) \quad \forall v \in L^2(\Gamma, \gamma). \tag{2.118}$$

Proof. The standard Gaussian measure γ is the image of γ_ϱ under the map τ_ϱ, *i.e.*, $\gamma(E) = \gamma_\varrho(\tau_\varrho^{-1}(E))$ for all $E \in \mathcal{B}(\Gamma)$. This is easily checked for sets $E = \{y \in \Gamma\,;\, y_m \le x\}$ with $x \in \mathbb{R}$ and $m \in \mathbb{N}$. Then (2.118) is the transformation theorem, and the rest of the claim is a direct consequence. □

Proposition 2.38. For all $\varrho \in \mathbb{R}$, $(H_\nu \circ \tau_\varrho)_{\nu\in\mathfrak{F}}$ is an orthonormal basis of $L^2(\Gamma, \gamma_\varrho)$.

Proof. The claim follows from Lemma 2.37 since $(H_\nu)_{\nu\in\mathfrak{F}}$ from (2.15) is an orthonormal basis of $L^2(\Gamma, \gamma)$: see Proposition 2.4 and Theorem 2.12. □

Corollary 2.39. Let $\varrho \ge 0$ and $f \in L^p(\Gamma, \gamma_\varrho; V')$ with $p > 2$. Then the solution u of (2.105) is of the form

$$u(y) = \sum_{\nu\in\mathfrak{F}} u_\nu H_\nu(\tau_\varrho(y)), \quad y \in \Gamma, \tag{2.119}$$

with convergence in $L^2(\Gamma, \gamma_\varrho; V)$, for the coefficients

$$u_\nu = \int_\Gamma u(\tau_\varrho^{-1}(y)) H_\nu(y)\gamma(\mathrm{d}y) \in V, \quad \nu \in \mathfrak{F}. \tag{2.120}$$

Furthermore, $\boldsymbol{u} := (u_\nu)_{\nu\in\mathfrak{F}} \in \ell^2(\mathfrak{F}; V)$ and

$$\|\boldsymbol{u}\|_{\ell^2(\mathfrak{F};V)} \le \widetilde{c}_{\varrho,p} \|f\|_{L^p(\Gamma,\gamma_\varrho;V')} \tag{2.121}$$

with the constant $\widetilde{c}_{\varrho,p}$ from (2.115).

Proof. By Theorem 2.36 with $q = 2$, the solution u of (2.105) is an element of $L^2(\Gamma, \gamma_\varrho; V)$. Then (2.119) is the expansion of u in the orthonormal basis from Proposition 2.38, and (2.120) follows from (2.118) since

$$u_\nu = \int_\Gamma u(y) H_\nu(\tau_\varrho(y))\gamma_\varrho(\mathrm{d}y) = \int_\Gamma u(\tau_\varrho^{-1}(y)) H_\nu(y)\gamma(\mathrm{d}y).$$

Equation (2.121) is a consequence of (2.114) and Parseval's identity. □

Weak formulation on a problem-dependent space

Since the diffusion coefficient $a(y, x)$ is not uniformly bounded in $y \in \Gamma$, simply integrating (2.105) over Γ with respect to γ does not lead to a well-posed linear variational problem on $L^2(\Gamma, \gamma; V)$. As shown below, this difficulty can be overcome by assuming sufficient integrability of f with respect to γ_ϱ for a parameter $\varrho > 0$.

Furthermore, if $a_*(x)$ is not bounded away from zero, then neither is $a(y, x)$. For this reason, we integrate (2.105) with respect to a measure that is stronger than γ in the sense of Proposition 2.33, but not by as much as γ_ϱ.

For parameters $0 \le \vartheta < 1$ and $\varrho > 0$, define

$$\begin{aligned} B_{\vartheta\varrho}(w,v) &:= \int_\Gamma b(y; w(y), v(y)) \gamma_{\vartheta\varrho}(\mathrm{d}y) \\ &= \int_\Gamma \int_D a(y,x) \nabla w(y,x) \cdot \nabla v(y,x) \, \mathrm{d}x \gamma_{\vartheta\varrho}(\mathrm{d}y) \end{aligned} \tag{2.122}$$

and, assuming that $y \mapsto f(y;\cdot) \in V'$ is $\mathcal{B}(\Gamma)$-measurable and sufficiently integrable,

$$F_{\vartheta\varrho}(v) := \int_\Gamma f(y; v(y)) \gamma_{\vartheta\varrho}(\mathrm{d}y) = \int_\Gamma \int_D f(x) v(y,x) \, \mathrm{d}x \gamma_{\vartheta\varrho}(\mathrm{d}y) \tag{2.123}$$

for suitable w and v.

We define the space

$$\mathcal{V}_{\vartheta\varrho} := \{ v : \Gamma \to V \ \mathcal{B}(\Gamma)\text{-measurable} \, ; \ B_{\vartheta\varrho}(v,v) < \infty \}. \tag{2.124}$$

More precisely, $\mathcal{V}_{\vartheta\varrho}$ contains equivalence classes of γ-a.e. identical functions.

Proposition 2.40. The space $\mathcal{V}_{\vartheta\varrho}$ endowed with the inner product $B_{\vartheta\varrho}(\cdot,\cdot)$ is a Hilbert space.

We refer to Gittelson (2010*a*, Proposition 3.6) for the proof of Proposition 2.40. The argument is analogous to a standard proof that $L^2(\mathbb{R})$ is a Hilbert space.

Lemma 2.41. For all $w, v \in L^2(\Gamma, \gamma_\varrho; V)$,

$$|B_{\vartheta\varrho}(w,v)| \le \widehat{c}_{\vartheta\varrho} \|w\|_{L^2(\Gamma,\gamma_\varrho;V)} \|v\|_{L^2(\Gamma,\gamma_\varrho;V)}$$

with

$$\widehat{c}_{\vartheta\varrho} = \left(\|a_*\|_{L^\infty(D)} + \|a_0\|_{L^\infty(D)} \exp\left(\frac{\mathrm{e}^{2\varrho\|\boldsymbol{\alpha}\|_{\ell^\infty}}}{4(1-\vartheta)\varrho} \|\boldsymbol{\alpha}\|_{\ell^1} \right) \right) \exp\big((1-\vartheta)\varrho\|\boldsymbol{\alpha}\|_{\ell^1}\big).$$

Proof. By continuity of $b(y;\cdot,\cdot)$ for $y \in \Gamma$,

$$\begin{aligned} |B_{\vartheta\varrho}(w,v)| &\le \int_\Gamma \frac{\zeta_{\vartheta\varrho}(y)}{\zeta_\varrho(y)} \widehat{a}(y) \|w(y)\|_V \|v(y)\|_V \gamma\varrho(\mathrm{d}y) \\ &\le \left\| \frac{\zeta_{\vartheta\varrho}}{\zeta_\varrho} \widehat{a} \right\|_{L^\infty(\Gamma,\gamma)} \|w\|_{L^2(\Gamma,\gamma_\varrho;V)} \|v\|_{L^2(\Gamma,\gamma_\varrho;V)}, \end{aligned}$$

and the claim follows from Lemmas 2.29 and 2.32 with $\eta = \vartheta\varrho$, $\chi = \varrho$ and $k = 1$. □

Lemma 2.42. For all $v \in L^2(\Gamma, \gamma; V)$ with $B_{\vartheta\varrho}(v,v) < \infty$,

$$B_{\vartheta\varrho}(v,v) \ge \check{c}_{\vartheta\varrho} \|v\|^2_{L^2(\Gamma,\gamma;V)}$$

with

$$\check{c}_{\vartheta\varrho} = \left(\operatorname*{ess\,inf}_{x\in D} a_*(x) + \check{a}_0 \exp\left(-\frac{\mathrm{e}^{2\vartheta\varrho\|\boldsymbol{\alpha}\|_{\ell^\infty}}}{4\vartheta\varrho}\|\boldsymbol{\alpha}\|_{\ell^1}\right)\right)\exp\bigl(-\vartheta\varrho\|\boldsymbol{\alpha}\|_{\ell^1}\bigr).$$

Proof. Using coercivity of $b(y;\cdot,\cdot)$ for $y\in\Gamma$, we obtain

$$\begin{aligned} B_{\vartheta\varrho}(v,v) &\geq \int_\Gamma \zeta_{\vartheta\varrho}(y)\check{a}(y)\|v(y)\|_V^2\gamma(\mathrm{d}y)\\ &\geq \operatorname*{ess\,inf}_{y\in\Gamma}\zeta_{\vartheta\varrho}(y)\check{a}(y)^{-1}\|v\|^2_{L^2(\Gamma,\gamma;V)}, \end{aligned}$$

and the claim follows from Lemmas 2.29 and 2.32 with $\eta=0$, $\chi=\vartheta\varrho$ and $k=1$. □

Proposition 2.43. If $\vartheta>0$, the Hilbert space $\mathcal{V}_{\vartheta\varrho}$ is related to Lebesgue–Bochner spaces by the continuous embeddings

$$L^2(\Gamma,\gamma;V)\supset\mathcal{V}_{\vartheta\varrho}\supset L^2(\Gamma,\gamma_\varrho;V).$$

For $\vartheta=0$, this still holds if $\operatorname{ess\,inf}_{x\in D} a_*(x)>0$.

Proof. Lemmas 2.41 and 2.42 imply

$$\check{c}_{\vartheta\varrho}\|v\|^2_{L^2(\Gamma,\gamma;V)}\leq B_{\vartheta\varrho}(v,v)\leq\widehat{c}_{\vartheta\varrho}\|v\|^2_{L^2(\Gamma,\gamma_\varrho;V)}$$

for all $v\in L^2(\Gamma,\gamma_\varrho;V)$. □

Also, using (2.111) with $\eta=\vartheta\varrho$ and $\chi=\varrho$, it follows that if $f\in L^2(\Gamma,\gamma_\varrho;V')$, then $F_{\vartheta\varrho}$ is in the dual of $\mathcal{V}_{\vartheta\varrho}$.

Theorem 2.44. If $F_{\vartheta\varrho}\in\mathcal{V}'_{\vartheta\varrho}$, then the solution u of (2.105) is the unique solution in $\mathcal{V}_{\vartheta\varrho}$ of the linear variational problem

$$B_{\vartheta\varrho}(u,v)=F_{\vartheta\varrho}(v)\quad\forall v\in\mathcal{V}_{\vartheta\varrho}. \tag{2.125}$$

Proof. By the Riesz isomorphism on the Hilbert space $\mathcal{V}_{\vartheta\varrho}$, (2.125) has a unique solution $u\in\mathcal{V}_{\vartheta\varrho}$. Fix a $w\in V$. Setting $v(y)=w1_E(y)$ for all $E\in\mathcal{B}(\Gamma)$ on which $\widehat{a}(y)$ is bounded, it follows that the solution of (2.125) satisfies

$$\int_E b(y;u,w)-f(y;w)\gamma_{\vartheta\varrho}(\mathrm{d}y)=0,$$

and since Γ is a countable union of such sets E, the integrand must vanish $\gamma_{\vartheta\varrho}$-a.e. on Γ. The claim follows since $w\in V$ is arbitrary. □

Galerkin approximation

Using the variational formulation (2.125) of (2.105), we can define Galerkin projections of u onto suitable spaces. Let $\mathcal{V}_N\subset L^2(\Gamma,\gamma_\varrho;V)\subset\mathcal{V}_{\vartheta\varrho}$ be finite-dimensional. Then the Galerkin projection of u onto $\mathcal{V}_N$ is the unique

element $u_N \in \mathcal{V}_N$ satisfying

$$B_{\vartheta\varrho}(u_N, v_N) = F_{\vartheta\varrho}(v_N) \quad \forall v_N \in \mathcal{V}_N. \tag{2.126}$$

This u_N is well-defined since, being finite-dimensional, $\mathcal{V}_N$ is a closed subspace of $\mathcal{V}_{\vartheta\varrho}$, and thus also a Hilbert space when endowed with the inner product $B_{\vartheta\varrho}(\cdot,\cdot)$.

Theorem 2.45. If $f \in L^p(\Gamma, \gamma_\varrho; V')$ for a $p > 2$, the Galerkin projection u_N satisfies

$$\|u - u_N\|_{L^2(\Gamma,\gamma;V)} \leq \sqrt{\frac{\widehat{c}_{\vartheta\varrho}}{\check{c}_{\vartheta\varrho}}} \inf_{v_N \in \mathcal{V}_N} \|u - v_N\|_{L^2(\Gamma,\gamma_\varrho;V)}. \tag{2.127}$$

Proof. Theorem 2.36 implies that $u \in L^2(\Gamma, \gamma_\varrho; V)$. By definition, u_N is the orthogonal projection of u onto $\mathcal{V}_N$ with respect to the inner product $B_{\vartheta\varrho}(\cdot,\cdot)$. Therefore, it minimizes the projection error in the norm induced by $B_{\vartheta\varrho}(\cdot,\cdot)$. Using Lemmas 2.41 and 2.42, we have

$$\begin{aligned}
\check{c}_{\vartheta\varrho}\|u - u_N\|^2_{L^2(\Gamma,\gamma;V)} &\leq B_{\vartheta\varrho}(u - u_N, u - u_N) \\
&= \inf_{v_N \in \mathcal{V}_N} B_{\vartheta\varrho}(u - v_N, u - v_N) \\
&\leq \widehat{c}_{\vartheta\varrho} \inf_{v_N \in \mathcal{V}_N} \|u - v_N\|^2_{L^2(\Gamma,\gamma_\varrho;V)},
\end{aligned}$$

and the claim follows. □

Remark 2.46. The errors on the two sides of the estimate (2.127) are measured in different norms. Therefore, Theorem 2.45 states that the Galerkin projection is almost quasi-optimal. Inserting the values of $\widehat{c}_{\vartheta\varrho}$ and $\check{c}_{\vartheta\varrho}$ from Lemmas 2.41 and 2.42, we see that the constant in (2.127) is

$$\sqrt{\frac{\widehat{c}_{\vartheta\varrho}}{\check{c}_{\vartheta\varrho}}} = \sqrt{\frac{\|a_*\|_{L^\infty(D)} + \|a_0\|_{L^\infty(D)} \exp\left(\frac{\mathrm{e}^{2\varrho\|\boldsymbol{\alpha}\|_{\ell^\infty}}}{4(1-\vartheta)\varrho} \|\boldsymbol{\alpha}\|_{\ell^1}\right)}{\operatorname{ess\,inf}_{x\in D} a_*(x) + \check{a}_0 \exp\left(-\frac{\mathrm{e}^{2\vartheta\varrho\|\boldsymbol{\alpha}\|_{\ell^\infty}}}{4\vartheta\varrho} \|\boldsymbol{\alpha}\|_{\ell^1}\right)}} \exp\left(\frac{\varrho}{2} \|\boldsymbol{\alpha}\|_{\ell^1}\right).$$

In particular, it tends to ∞ as ϱ approaches 0 or ∞, or if ϑ approaches 1. If a_* is not bounded away from 0, then the constant also tends to ∞ as ϑ approaches 0.

Motivated by Corollary 2.39, we consider in particular spaces $\mathcal{V}_N$ of the form

$$\mathcal{V}_N := \{v \in L^2(\Gamma, \gamma_\varrho; V)\,;\; v_\nu \in V_{N,\nu}\ \forall \nu \in\in \mathrm{d}x\}, \tag{2.128}$$

where $V_{N,\nu} \subset V$ is a finite-dimensional subspace for all $\nu \in \mathfrak{F}$, and $V_{N,\nu} = \{0\}$ for all but finitely many $\nu \in \mathfrak{F}$. In (2.128), $(v_\nu)_{\nu\in\mathfrak{F}}$ are the Hermite coefficients of $v \in L^2(\Gamma, \gamma_\varrho; V)$ with respect to the scaled Hermite polynomials

$(H_\nu \circ \tau_\varrho)_{\nu\in\mathfrak{F}}$ from Proposition 2.38, *i.e.*,

$$v_\nu = \int_\Gamma v(\tau_\varrho^{-1}(y))H_\nu(y)\gamma(\mathrm{d}y), \quad \nu \in \mathfrak{F}. \tag{2.129}$$

Then $\mathcal{V}_N$ is a finite-dimensional subspace of $L^2(\Gamma,\gamma_\varrho;V)$, and its dimension is the sum of the dimensions of $V_{N,\nu}$ over $\nu \in \mathfrak{F}$.

Corollary 2.47. If $f \in L^p(\Gamma,\gamma_\varrho;V')$ for some $p > 2$, and $\mathcal{V}_N$ is of the form (2.128), then the Galerkin projection u_N satisfies

$$\|u - u_N\|_{L^2(\Gamma,\gamma;V)} \le \sqrt{\frac{\widehat{c}_{\vartheta\varrho}}{\check{c}_{\vartheta\varrho}}}\left(\sum_{\nu\in\mathfrak{F}} \inf_{v_N\in V_{N,\nu}} \|u_\nu - v_N\|_V^2\right)^{1/2}. \tag{2.130}$$

Proof. The claim follows from Theorem 2.45 and Parseval's identity since $(H_\nu \circ \tau_\varrho)_{\nu\in\mathfrak{F}}$ is an orthonormal basis of $L^2(\Gamma,\gamma_\varrho;V)$ by Proposition 2.38. □

3. Optimal convergence rates of stochastic Galerkin approximations

We have seen in the preceding sections, in Proposition 2.27 and in Corollary 2.47, that the Galerkin approximations of solutions for the parametric, deterministic formulations of the stochastic problems are well-defined and, in the mean square sense, quasi-optimal. The natural questions that arise from these results for *computable* numerical GPC approximations are: (a) What is the best possible rate achievable by polynomial chaos expansions that are truncated to at most N terms? and (b) How does one obtain in a *constructive* fashion index sets $\Lambda \subset \mathfrak{F}$ of 'active' polynomial chaos coefficients whose cardinality does not exceed N, *i.e.*, $\#\Lambda \le N$?

Question (a) is addressed in the present section, whereas (b) will be discussed in the subsequent section. We deal with question (a) in the case of Legendre chaos. The general approach to establishing the convergence rate of N-term truncated GPC approximations consists in a careful analysis of the regularity of the unknown solution, in terms of summability of the GPC coefficient series.

There are several ways to obtain *a priori* estimates on GPC coefficients. The most straightforward way is by a bootstrapping argument in the spirit of regularity theory for PDEs. This approach was taken in Todor and Schwab (2007) and Cohen, DeVore and Schwab (2010). We found, however, approaches based on the *analytic continuation* of the parametric, deterministic PDEs resulting from GPC expansions of the input random field in the parameter space of these expansions to yield, in general, sharper bounds. We therefore present *a priori* estimates of GPC coefficients obtained from the analytic continuation. This approach is rather general (we refer to

Schwab and Stuart (2011) for an application of this approach to sparse approximation of posterior densities in Bayesian inverse problems).

3.1. Elliptic problems

We consider the stochastic elliptic problem (2.49) with coefficient $a(x,\omega)$ depending in an affine fashion (2.55) on the coordinates $Y_m(\omega)$. Then, by Lemma 2.20, the functions $\varphi_m(x)$ that characterize the spatial heterogeneity of the random coefficient $a(\omega,x)$ in (2.49) can be rescaled via the scaling factors α_m in Lemma 2.20 such that the random coefficients $Y_m(\omega)$ satisfy $Y_m(\omega) \in [-1,1]$ for all $\omega \in \Omega$ and $m \in \mathbb{N}$. *In what follows we shall assume that α_m has been chosen in this way and denote the scaled coefficient functions by $\psi_j(x)$, i.e.,*

$$\psi_j(x) = \alpha_j \varphi_j(x), \quad j = 1,2,\ldots.$$

Then the abstract parametric deterministic operator equation (2.42), (2.44) reads, formally, as

$$-\nabla\cdot(a(y,x)\nabla u) = f \quad \text{in} \quad D, \quad u(y,\cdot)|_{\partial D} = 0, \quad y \in \Gamma. \tag{3.1}$$

Basic assumptions and preliminaries
We impose in addition the following conditions on the coefficient functions ψ_j.

(C1) For all $j \in \mathbb{N}$, $\psi_j \in L^\infty(D)$, and $\psi_j(x)$ is defined for all $x \in D$.

(C2) The $y = (y_1, y_2, \ldots)$ to be considered are all in the set $\Gamma = [-1,1]^{\mathbb{N}}$, *i.e.*, the unit ball of the sequence space $\ell^\infty(\mathbb{N})$ (with $\mathbb{N}$ replaced by $\{1,\ldots,K\}$ when the number K of random parameters is finite).

(C3) For each $a(x,y)$ to be considered, we have for every $x \in D$ and every $y \in \Gamma$

$$a(x,y) = \bar{a}(x) + \sum_{j\geq 1} y_j \psi_j(x). \tag{3.2}$$

Under these assumptions, we consider the map $y \mapsto u(y)$ from Γ to V, where $u(y)$ is the solution of (3.1) with coefficient given by (3.2).

The variational formulation of (3.1) is set in the Sobolev space $V := H_0^1(D)$, called the *energy space*, which is the set of all functions v whose trace vanishes on the boundary of D and whose *energy norm* $\|v\|_V := \|\nabla v\|_{L^2(D)}$ is finite. The dual of V is denoted by $V' = H^{-1}(D)$. The solution of the parametric problem (3.1) is defined for any $f \in V'$ as a measurable mapping $u : \Gamma \to V$ which satisfies the parametric elliptic PDE (3.1) for a.e. parameter vector $y \in \Gamma$, in variational form. If we let $A_m = -\nabla\cdot\psi_m(x)\nabla \in \mathcal{L}(V,V')$, then, by Proposition 2.22, the parametric problem (3.1) admits a unique solution.

As indicated before, estimates of the Legendre expansion coefficients of the parametric, deterministic solution will be obtained by tools from the theory of complex variables. It will be crucial to verify that the results from Section 2.3 will also hold for complex extensions of the parameter vector.

To this end, we now recapitulate results on well-posedness of variational elliptic problems. Consider the generic diffusion problem with coefficient $\alpha(x)$ in variational form

$$\int_D \alpha(x)\nabla u(x)\cdot\nabla v(x)\,\mathrm{d}x = \int_D f(x)v(x)\,\mathrm{d}x, \quad \text{for all } \ v\in V, \tag{3.3}$$

where $\alpha(x)$ satisfies the ellipticity condition

$$0 < r \le \alpha(x) \le R < \infty, \quad x\in D. \tag{3.4}$$

Then the Lax–Milgram lemma implies existence and uniqueness of the solution u of (3.3) in V and this solution satisfies the *a priori* estimate

$$\|u\|_V \le \frac{\|f\|_{V'}}{r}. \tag{3.5}$$

Now consider the case that the coefficient function α is complex-valued. In this case, the weak solution of (3.3) will be a complex-valued function. Therefore, *we assume from now on that all function spaces V, their duals, etc., are spaces of complex-valued functions and duality is understood with respect to the antilinear dual pairing.* We shall not distinguish this generalization notationally. In this case,

$$0 < r \le \mathrm{Re}\,(\alpha(x)) \le |\alpha(x)| \le R < \infty, \quad x\in D. \tag{3.6}$$

For the parametric coefficient $a(y,x)$, ellipticity is ensured by the *uniform ellipticity assumption*, as follows.

Uniform ellipticity assumption. There exist $0 < r \le R < \infty$ such that, for all $x\in D$ and for all $y\in\Gamma$,

$$0 < r \le a(x,y) \le R < \infty. \tag{3.7}$$

We shall refer to assumption (3.7) in the following as UEA(r,R). We note that UEA(r,R) implies $r \le \bar{a}(x) \le R$ for all $x\in D$, since we can choose $y_j = 0$ for all $j\in\mathbb{N}$. We note also that the uniform in Γ validity of the lower and upper inequality in (3.7) is equivalent to the conditions that

$$\sum_{j\ge1} |\psi_j(x)| \le \bar{a}(x) - r, \quad x\in D, \tag{3.8}$$

and

$$\sum_{j\ge1} |\psi_j(x)| \le R - \bar{a}(x), \quad x\in D. \tag{3.9}$$

To ensure well-posedness of the parametric deterministic PDE (3.1) to complex values of the parameters y, we impose a complex analogue of hypothesis UEA(r, R), as follows.

Uniform ellipticity assumption in $\mathbb{C}$. There exist $0 < r \leq R < \infty$ such that, for all $x \in D$ and all $z \in \mathcal{U}$,

$$0 < r \leq \operatorname{Re}(a(x,z)) \leq |a(x,z)| \leq R < \infty. \tag{3.10}$$

We refer to (3.10) as UEAC(r, R). We extend the definition of $u(y)$ to $u(z)$ for the complex variable $z = (z_j)_{j\geq 1}$ (by using the z_j instead of y_j in the definition of a by (3.2)) where $|z_j| \leq 1$ for all j. Therefore, the parameter vector z belongs to the polydisc

$$\mathcal{U} := \prod_{j\geq 1} \{z_j \in \mathbb{C}\,;\ |z_j| \leq 1\} \supset \Gamma. \tag{3.11}$$

Using (3.8) and (3.9), it is readily seen that when $\bar{a}$ and ψ_j are real-valued, then UEA(r, R) implies that for all $x \in D$ and $z \in \mathcal{U}$,

$$0 < r \leq \operatorname{Re}(a(x,z)) \leq |a(x,z)| \leq 2R, \tag{3.12}$$

and therefore the corresponding solution $u(z)$ is well-defined in V for all $z \in \mathcal{U}$ according to the complex-valued version of the Lax–Milgram lemma. We leave to the reader the verification of Lemma 2.21 under UEAC(r, R) in the complex parameter case.

For the derivation of the convergence rates of best N-term truncated GPC expansions, the following observation, due to Stechkin, will be used repeatedly. Let $(\gamma_n)_{n\geq 1}$ denote a decreasing sequence of non-negative integers. Then, for any $0 < p \leq q \leq \infty$ and any $N \in \mathbb{N}$,

$$\left(\sum_{n\geq N} \gamma_n^q\right)^{\frac{1}{q}} \leq N^{\frac{1}{q}-\frac{1}{p}} \left(\sum_{n\geq 1} \gamma_n^p\right)^{\frac{1}{p}}. \tag{3.13}$$

For $q < \infty$ this is easily proved by combining the two estimates

$$\sum_{n\geq N} \gamma_n^q \leq \gamma_N^{q-p} \sum_{n\geq N} \gamma_n^p \leq \gamma_N^{q-p} \sum_{n\geq 1} \gamma_n^p \quad \text{and} \quad N\gamma_N^p \leq \sum_{n\leq N} \gamma_n^p \leq \sum_{n\geq 1} \gamma_n^p.$$

Verification of the case $q = \infty$ is left to the reader.

We shall use standard multivariate notation. As in (2.27), the countable set of finitely supported sequences of non-negative integers is denoted by

$$\mathfrak{F} := \{\nu = (\nu_1, \nu_2, \ldots)\,;\ \nu_j \in \mathbb{N},\ \text{and } \nu_j \neq 0 \text{ for only a finite number of } j\}, \tag{3.14}$$

which implies that

$$|\nu| := \sum_{j\geq 1} |\nu_j| \tag{3.15}$$

is finite if and only if $\nu \in \mathfrak{F}$. For $\nu \in \mathfrak{F}$ supported in $\{1, \dots, J\}$, we define the partial derivative

$$\partial^\nu u = \frac{\partial^{|\nu|} u}{\partial^{\nu_1} y_1 \cdots \partial^{\nu_J} y_J},$$

and the multi-factorial

$$\nu! := \prod_{j \geq 1} \nu_j! \quad \text{where} \quad 0! := 1.$$

If $\alpha = (\alpha_j)_{j \geq 1}$ is a sequence of complex numbers, we define for all $\nu \in \mathfrak{F}$

$$\alpha^\nu := \prod_{j \geq 1} \alpha_j^{\nu_j},$$

With this notation, the proof of summability of GPC coefficients of the solution will use the following result on sequence summability proved in Cohen *et al.* (2010).

Theorem 3.1. For $0 < p < 1$, we have $\left(\frac{|\nu|!}{\nu!} b^\nu\right)_{\nu \in \mathfrak{F}} \in \ell^p(\mathfrak{F})$ if and only if (i) $\sum_{j \geq 1} b_j < 1$, and (ii) $(b_j) \in \ell^p(\mathbb{N})$.

Main result on analyticity

We can now state the main result from Cohen, DeVore and Schwab (2011) on analyticity of the parametric solution $u(z)$ of (3.1) for parameter vectors z belonging to the polydisc $\mathcal{U}$. The result also shows convergence rates for truncated Taylor expansions which include the N most significant terms, in a sense made precise in the statement of the following theorem. Such results could become significant in the context of sensitivity analysis of PDEs on high-dimensional parameter spaces.

Theorem 3.2. If $a(x, z)$ satisfies UEAC(r, R) for some $0 < r \leq R < \infty$, and if $(\|\psi_j\|_{L^\infty(D)})_{j \geq 1} \in \ell^p(\mathbb{N})$ for some $0 < p < 1$, then $u(z)$ is analytic as a mapping from $\mathcal{U}$ into V. Moreover, for all $z \in \mathcal{U}$, $u(z)$ can be expanded in the Taylor series

$$u(z) = \sum_{\nu \in \mathfrak{F}} t_\nu z^\nu \quad \text{in} \quad V, \tag{3.16}$$

where the Taylor coefficients $t_\nu \in V$ are defined as

$$t_\nu := \frac{1}{\nu!} \partial^\nu u(0), \quad \nu \in \mathfrak{F}$$

and where $t_\nu \in V$ and $(\|t_\nu\|_V)_{\nu \in \mathfrak{F}} \in \ell^p(\mathfrak{F})$ for the same value of p.

The convergence of the series (3.16) is unconditional in the following sense. If $(\Lambda_N)_{N \geq 1}$ is any increasing sequence of finite sets which exhausts $\mathfrak{F}$, the

partial sums $S_{\Lambda_N} u(z) = \sum_{\nu \in \Lambda_N} t_\nu z^\nu$ satisfy

$$\lim_{N \to +\infty} \sup_{z \in \mathcal{U}} \|u(z) - S_{\Lambda_N} u(z)\|_V = 0. \tag{3.17}$$

If in addition Λ_N is a set of $\nu \in \mathfrak{F}$ corresponding to indices of N Taylor coefficients with largest norms $\|t_\nu\|_V$, we have the convergence estimate

$$\sup_{z \in \mathcal{U}} \|u(z) - S_{\Lambda_N} u(z)\|_V \leq \|(\|t_\nu\|_V)\|_{\ell^p(\mathfrak{F})} N^{-s}, \quad s := \frac{1}{p} - 1. \tag{3.18}$$

In the statement of the preceding theorem, a sequence $(\Lambda_N)_{N \geq 1} \subset \mathfrak{F}$ is said to *exhaust* $\mathfrak{F}$ if any finite subset $\Lambda \subset \mathfrak{F}$ is contained in all sets Λ_N, $N \geq N_0$ for some N_0. We next give the proof of this theorem, drawing upon Cohen *et al.* (2011).

Proof of analyticity of $u(z)$

There are several approaches to proving analyticity of $u(z)$ as a V-valued function: a power series approach would require a bootstrapping argument for real parameter values as in Cohen *et al.* (2010). Here, we establish strong differentiability of u with respect to the complex coordinates z_j by a difference quotient argument.

The key to proving Theorem 3.2 is the observation that if UEAC(r, R) holds, then $z \mapsto u(z)$ is a V-valued and bounded analytic function in certain domains which are larger than $\mathcal{U}$: for $0 < \delta \leq 2R < \infty$ we define the set

$$\mathcal{A}_\delta = \{z \in \mathbb{C}^{\mathbb{N}} \,;\, \delta \leq \mathrm{Re}\,(a(x, z)) \leq |a(x, z)| \leq 2R \ \text{ for every } \ x \in D\}. \tag{3.19}$$

Clearly, if UEAC(r, R) holds, then for any $0 < \delta < r$ the domain $\mathcal{A}_\delta$ contains $\mathcal{U}$. By the Lax–Milgram lemma, for any $z \in \mathcal{A}_\delta$ there exists a unique solution $u(z) \in V$ of the parametric problem (3.1) which satisfies the *a priori* estimate

$$\|u(z)\|_V \leq \frac{\|f\|_{V'}}{\delta} \quad \text{for all} \ \ z \in \mathcal{A}_\delta. \tag{3.20}$$

The difference quotient argument for establishing holomorphy of $u(z)$ as a V-valued function on $\mathcal{A}_\delta$ is based on the following perturbation lemma, whose proof is straightforward. We start from a stability result, which is also used further in this section.

Lemma 3.3. If u and $\widetilde{u}$ are solutions of (3.3) with the same right-hand side f and with coefficients α and $\widetilde{\alpha}$, respectively, and if these coefficients both satisfy the assumption (3.6), then

$$\|u - \widetilde{u}\|_V \leq \frac{\|f\|_{V'}}{r^2} \|\alpha - \widetilde{\alpha}\|_{L^\infty(D)}. \tag{3.21}$$

Based on Lemma 3.3, we are now in a position to prove holomorphy of the mapping $z \mapsto u(z)$.

Lemma 3.4. At any $z \in \mathcal{A}_\delta$, the function $z \mapsto u(z)$ admits a complex derivative $\partial_{z_j} u(z) \in V$ with respect to each variable z_j. This derivative is the weak solution of the parametric problem: find $\partial_{z_j} u(z) \in V$ such that, for all $v \in V$ and for all $z \in \mathcal{A}_\delta$,

$$\int_D a(x,z) \nabla \partial_{z_j} u(x,z) \cdot \nabla v(x)\, \mathrm{d}x = L_0(v) := -\int_D \psi_j(x) \nabla u(x,z) \cdot \nabla v(x)\, \mathrm{d}x. \tag{3.22}$$

To prove Lemma 3.4, we fix $j \geq 1$ and $z \in \mathcal{A}_\delta$ and denote by e_j the Kronecker sequence with 1 at index j and 0 at other indices. For $h \in \mathbb{C}\backslash\{0\}$, consider the difference quotient

$$w_h(z) = \frac{u(z+he_j) - u(z)}{h} \in V. \tag{3.23}$$

We notice that $w_h(z)$ is well-defined if $|h|\|\psi_j\|_{L^\infty(D)} \leq \frac{\delta}{2}$, since then

$$\frac{\delta}{2} \leq \mathrm{Re}\,(a(x, z+he_j)) \leq |a(x, z+he_j)| \leq 2R + \frac{\delta}{2}, \quad x \in D.$$

A short calculation shows that the difference quotient w_h is the unique solution to the variational problem

$$\int_D a(x,z) \nabla w_h(x,z) \cdot \nabla v(x)\, \mathrm{d}x = L_h(v), \quad \text{for all } \ v \in V,$$

where $L_h : v \to L_h(v) := -\int_D \psi_j \nabla u(z+he_j) \cdot \nabla v$ is a continuous, linear functional on V. The linear functional $L_h(\cdot)$ varies continuously in V' with h as h tends to 0 since the stability estimate (3.21) implies

$$\|u(z+he_j) - u(z)\|_V = \|\nabla u(z+he_j) - \nabla u(z)\|_{L^2(D)} \leq |h|\|\psi_j\|_{L^\infty(D)} \frac{4\|f\|_{V'}}{\delta^2}.$$

Therefore L_h converges towards L_0 in V' as $h \to 0$, which implies that w_h converges in V towards a limit $w_0 \in V$ which is the solution to

$$\int_D a(z,x) \nabla w_0(z) \cdot \nabla v = L_0(v), \quad \text{for all } \ v \in V.$$

Hence $\partial_{z_j} u(z) = w_0$ exists in V and is the unique solution of the variational problem (3.22).

For our further development, it is crucial to note that the analyticity domains $\mathcal{A}_\delta$ contain polydiscs: we let $\varrho := (\varrho_j)_{j\geq 1}$ be a sequence of positive radii and define the polydiscs

$$\mathcal{U}_\varrho = \prod_{j\geq 1} \{z_j \in \mathbb{C}\,;\ |z_j| \leq \varrho_j\} = \{z_j \in \mathbb{C}\,;\ z = (z_j)_{j\geq 1},\ |z_j| \leq \varrho_j\}. \tag{3.24}$$

We say that a sequence $\varrho = (\varrho_j)_{j\geq 1}$ is δ-*admissible* if and only if, for every $x \in D$,

$$\sum_{j\geq 1} \varrho_j |\psi_j(x)| \leq \mathrm{Re}\,(\bar{a}(x)) - \delta. \tag{3.25}$$

If the sequence ϱ is δ-admissible, then the polydisc $\mathcal{U}_\varrho$ is contained in $\mathcal{A}_\delta$. We also notice that the validity of the lower inequality in (3.10) for all $z \in \mathcal{U}$ is equivalent to the condition that

$$\sum_{j\geq 1} |\psi_j(x)| \leq \mathrm{Re}\,(\bar{a}(x)) - r, \quad x \in D. \tag{3.26}$$

Hence the sequence $\varrho_j = 1$ is δ-admissible for all $0 < \delta \leq r$, and that for $\delta < r$ there exist δ-admissible sequences such that $\varrho_j > 1$ for all $j \geq 1$, *i.e.*, such that the polydisc $\mathcal{U}_\varrho$ is strictly larger than $\mathcal{U}$ in every variable. These increasing δ-admissible sequences ϱ will next be exploited to obtain bounds on Taylor and Legendre coefficients by shifting paths of integration in the complex domain into the polydiscs $\mathcal{U}_\varrho$.

Estimates of Taylor coefficients

Estimates for the Taylor coefficients t_ν for $\nu \in \mathfrak{F}$ are given by the following result.

Lemma 3.5. If UEAC(r, R) holds for some $0 < r \leq R < \infty$ and if $\varrho = (\varrho_j)_{j\geq 1}$ is a δ-admissible sequence for some $0 < \delta < r$, then for any $\nu \in \mathfrak{F}$ we have the estimate

$$\|t_\nu\|_V \leq \frac{\|f\|_{V'}}{\delta} \prod_{j\geq 1} \varrho_j^{-\nu_j} = \frac{\|f\|_{V'}}{\delta} \varrho^{-\nu}, \tag{3.27}$$

where we use the convention that $t^{-0} = 1$ for any $t \geq 0$.

To prove Lemma 3.5, let $\nu = (\nu_j)_{j\geq 1} \in \mathfrak{F}$ and $J = \max\{j \in \mathbb{N}\,;\ \nu_j \neq 0\}$. For J and for $z \in \mathcal{U}$, we define the set $E_J = \{1, \ldots, J\}$ and the parameter vector z_{E_J} obtained from z by setting to 0 all entries z_j for $j > J$. We then have

$$\partial^\nu u(0) = \frac{\partial^{|\nu|} u_J}{\partial z_1^{\nu_1} \cdots \partial z_J^{\nu_J}}(0, \ldots, 0).$$

From the assumption that ϱ is δ-admissible, we have that

$$\|u_J(z_1, \ldots, z_J)\|_V \leq \frac{\|f\|_{V'}}{\delta}, \tag{3.28}$$

for all $(z_1, \ldots, z_J)$ in the J-dimensional polydisc

$$\mathcal{U}_{\varrho,J} := \prod_{1\leq j\leq J} \{z_j \in \mathbb{C}\,;\ |z_j| \leq \varrho_j\}. \tag{3.29}$$

Introducing the sequence $\widetilde{\varrho}$ defined by

$$\widetilde{\varrho}_j = \varrho_j + \varepsilon \text{ if } j \le J, \quad \widetilde{\varrho}_j = \varrho_j \text{ if } j > J, \quad \varepsilon := \frac{\delta}{2\|\sum_{j\le J} |\psi_j|\|_{L^\infty(D)}},$$

it is easily checked that $\widetilde{\varrho}$ is $\frac{\delta}{2}$-admissible and therefore $\mathcal{U}_{\widetilde{\varrho}} \subset \mathcal{A}_{\delta/2}$. We infer from Lemma 3.4 that for each $z \in \mathcal{U}_{\widetilde{\varrho}}$, u is holomorphic in each variable z_j.

Therefore u_J is strongly holomorphic as a V-valued function function with respect to each of the variables $z_1, \dots, z_J$ on the polydisc $\prod_{1\le j\le J}\{|z_j| < \widetilde{\varrho}_j\}$. This polydisc is an open neighbourhood of $\mathcal{U}_{\varrho,J}$. In this disc, we apply a suitable version of Cauchy's integral formula (*e.g.*, Theorem 2.1.2 of Hervé (1989)) with respect to each z_j, and write

$$u_J(\widetilde{z}_1, \dots, \widetilde{z}_J) = (2\pi\mathrm{i})^{-J} \int_{|z_1|=\varrho_1} \cdots \int_{|z_J|=\varrho_J} \frac{u_J(z_1, \dots, z_J)}{(\widetilde{z}_1 - z_1)\cdots(\widetilde{z}_J - z_J)} \,\mathrm{d}z_1 \cdots \mathrm{d}z_J.$$

Differentiating this expression with respect to z_j, we find

$$\frac{\partial^{|\nu|}}{\partial z_1^{\nu_1}\cdots\partial z_J^{\nu_J}} u_J(0, \dots, 0) = \nu!(2\pi\mathrm{i})^{-J} \int_{|z_1|=\varrho_1} \cdots \int_{|z_J|=\varrho_J} \frac{u_J(z_1, \dots, z_J)}{z_1^{\nu_1}\cdots z_J^{\nu_J}} \,\mathrm{d}z_1 \cdots \mathrm{d}z_J,$$

and therefore, using (3.28), we obtain the estimate

$$\left\| \frac{\partial^{|\nu|} u_J}{\partial z_1^{\nu_1}\cdots\partial z_J^{\nu_J}}(0, \dots, 0) \right\|_V \le \nu! \frac{\|f\|_{V'}}{\delta} \prod_{j\le J} \varrho_j^{-\nu_j},$$

which is equivalent to (3.27). □

Proof of Theorem 3.2

With the analyticity of the mapping $z \mapsto u(z)$ on the domains $\mathcal{A}_\delta$ in hand, the proof of Theorem 3.2 under the uniform ellipticity assumption UEAC(r, R) involves two steps: (a) a particular choice of $r/2$-admissible sequences ϱ, and (b) establishing $\ell^p(\mathfrak{F})$-summability of the Taylor coefficient sequence. With $\delta = r/2$, (3.27) of Lemma 3.5 reads

$$\|t_\nu\|_V \le \frac{2\|f\|_{V'}}{r} \prod_{j\ge 1} \varrho_j^{-\nu_j} = \frac{2\|f\|_{V'}}{r} \varrho^{-\nu}. \tag{3.30}$$

There are many sequences ϱ that are δ-admissible. We now indicate one such choice from Cohen *et al.* (2011), which is ν-dependent, in order to

yield possibly sharp coefficient bounds. We begin our choice by selecting $J_0 \in \mathbb{N}$ so large that

$$\sum_{j>J_0} \|\psi_j\|_{L^\infty(D)} \leq \frac{r}{12}, \tag{3.31}$$

Such a J_0 exists under the assumptions of Theorem 3.2 because

$$(\|\psi_j\|_{L^\infty(D)})_{j\geq 1} \in \ell^p(\mathbb{N}) \subset \ell^1(\mathbb{N}).$$

Without loss of generality, the basis elements ψ_j of the sequence are assumed to be enumerated in such a way that $(\|\psi_j\|_{L^\infty})_{j\geq 1}$ is non-increasing.

To construct a $\delta = r/2$ admissible vector ϱ of weights, we partition $\mathbb{N}$ into two sets $E := \{1 \leq j \leq J_0\}$ and $F := \mathbb{N}\backslash E$. Next we choose $\kappa > 1$ such that

$$(\kappa - 1) \sum_{j\leq J_0} \|\psi_j\|_{L^\infty(D)} \leq \frac{r}{4}. \tag{3.32}$$

For each multi-index $\nu \in \mathfrak{F}$ we select $\varrho = \varrho(\nu)$ by

$$\varrho_j := \kappa, \quad j \in E; \quad \varrho_j := \max\left\{1, \frac{r\nu_j}{4|\nu_F|\|\psi_j\|_{L^\infty(D)}}\right\}, \quad j \in F. \tag{3.33}$$

Here, ν_E denotes the restriction of ν to a set E and $|\nu_F| := \sum_{j>J_0} \nu_j$. We also make the convention that $\frac{\nu_j}{|\nu_F|} = 0$ when $|\nu_F| = 0$. It can be verified that the sequence ϱ defined in (3.33) is $\frac{r}{2}$-admissible (see Cohen *et al.* (2011) for details).

The general bound (3.30) is in particular valid for this sequence ϱ, for which it takes the form (with the convention that a factor equals 1 if $\nu_j = 0$)

$$\|t_\nu\|_V \leq \frac{2\|f\|_{V'}}{r} \left(\prod_{j\in E} \eta^{\nu_j}\right) \left(\prod_{j\in F} \left(\frac{|\nu_F| d_j}{\nu_j}\right)^{\nu_j}\right), \tag{3.34}$$

where $\eta := \frac{1}{\kappa} < 1$ and

$$d_j := \frac{4\|\psi_j\|_{L^\infty}}{r}.$$

We note that from (3.31)

$$\|d\|_{\ell^1} = \sum_{j>J_0} d_j \leq \frac{1}{3}. \tag{3.35}$$

To prove $\ell^p(\mathfrak{F})$ summability of the Taylor coefficients t_ν, we observe that the estimate (3.34) has the general form

$$\|t_\nu\|_V \leq C_r \alpha(\nu_E)\, \beta(\nu_F). \tag{3.36}$$

We let $\mathfrak{F}_E$ (respectively $\mathfrak{F}_F$) be the collection of $\nu \in \mathfrak{F}$ supported on E (respectively on F). Then, for any $0 < p < \infty$, we have

$$\sum_{\nu\in\mathfrak{F}} \|t_\nu\|_V^p \le C_r^p \sum_{\nu\in\mathfrak{F}} \alpha(\nu_E)^p \beta(\nu_F)^p = C_r^p\, A_E\, A_F, \tag{3.37}$$

where

$$A_E := \left(\sum_{\nu\in\mathfrak{F}_E} \alpha(\nu)^p\right), \quad A_F := \left(\sum_{\nu\in\mathfrak{F}_F} \beta(\nu)^p\right).$$

The first factor A_E is estimated as follows:

$$A_E = \sum_{\nu\in\mathfrak{F}_E} \alpha(\nu)^p = \sum_{\nu\in\mathfrak{F}_E} \prod_{j\in E} \eta^{p\nu_j} = \prod_{j\in E}\left(\sum_{n\ge 0} \eta^{np}\right) = \left(\frac{1}{1-\eta^p}\right)^{J_0} < \infty. \tag{3.38}$$

To show that A_F is finite, we observe

$$\beta(\nu) := \prod_{j\in F}\left(\frac{|\nu_F| d_j}{\nu_j}\right)^{\nu_j} \le \frac{|\nu_F|^{|\nu_F|}}{\prod_{j\in F} \nu_j^{\nu_j}} d^{\nu_F}, \quad \nu \in \mathfrak{F}_F, \tag{3.39}$$

where $d^{\nu_F} = \prod_{j\in F} d_j^{\nu_j}$ and $0^0 := 1$. By Stirling estimates,

$$\frac{n!\mathrm{e}^n}{\mathrm{e}\sqrt{n}} \le n^n \le \frac{n!\mathrm{e}^n}{\sqrt{2\pi}\sqrt{n}}, \tag{3.40}$$

which hold for all $n \ge 1$, we obtain

$$|\nu_F|^{|\nu_F|} \le\ |\nu_F|!\mathrm{e}^{|\nu_F|}.$$

On the other hand, using the left inequality in (3.40), we obtain

$$\prod_{j\in F} \nu_j^{\nu_j} \ge \frac{\nu_F!\mathrm{e}^{|\nu_F|}}{\prod_{j\in F} \max\{1, \mathrm{e}\sqrt{\nu_j}\}}.$$

With these estimates we obtain from (3.39) that

$$\beta(\nu) \le \frac{|\nu_F|!}{\nu_F!} d^{\nu_F} \prod_{j\in F} \max\{1, \mathrm{e}\sqrt{\nu_j}\} \le \frac{|\nu_F|!}{\nu_F!} \bar{d}^{\nu_F}, \tag{3.41}$$

where $\bar{d}_j := \mathrm{e}d_j, \quad j \in F$. We conclude by noticing that $\|\bar{d}\|_{\ell^1} = \mathrm{e}\|d\|_{\ell^1} \le \frac{\mathrm{e}}{3} < 1$. Since $\bar{d}$ is $\ell^p(\mathbb{N})$ summable, we may apply Theorem 3.1 to conclude the $\ell^p(\mathfrak{F})$ summability of t_ν.

With the $\ell^p(\mathfrak{F})$ summability of t_ν, the best N-term convergence rate estimate (3.18) follows from (3.13). □.

Convergence rates of Legendre expansions

The analyticity result Theorem 3.2 contains, as a special case, the rate of convergence of best N-term truncations of Taylor expansions in the

stochastic coordinates y_j. This result and the proof, however, also allow us to obtain corresponding results for Legendre expansions, as we shall show next. We shall obtain bounds in L^2 and pointwise bounds in the parameter vector y.

To this end, it will be convenient to introduce two types of Legendre expansions with different normalization of the Legendre basis: the Legendre basis $(P_n)_{n\geq 0}$ with L^∞ normalization

$$\|P_n\|_{L^\infty([-1,1])} = P_n(1) = 1 \tag{3.42}$$

and the L^2 normalized sequence $L_n(t) = \sqrt{2n+1}P_n(t)$, which satisfies

$$\int_{-1}^{1} |L_n(t)|^2 \frac{\mathrm{d}t}{2} = 1.$$

We recall that $L_0 = P_0 = 1$ and, for $\nu \in \mathfrak{F}$,

$$P_\nu(y) := \prod_{j\geq 1} P_{\nu_j}(y_j) \quad \text{and} \quad L_\nu(y) := \prod_{j\geq 1} L_{\nu_j}(y_j). \tag{3.43}$$

Note that $(L_\nu)_{\nu\in\mathfrak{F}}$ is an orthonormal basis of $L^2(\Gamma,\mu)$ where $\mathrm{d}\mu$ denotes the tensor product of the (probability) measures $\frac{\mathrm{d}y_j}{2}$ on $[-1,1]$ and is therefore a probability measure on $\Gamma = [-1,1]^{\mathbb{N}}$.

Since $u \in L^\infty(\Gamma,\mu;V) \subset L^2(\Gamma,\mu;V)$, it admits unique expansions

$$u(y) = \sum_{\nu\in\mathfrak{F}} u_\nu P_\nu(y) = \sum_{\nu\in\mathfrak{F}} v_\nu L_\nu(y), \tag{3.44}$$

that converge in $L^2(\Gamma,\mu;V)$, where the coefficients $u_\nu, v_\nu \in V$ are defined by

$$v_\nu := \int_\Gamma u(y)L_\nu(y)\mu(\mathrm{d}y) \quad \text{and} \quad u_\nu := \left(\prod_{j\geq 1}(1+2\nu_j)\right)^{1/2} v_\nu. \tag{3.45}$$

Once again, the key step in establishing sharp rates of convergence of best N-term Legendre GPC approximations of the solution of the parametric, deterministic problem are sharp *a priori* bounds on the Legendre coefficients. In order to prove their $\ell^p(\mathfrak{F})$ summability, we estimate the quantities $\|u_\nu\|_V$ and $\|v_\nu\|_V$. By (3.45),

$$\|u_\nu\|_V = \left(\prod_{j\geq 1}(1+2\nu_j)\right)^{\frac{1}{2}} \|v_\nu\|_V, \quad \nu \in \mathfrak{F}. \tag{3.46}$$

Therefore $\|v_\nu\|_V \leq \|u_\nu\|_V$ and it will be sufficient to prove the ℓ^p summability of $(\|u_\nu\|_V)_{\nu\in\mathfrak{F}}$. We have the following analogue to Lemma 3.5 from Cohen *et al.* (2011).

Lemma 3.6. Assume that UEAC(r, R) holds for some $0 < r \le R < \infty$. Let $\varrho = (\varrho_j)_{j\ge1}$ be a δ-admissible sequence for some $0 < \delta < r$ that satisfies $\varrho_j > 1$ for all j such that $\nu_j \neq 0$. Then, for any $\nu \in \mathfrak{F}$ we have the estimate

$$\|u_\nu\|_V \le \frac{\|f\|_{V'}}{\delta} \prod_{j\ge1,\nu_j\neq0} \varphi(\varrho_j)(2\nu_j+1)\varrho_j^{-\nu_j}, \tag{3.47}$$

where $\varphi(t) := \frac{\pi t}{2(t-1)}$ for $t > 1$.

Based on Lemma 3.6 and a judicious choice of δ-admissible sequence, the following theorem was shown in Cohen *et al.* (2011). It is the analogue to Theorem 3.2 for Legendre GPC expansions.

Theorem 3.7. If $a(x, z)$ satisfies UEAC(r, R) for some $0 < r \le R < \infty$ and if $(\|\psi_j\|_{L^\infty})_{j\ge1} \in \ell^p(\mathbb{N})$ for some $p < 1$, then the sequences $(\|u_\nu\|_V)_{\nu\in\mathfrak{F}}$ and $(\|v_\nu\|_V)_{\nu\in\mathfrak{F}}$ belong to $\ell^p(\mathfrak{F})$ for the same value of p. The Legendre expansions (3.44) converge in $L^\infty(\Gamma, \mu; V)$ in the following sense. If $(\Lambda_N)_{N\ge1}$ is any sequence of finite sets which exhausts $\mathfrak{F}$, then the partial sums $S_{\Lambda_N}u(y) := \sum_{\nu\in\Lambda_N} u_\nu(x)P_\nu(y) = \sum_{\nu\in\Lambda_N} v_\nu(x)L_\nu(y)$ satisfy

$$\lim_{N\to+\infty} \sup_{y\in\Gamma} \|u(y) - S_{\Lambda_N}u(y)\|_V = 0. \tag{3.48}$$

If Λ_N is a set of $\nu \in \mathfrak{F}$ corresponding to indices of N maximal $\|u_\nu\|_V$,

$$\sup_{y\in\Gamma} \|u(y) - S_{\Lambda_N}u(y)\|_V \le \|(\|u_\nu\|_V)\|_{\ell^p(\mathfrak{F})} N^{-s}, \quad s := \frac{1}{p} - 1. \tag{3.49}$$

If Λ_N is a set of $\nu \in \mathfrak{F}$ corresponding to indices of N maximal $\|v_\nu\|_V$,

$$\|u - S_{\Lambda_N}u\|_{L^2(\Gamma,\mu;V)} \le \|(\|v_\nu\|_V)\|_{\ell^p(\mathfrak{F})} N^{-s}, \quad s := \frac{1}{p} - \frac{1}{2}. \tag{3.50}$$

Spatial regularity and finite element discretization

So far, we have considered approximations of $u(y)$ with respect to the parameter vector $y \in \Gamma$ under the assumption that the coefficients t_ν, v_ν and u_ν could be obtained exactly. In practice, however, such coefficients must be approximated by a discretization scheme such as the finite element method. An additional *discretization error* arises in doing so which can be analysed using standard convergence results for finite element approximations. It is interesting to note that the regularity required of the solution $u(y)$ must then involve both smoothness in the stochastic parameter vector y *and* in the spatial domain D. We will now present some convergence results of this type. We assume that D is a bounded Lipschitz polyhedron D and that in D we are given a one-parameter, affine family of continuous, piecewise linear finite element spaces $(V_h)_{h>0}$ on a shape-regular family of simplicial triangulations of mesh width $h > 0$ in the sense of Ciarlet (1978).

To obtain regularity of the parametric solution $u(y)$ in D, additional regularity on f is required: we shall assume $f \in L^2(D) \subset V'$. Then

$$\|f\|_{V'} \leq C_P \|f\|_{L^2(D)}, \tag{3.51}$$

where C_P is the Poincaré constant of D (*i.e.*, $C_P = 1/\sqrt{\lambda_1}$ with λ_1 being the smallest eigenvalue of the Dirichlet Laplacian in D). Then the smoothness space $W \subset V$ is the space of all solutions to the Dirichlet problem

$$-\Delta u = f \quad \text{in} \quad D, \quad u|_{\partial D} = 0, \tag{3.52}$$

with $f \in L^2(D)$

$$W = \{v \in V\,;\ \Delta v \in L^2(D)\}. \tag{3.53}$$

We define the W-seminorm and the W-norm by

$$|v|_W = \|\Delta v\|_{L^2(D)}, \quad \|v\|_W := \|v\|_V + |v|_W. \tag{3.54}$$

It is well known that $W = H^2(D) \cap V$ for convex $D \subset \mathbb{R}^d$. Then any $w \in W$ may be approximated in V with convergence rate $\mathcal{O}(h)$ by continuous, piecewise linear finite element approximations on regular quasi-uniform simplicial partitions of D of mesh width h (see, *e.g.*, Ciarlet (1978), Braess (2007), Brenner and Scott (2002)). Therefore, denoting by $M = \dim(V_h) \sim h^{-d}$ the dimension of the finite element space, we have for all $w \in W$ the convergence rate, as $M = \dim(V_h) \to \infty$, of

$$\inf_{v_h \in V_h} \|w - v_h\|_V \leq C_t M^{-t} |w|_W, \tag{3.55}$$

with some $0 < t \leq 1/d$ (with $t = 1/d$ if $W \subset H^2(D)$).

Spatial regularity of the parametric solution $u(y)$ now takes the form of p-summability of W-norms of the t_ν, u_ν and v_ν.

In addition to the requirement $f \in L^2(D)$, we add a fourth assumption to conditions (C1)–(C3) on the coefficient $a(z, x)$.

(C4) The gradients of the functions $\bar{a}$ and ψ_j, for $j \geq 1$, are defined for every $x \in D$ and belong to $L^\infty(D)$.

Then the following regularity holds (see Cohen *et al.* (2011)).

Theorem 3.8. Let $f \in L^2(D)$ and let $a(z, x)$ satisfy UEAC(r, R) for some $0 < r \leq R < \infty$. If $(\|\psi_j\|_{L^\infty(D)})_{j \geq 1} \in \ell^p(\mathbb{N})$ and $(\|\nabla \psi_j\|_{L^\infty(D)})_{j \geq 1} \in \ell^p(\mathbb{N})$ for some $0 < p < 1$, then $(\|t_\nu\|_W)_{\nu \in \mathfrak{F}}$, $(\|u_\nu\|_W)_{\nu \in \mathfrak{F}}$ and $(\|v_\nu\|_W)_{\nu \in \mathfrak{F}}$ belong to $\ell^p(\mathfrak{F})$.

We now obtain bounds on the convergence rates of the *fully discrete* approximation of u by linear combinations

$$\sum_{\nu \in \Lambda} \widetilde{t}_\nu y^\nu, \quad \sum_{\nu \in \Lambda} \widetilde{u}_\nu P_\nu(y), \quad \text{or} \quad \sum_{\nu \in \Lambda} \widetilde{v}_\nu L_\nu(y),$$

where $\Lambda \subset \mathfrak{F}$ is finite and when the coefficients $\widetilde{t}_\nu$, $\widetilde{u}_\nu$ and $\widetilde{v}_\nu$ are finite element approximations of t_ν, u_ν and v_ν, respectively, from finite element spaces $(V_\nu)_{\nu\in\Lambda}$.

For efficiency of approximation as well as in the specific sparse tensor approximation schemes, it will be crucial that for given $\nu \in \Lambda \subset \mathfrak{F}$, the approximation space V_ν may depend on ν. To this end, we introduce the vector $\mathcal{M} = (M_\nu)_{\nu\in\Lambda}$ of the dimensions $M_\nu = \dim V_\nu$, $\nu \in \Lambda$, of the finite element approximation spaces V_ν used for approximating the t_ν. Without loss of generality, we may assume that the error bound (3.55) holds for all such M up to increasing C_t, and express the approximation rate in terms of N_{dof}, *i.e.*, the *total number of degrees of freedom involved*:

$$N_{\mathrm{dof}} := \sum_{\nu\in\Lambda} M_\nu. \tag{3.56}$$

Then we may estimate

$$\sup_{y\in\Gamma}\left\| u(y) - \sum_{\nu\in\Lambda} \widetilde{t}_\nu y^\nu \right\|_V \le \sum_{\nu\in\Lambda} \|t_\nu - \widetilde{t}_\nu\|_V + \sum_{\nu\notin\Lambda} \|t_\nu\|_V. \tag{3.57}$$

The first term on the right-hand side of (3.57) corresponds to the error occurring from the finite element discretization of the t_ν; the second term on the right-hand side corresponds to the error incurred by truncating the Taylor series. By taking $\Lambda := \Lambda_N$, the set of indices corresponding to N maximal $\|t_\nu\|_W$, it is bounded by

$$\sum_{\nu\notin\Lambda} \|t_\nu\|_W \le C_V N^{-s}, \quad s := \frac{1}{p} - 1. \tag{3.58}$$

The global error can then be bounded by

$$\sup_{y\in\Gamma}\left\| u(y) - \sum_{\nu\in\Lambda_N} \widetilde{t}_\nu y^\nu \right\|_W \le C_t \sum_{\nu\in\Lambda_N} M_\nu^{-t} |t_\nu|_W + C_V N^{-s}. \tag{3.59}$$

We now have an optimization problem: minimize the degrees of freedom M_ν such that N_{dof} is minimized for a fixed contribution $C_t \sum_{\nu\in\Lambda_N} M_\nu^{-t} |t_\nu|_W$ to the error, *i.e.*, we consider the minimization

$$\min\left\{ \sum_{\nu\in\Lambda_N} M_\nu \,;\, \sum_{\nu\in\Lambda_N} M_\nu^{-t} |t_\nu|_W \le N^{-s} \right\}. \tag{3.60}$$

In Cohen *et al.* (2011), this minimization problem is solved and the following result is obtained. To state it, we let $\widetilde{t}_\nu$, $\widetilde{u}_\nu$ and $\widetilde{v}_\nu$ denote the V-projection of t_ν, u_ν and v_ν, respectively, onto V_ν.

Theorem 3.9. Assume that the finite element spaces have the approximation property (3.55). Then, under the same assumptions as in Theorem 3.8, the following hold.

(a) Let Λ_N be a set of indices corresponding to N maximal $\|t_\nu\|_W$. Then there exists a choice of finite element spaces V_ν of dimension M_ν, $\nu \in \Lambda_N$, such that

$$\sup_{y\in\Gamma}\left\|u(y) - \sum_{\nu\in\Lambda_N} \widetilde{t}_\nu y^\nu\right\|_V \le C N_{\mathrm{dof}}^{-\min\{s,t\}}, \quad s := \frac{1}{p} - 1,$$

where $N_{\mathrm{dof}} = \sum_{\nu\in\Lambda_N} M_\nu$, $C = (\bar{C}_t + \|(\|t_\nu\|_V)\|_{\ell^p(\mathfrak{F})})\|(\|t_\nu\|_W)\|_{\ell^p(\mathfrak{F})}$.

(b) Let Λ_N be a set of indices corresponding to N maximal $\|u_\nu\|_W$. Then there exists a choice of finite element spaces V_ν of dimension M_ν, $\nu \in \Lambda_N$, such that

$$\sup_{y\in\Gamma}\left\|u(y) - \sum_{\nu\in\Lambda_N} \widetilde{u}_\nu P_\nu(y)\right\|_V \le C N_{\mathrm{dof}}^{-\min\{s,t\}}, \quad s := \frac{1}{p} - 1,$$

where $N_{\mathrm{dof}} = \sum_{\nu\in\Lambda_N} M_\nu$, $C = (\bar{C}_t + \|(\|u_\nu\|_V)\|_{\ell^p(\mathfrak{F})})\|(\|u_\nu\|_W)\|_{\ell^p(\mathfrak{F})}$.

(c) Let Λ_N be a set of indices corresponding to N maximal $\|v_\nu\|_W$. Then there exists a choice of finite element spaces V_ν of dimension M_ν, $\nu \in \Lambda_N$, such that

$$\left\|u - \sum_{\nu\in\Lambda_N} \widetilde{v}_\nu L_\nu\right\|_{L^2(\Gamma,\mu;V)} \le C N_{\mathrm{dof}}^{-\min\{s,t\}}, \quad s := \frac{1}{p} - \frac{1}{2},$$

where $N_{\mathrm{dof}} = \sum_{\nu\in\Lambda_N} M_\nu$, $C = (\bar{C}_t^2 + \|(\|v_\nu\|_V)\|^2_{\ell^p(\mathfrak{F})})^{\frac{1}{2}}\|(\|v_\nu\|_W)\|_{\ell^p(\mathfrak{F})}$.

3.2. Parabolic problems

A class of random parabolic problems

For $0 < T < \infty$, we consider in the bounded time interval $I = (0, T)$ linear, parabolic initial boundary value problems with random coefficients where, for ease of exposition, we assume that these coefficients are independent of t. We still denote by $D \subset \mathbb{R}^d$ a bounded Lipschitz domain and we denote the associated space–time cylinder by $Q_T = I \times D$. In Q_T, we consider the random parabolic initial boundary value problem

$$\frac{\partial u}{\partial t} - \nabla\cdot(a(x,\omega)\nabla u) = g(t,x), \quad u|_{\partial D\times I} = 0, \quad u|_{t=0} = h(x). \tag{3.61}$$

As before, we make the following assumption.

Assumption 3.10. There exist constants $0 < a_- \le a_+ < \infty$ such that

$$\forall x \in D,\ \forall \omega \in \Omega: \quad 0 < a_- \le a(x,\omega) \le a_+.$$

It will be convenient to impose a stronger requirement.

Assumption 3.11. The functions $\bar{a}$ and ψ_j satisfy

$$\sum_{j\geq 1} \|\psi_j\|_{L^\infty(D)} \leq \frac{\kappa}{1+\kappa}\bar{a}_-,$$

with $\bar{a}_- = \min_{x\in D} \bar{a}(x) > 0$ and $\kappa > 0$.

Assumption 3.10 is then satisfied by choosing

$$a_- := \bar{a}_- - \frac{\kappa}{1+\kappa}\bar{a}_- = \frac{1}{1+\kappa}\bar{a}_-. \tag{3.62}$$

We consider a *space–time variational formulation* of problem (3.61). To state it, we denote by $V = H_0^1(D)$ and $H = L^2(D)$ and identify H with its dual: $H \simeq H'$. Then $V \subset H \simeq H \subset V' = H^{-1}(D)$ is a Gelfand evolution triple. For the variational formulation of (3.61), we introduce the Bochner spaces

$$\mathcal{X} = L^2(I;V) \cap H^1(I;V') \quad \text{and} \quad \mathcal{Y} = L^2(I;V) \times H. \tag{3.63}$$

We equip $\mathcal{X}$ and $\mathcal{Y}$ with norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Y}}$, respectively, which are for $u \in \mathcal{X}$ and $v = (v_1, v_2) \in \mathcal{Y}$ given by

$$\|u\|_{\mathcal{X}} = (\|u\|^2_{L^2(I;V)} + \|u\|^2_{H^1(I;V')})^{1/2} \quad \text{and} \quad \|v\|_{\mathcal{Y}} = (\|v_1\|^2_{L^2(I;V)} + \|v_2\|^2_H)^{1/2}.$$

Given a coefficient realization $a(\omega,\cdot)$ with $\omega \in \Omega$, a *weak solution* of problem (3.61) is a function $u(\cdot,\cdot,\omega) \in \mathcal{X}$ such that

$$\int_I \left\langle \frac{\mathrm{d}u}{\mathrm{d}t}, v_1 \right\rangle_H \mathrm{d}t + \int_I \int_D a(x,\omega)\nabla u(t,x,\omega)\cdot\nabla v_1(t,x)\,\mathrm{d}x\,\mathrm{d}t + \langle u(0,\cdot,\omega), v_2\rangle_H = \int_I \langle g(t,\cdot), v_1\rangle\,\mathrm{d}t + \langle h, v_2\rangle_H, \quad \forall v \in \mathcal{Y}. \tag{3.64}$$

The following proposition from Schwab and Stevenson (2009) guarantees its well-posedness for all $\omega \in \Omega$, under Assumption 3.10.

Proposition 3.12. Assume that $g \in L^2(I;V')$, $h \in L^2(D)$ and that Assumption 3.10 holds. Then, for every $\omega \in \Omega$, the parabolic operator $B \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ induced by (3.61) in the weak form (3.64) is an isomorphism: for given $(g,h) \in \mathcal{Y}'$ and every $\omega \in \Omega$, problem (3.64) has a unique solution $u(\cdot,\cdot,\omega)$, which satisfies the *a priori* estimate

$$\|u\|_{\mathcal{X}} \leq C\big(\|g\|_{L^2(I;V')} + \|h\|_{L^2(D)}\big), \tag{3.65}$$

where the constant C is bounded uniformly for all realizations.

As before, we assume that the coefficient a in (3.61) is characterized by a sequence of random variables $(y_j)_{j\geq 1}$, *i.e.*, that

$$a(x,\omega) = \bar{a}(x) + \sum_{j\geq 1} y_j(\omega)\psi_j(x). \tag{3.66}$$

We again assume that the ψ_j are scaled in $L^\infty(D)$ such that $y_j : \Omega \to \mathbb{R}$, $j = 1, 2, \ldots$ are distributed identically and uniformly, and that the ψ_j are scaled in $L^\infty(D)$ such that the range of the Y_j is $[-1, 1]$ (see Lemma 2.20).

Parametric deterministic parabolic problems

As before, with (3.61) we associate the following parametric family of deterministic parabolic problems: given a source term $g(t,x)$ and initial data $h(x)$, for $y \in \Gamma$, find $u(t,x,y)$ such that

$$\begin{gathered}\frac{\partial u}{\partial t}(t,x,y) - \nabla_x \cdot [a(x,y)\nabla_x u(t,x,y)] = g(t,x), \quad \text{in } Q_T, \\ u(t,x,y)|_{\partial D\times I} = 0, \quad u|_{t=0} = h(x),\end{gathered} \tag{3.67}$$

where, for every $y = (y_1, y_2, \ldots) \in \Gamma$,

$$a(x,y) = \bar{a}(x) + \sum_{j=1}^{\infty} y_j\psi_j(x)$$

in $L^\infty(D)$. For the weak formulation of (3.67), we follow (3.64) and define for $y \in \Gamma$ the parametric family of bilinear forms $\Gamma \ni y \to b(y; w, (v_1, v_2)) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ by

$$\begin{aligned} b(y; w, (v_1, v_2)) = &\int_I \left\langle \frac{\mathrm{d}w}{\mathrm{d}t}, v_1(t,\cdot)\right\rangle_H \mathrm{d}t \\ &+ \int_D \int_I a(x,y)\nabla w(t,x)\cdot\nabla v_1(t,x)\,\mathrm{d}x\,\mathrm{d}t + \langle w(0), v_2\rangle_H. \end{aligned} \tag{3.68}$$

We also define the linear form

$$f(v) = \int_I \langle g(t), v_1(t)\rangle_H \,\mathrm{d}t + \langle h, v_2\rangle_H, \quad v = (v_1, v_2) \in \mathcal{Y}. \tag{3.69}$$

The variational form for the parametric, deterministic parabolic problems (3.67) reads: given $f \in \mathcal{Y}'$, find $u(y) : \Gamma \ni y \to \mathcal{X}$ such that

$$b(y; u, v) = f(v) \quad \forall v = (v_1, v_2) \in \mathcal{Y}, \quad y \in \Gamma. \tag{3.70}$$

Proposition 3.13. For each $y \in \Gamma$, the operator $B(y) \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ defined by $(B(y)w)(v) = b(y, w, v)$ is boundedly invertible. The norms of $B(y)$ and $B(y)^{-1}$ can be bounded uniformly by constants which only depend on a_-, a_+, T and the spaces $\mathcal{X}$ and $\mathcal{Y}$. In particular, the solution u of problem (3.70) is uniformly bounded in $\mathcal{X}$ for all $y \in \Gamma$. Moreover, the map $u(\cdot,\cdot,y) : \Gamma \to \mathcal{X}$ is measurable as a Bochner function.

The proof of this theorem can be found in Appendix A of Schwab and Stevenson (2009), or in Hoang and Schwab (2010*b*).

The variational formulation (3.70) is pointwise in the parameter and is the basis for sampling methods, such as the Monte Carlo method. For stochastic Galerkin approximation, (3.70) needs to be extended to the parameter space. To this end, we introduce Bochner spaces

$$\underline{\mathcal{X}} = L^2(\Gamma, \mu; \mathcal{X}) \quad \text{and} \quad \underline{\mathcal{Y}} = L^2(\Gamma, \mu; \mathcal{Y})$$

and note

$$\underline{\mathcal{X}} \simeq L^2(\Gamma, \mu) \otimes \mathcal{X}, \quad \underline{\mathcal{Y}} \simeq L^2(\Gamma, \mu) \otimes \mathcal{Y}.$$

With the bilinear form $B(\cdot,\cdot) : \underline{\mathcal{X}} \times \underline{\mathcal{Y}} \to \mathbb{R}$ and the linear form $F(\cdot) : \underline{\mathcal{Y}} \to \mathbb{R}$ defined by

$$B(u,v) = \int_\Gamma b(y,u,v)\mu(\mathrm{d}y) \quad \text{and} \quad F(v) = \int_\Gamma f(v)\mu(\mathrm{d}y). \tag{3.71}$$

we consider the variational problem: find

$$u \in \underline{\mathcal{X}} \quad \text{such that} \quad B(u,v) = F(v) \quad \text{for all} \quad v \in \underline{\mathcal{Y}}. \tag{3.72}$$

The Galerkin formulation is well-posed, as we see in the next result (see Hoang and Schwab (2010*b*)).

Proposition 3.14. Under Assumption 3.10, for every f as in (3.69) with $g \in L^2(I; V')$ and $h \in L^2(D)$, the parametric deterministic variational problem (3.72) admits a unique solution $u \in \underline{\mathcal{X}}$. Since the family $\{L_\nu\}_{\nu\in\mathfrak{F}}$ of tensor product polynomials forms a complete orthonormal system of $L^2(\Gamma, \mu)$, each $u \in \underline{\mathcal{X}}$ can be represented as

$$u = \sum_{\nu\in\mathcal{F}} u_\nu L_\nu, \tag{3.73}$$

where the coefficients $u_\nu \in \mathcal{X}$ are defined by

$$u_\nu = \int_\Gamma u(\cdot,\cdot,y)L_\nu(y)\mu(\mathrm{d}y) \in \mathcal{X},$$

the integral being understood as a Bochner integral of $\mathcal{X}$-valued functions over Γ.

With this result, we recover from the parametric, deterministic solution the random solution $u(t,x,\omega)$ by inserting the random variables $Y_m(\omega)$ for the coordinate vector $y \in \Gamma$.

Theorem 3.15. Under Assumptions 3.10, 3.11, for given $g \in L^2(I; V')$ and $h \in H$, the following variational problem admits a unique solution.

Find $u \in L^2(\Omega;\mathcal{X})$ such that, for every $v(t,x,\omega) = (v_1(t,x,\omega), v_2(x,\omega)) \in L^2(\Omega;\mathcal{Y})$,

$$\begin{aligned}\mathbb{E}\left\{\int_I \left\langle \frac{\mathrm{d}u}{\mathrm{d}t}(t,\cdot,\cdot), v_1(t,\cdot,\cdot)\right\rangle_H \mathrm{d}t\right\} \\ + \mathbb{E}\left\{\int_I \int_D a(x,\omega)\nabla u(t,x,\omega)\cdot\nabla v_1(t,x,\omega)\,\mathrm{d}x\,\mathrm{d}t\right\} \\ + \mathbb{E}\left\{\int_D u(0,x,\omega)v_2(x,\omega)\,\mathrm{d}x\right\} \\ = \mathbb{E}\left\{\int_I \int_D g(t,x)v_1(t,x,\omega)\,\mathrm{d}x\,\mathrm{d}t\right\} + \mathbb{E}\left\{\int_D h(x)v_2(x,\omega)\,\mathrm{d}x\right\}. \quad (3.74)\end{aligned}$$

This unique solution satisfies the *a priori* estimate

$$\|u\|_{L^2(\Omega;\underline{\mathcal{X}})} \le C(a)\big(\|g\|_{L^2(I;V')} + \|h\|_H\big). \quad (3.75)$$

Galerkin approximation

As in the elliptic case, we obtain GPC approximations by Galerkin projections onto suitable spaces of polynomials in $y \in \Gamma$. For every finite subset $\Lambda \subset \mathfrak{F}$ of cardinality not exceeding N, we define spaces of $\mathcal{X}$- and $\mathcal{Y}$-valued polynomial expansions

$$\underline{\mathcal{X}}_\Lambda = \left\{u_\Lambda(t,x,y) = \sum_{\nu\in\Lambda} u_\nu(t,x)L_\nu(y)\,;\ u_\nu \in \mathcal{X}\right\} \subset \underline{\mathcal{X}},$$

and

$$\underline{\mathcal{Y}}_\Lambda = \left\{v_\Lambda(t,x,y) = \sum_{\nu\in\Lambda} v_\nu(t,x)L_\nu(y)\,;\ v_\nu \in \mathcal{X}\right\} \subset \underline{\mathcal{Y}}.$$

In the Legendre basis $(L_\nu)_{\nu\in\mathfrak{F}}$, we write

$$v_{1\Lambda}(t,x,y) = \sum_{\nu\in\Lambda} v_{1\nu}(t,x)L_\nu(y) \quad \text{and} \quad v_{2\Lambda}(x,y) = \sum_{\nu\in\Lambda} v_{2\nu}(x)L_\nu(y),$$

respectively, where $v_\nu = (v_{1\nu}, v_{2\nu}) \in \mathcal{Y}$ for all $\nu \in \mathfrak{F}$. We consider the (semidiscrete) Galerkin approximation: find

$$u_\Lambda \in \underline{\mathcal{X}}_\Lambda \quad \text{such that} \quad B(u_\Lambda, v_\Lambda) = F(v_\Lambda) \quad \forall\, v_\Lambda \in \underline{\mathcal{Y}}_\Lambda. \quad (3.76)$$

Theorem 3.16. For any finite subset $\Lambda \subset \mathfrak{F}$ of cardinality exactly equal to N, the problem (3.76) corresponds to a coupled system of $N = \#\Lambda$ linear parabolic equations. Under Assumptions 3.10, 3.11, this coupled system of parabolic equations is stable uniformly with respect to $\Lambda \subset \mathfrak{F}$: for any $\Lambda \subset \mathfrak{F}$, problem (3.76) admits a unique solution $u_\Lambda \in \underline{\mathcal{X}}_\Lambda$ which satisfies

the *a priori* error bound

$$\|u - u_\Lambda\|_{\underline{\mathcal{X}}} \le c\left(\sum_{\nu\notin\Lambda} \|u_\nu\|_{\mathcal{X}}^2\right)^{1/2}.$$

Here, $u_\nu \in \mathcal{X}$ are the Legendre coefficients of the solution of the parametric problem in (3.73) and the constant c is independent of Λ.

Proof. To prove the uniform well-posedness of the coupled parabolic system resulting from the Galerkin discretization in Γ, we prove that the following inf-sup condition holds: there exist $\alpha, \beta > 0$ such that for *any* $\Lambda \subset \mathfrak{F}$,

$$\sup_{u_\Lambda\in\underline{\mathcal{X}}_\Lambda, v_\Lambda\in\underline{\mathcal{Y}}_\Lambda} \frac{|B(u_\Lambda, v_\Lambda)|}{\|u_\Lambda\|_{\underline{\mathcal{X}}}\|v_\Lambda\|_{\underline{\mathcal{Y}}}} \le \alpha < \infty, \tag{3.77}$$

$$\inf_{0\neq u_\Lambda\in\underline{\mathcal{X}}_\Lambda} \sup_{0\neq v_\Lambda\in\underline{\mathcal{Y}}_\Lambda} \frac{|B(u_\Lambda, v_\Lambda)|}{\|u_\Lambda\|_{\underline{\mathcal{X}}}\|v_\Lambda\|_{\underline{\mathcal{Y}}}} \ge \beta > 0, \tag{3.78}$$

$$\forall 0 \neq v_\Lambda \in \underline{\mathcal{Y}}_\Lambda : \sup_{0\neq u_\Lambda\in\underline{\mathcal{X}}_\Lambda} |B(u_\Lambda, v_\Lambda))| > 0, \tag{3.79}$$

where the constants α, β are in particular independent of the choice of $\Lambda \subset \mathfrak{F}$ (a proof can be found in the Appendix of Hoang and Schwab (2010*b*)).

The projected parametric deterministic parabolic problem (3.76) has a unique solution, and, in virtue of the independence of α, β from Λ, is well-posed and stable with stability bounds which are independent of the choice of $\Lambda \subset \mathfrak{F}$. Hence, the error incurred by this projection is quasi-optimal:

$$\begin{aligned}\|u - u_\Lambda\|_{\underline{\mathcal{X}}} &\le (1 + \beta^{-1}(\|g\|_{L^2(I;V')} + \|h\|_{L^2(D)})) \inf_{v_\Lambda\in\mathcal{X}} \|u - v_\Lambda\|_{\underline{\mathcal{X}}} \\ &\le c\left\|u - \sum_{\nu\in\Lambda} u_\nu L_\nu\right\|_{\underline{\mathcal{X}}} = c\left\|\sum_{\nu\notin\Lambda} u_\nu L_\nu\right\|_{\underline{\mathcal{X}}}.\end{aligned}$$

By the normalization of the tensorized Legendre polynomials L_ν and by Parseval's equality,

$$\left\|\sum_{\nu\notin\Lambda} u_\nu L_\nu\right\|_{\underline{\mathcal{X}}}^2 = \sum_{\nu\notin\Lambda} \|u_\nu\|_{\mathcal{X}}^2.$$

The conclusion then follows with $c = 1 + \beta^{-1}(\|g\|_{L^2(I;V')} + \|h\|_{L^2(D)})$. □

Best N-term GPC approximations

As in the elliptic case, Theorem 3.16 again suggests choosing $\Lambda \subset \mathfrak{F}$ to be the set of the largest N coefficients $\|u_\nu\|_{\mathcal{X}}$. Once (sharp and computable) *a priori* bounds for u_ν in $\mathcal{X}$ are known, one algorithmic strategy could be to optimize the sets $\Lambda \subset \mathfrak{F}$ according to these *a priori* bounds (one such strategy is outlined in Section 4.1 for the elliptic case). Alternatively,

an optimal, adaptive Galerkin method will yield iteratively quasi-optimal sequences Λ_N of active indices. We now determine such *a priori* bounds.

A best N-term convergence rate estimate in terms of N will result from these bounds once more using (3.13). Therefore, the convergence rate of spectral approximations such as (3.76) of the parabolic problem on the infinite-dimensional parameter space Γ is determined by the summability of the Legendre coefficient sequence $(\|u_\nu\|_{\underline{\mathcal{X}}})_{\nu\in\mathfrak{F}}$. We shall now prove that summability of this sequence is determined by that of the sequence $(\psi_j(x))_{j\in\mathbb{N}}$ in the input's fluctuation expansion (3.66). Throughout, Assumptions 3.10 and 3.11 will be required to hold. In addition, we shall make the following requirement.

Assumption 3.17. There exists $0 < p < 1$ such that

$$\sum_{j=1}^{\infty} \|\psi_j\|^p_{L^\infty(D)} < \infty. \tag{3.80}$$

Based on this assumption, in Hoang and Schwab (2010*b*) the following result was proved, with a proof along similar lines to that of the elliptic result Theorem 3.7.

Theorem 3.18. If Assumptions 3.10, 3.11 and 3.17 hold for some $0 < p < 1$, $\sum_{\nu\in\mathcal{F}} \|u_\nu\|^p_{\mathcal{X}}$ is finite.

Moreover, there is a sequence $(\Lambda_N)_{N\in\mathbb{N}} \subset \mathfrak{F}$ of index sets with cardinality not exceeding N such that the solutions u_{Λ_N} of the Galerkin semidiscretized problems (3.76) satisfy

$$\|u - u_{\Lambda_N}\|_{\underline{\mathcal{X}}} \leq CN^{-\sigma}, \quad \sigma = \frac{1}{p} - \frac{1}{2}.$$

This establishes a rate of convergence for best N-term GPC approximations in the stochastic Galerkin semidiscretization (3.76), which is analogous to the estimate (3.50) for the parametric elliptic problem. Note that Theorem 3.18 holds in the *semidiscrete setting*, *i.e.*, under the assumption that the Galerkin projections (3.76) can be computed exactly. To obtain actually *computable realizations* of such approximations, however, the coefficients need to be approximated in a hierarchical family of finite element spaces in the domain D. In order to obtain convergence rates analogous to Theorem 3.9 in the elliptic setting, regularity results for the parametric, parabolic problems (3.67) are required. Such results on best N-term approximation of expansions whose coefficients are measured in scales of spaces with additional smoothness in x and t are obtained in Hoang and Schwab (2010*b*).

3.3. Second-order hyperbolic problems

Analogous results may also be proved for linear, second-order hyperbolic problems with random coefficients. Such equations arise, for example, in the mathematical description of wave propagation in media with uncertain material properties. Below we recapitulate recent results from Hoang and Schwab (2010*a*) which are analogous to those for the elliptic and parabolic cases. One important distinction to the parabolic case, however, is that a space–time variational principle is not available so that some of the proofs in Hoang and Schwab (2010*a*) are significantly different from the elliptic case.

A class of wave equations with random coefficients

For $0 < T < \infty$, we consider in $I = (0,T)$ the following class of linear, second-order hyperbolic equations with random coefficients: let D be a bounded Lipschitz domain in $\mathbb{R}^d$. We define the space–time cylinder $Q_T = I \times D$. In Q_T, we consider the stochastic wave equation

$$\frac{\partial^2 u}{\partial t^2} - \nabla \cdot (a(x,\omega)\nabla u) = g(t,x), \quad u|_{\partial D \times I} = 0, \quad u|_{t=0} = g_1, \quad u_t|_{t=0} = g_2. \tag{3.81}$$

As before, we assume the coefficient $a(\omega, x)$ to be a random field on a probability space $(\Omega, \Sigma, \mathbb{P})$ over $L^\infty(D)$. The forcing g and initial data g_1 and g_2 are assumed to be deterministic. To ensure well-posedness of (3.81), we once more require Assumptions 3.10 and 3.11. To state the weak form of the initial boundary value problem (3.81), we let $V = H_0^1(D)$ and $H = L^2(D)$, and require

$$g \in L^2(I;H), \quad g_1 \in V, \quad g_2 \in H. \tag{3.82}$$

For the variational formulation of (3.81), we introduce the Bochner spaces

$$\mathcal{X} = L^2(I;V) \cap H^1(I;H) \cap H^2(I;V'), \quad \mathcal{Y} = L^2(I;V) \times V \times H. \tag{3.83}$$

A *weak solution of the hyperbolic initial boundary value problem* (3.81) is any function $u \in \mathcal{X}$ such that, for every $v = (v_0, v_1, v_2) \in \mathcal{Y}$,

$$\begin{aligned} \int_I \left\langle \frac{\mathrm{d}^2 u}{\mathrm{d}t^2}(t,\cdot), v_0(t,\cdot) \right\rangle_H \mathrm{d}t + \langle u(0), v_1 \rangle_V + \langle u_t(0), v_2 \rangle_H \\ + \int_I \int_D a(x,\omega)\nabla u(t,x,\omega) \cdot \nabla v_0(t,x)\, \mathrm{d}x\, \mathrm{d}t \\ = \int_I \int_D g(t,x) v_0(t,x)\, \mathrm{d}x\, \mathrm{d}t + \langle g_1, v_1 \rangle_V + \langle g_2, v_2 \rangle_H. \end{aligned} \tag{3.84}$$

We have the following result, from Hoang and Schwab (2010*a*).

Proposition 3.19. Under Assumption 3.10 and under condition (3.82), for every $\omega \in \Omega$, the initial boundary value problem (3.84) admits a unique

weak solution $u \in \mathcal{X}$. The following estimate holds:

$$\|u\|_{\mathcal{X}} \leq C(\|g\|_{L^2(I;H)} + \|g_1\|_V + \|g_2\|_H), \tag{3.85}$$

where the constant C depends only on T and on a_- and a_+ in Assumption 3.10.

We next present results on the stochastic Galerkin approximation of the initial boundary value problem (3.84).

Parametric deterministic wave equations
Given a forcing function $g(t,x)$ and initial data $g_1(x)$ and $g_2(x)$ satisfying (3.82), for each $y \in \Gamma$ we consider the parametric, deterministic initial boundary value problem

$$\begin{aligned} &\frac{\partial^2 u(t,x,y)}{\partial t^2} - \nabla_x \cdot (a(x,y)\nabla_x u(t,x,y)) = g(t,x) \text{ in } \ Q_T, \\ &u(t,x,y)|_{\partial D \times I} = 0, \ u|_{t=0} = g_1, \ u_t|_{t=0} = g_2, \end{aligned} \tag{3.86}$$

where the parametric coefficient $a(x,y)$ is defined as in the elliptic and parabolic cases in (3.66). Again, for each $y \in \Gamma$, we define the bilinear map $b : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ by

$$\begin{aligned} b(y; w, (v_0, v_1, v_2)) = &\int_I \left\langle \frac{\mathrm{d}^2 w}{\mathrm{d}t^2}(t,\cdot), v_0(t,\cdot) \right\rangle_H \mathrm{d}t \\ &+ \int_I \int_D a(x,y)\nabla w(t,x) \cdot \nabla v_0(t,x)\,\mathrm{d}x\,\mathrm{d}t \\ &+ \langle u(0), v_1\rangle_V + \langle u_t(0), v_2\rangle_H. \end{aligned} \tag{3.87}$$

We also define the linear form on $\mathcal{Y}$,

$$f(v) = \int_I \int_D g(t,x)v_0(t,x)\,\mathrm{d}x\,\mathrm{d}t + \langle g_1, v_1\rangle_V + \langle g_2, v_2\rangle_H.$$

The *pointwise parametric* (with respect to $y \in U$) variational formulation of the parametric, deterministic problem (3.86) then reads: find

$$u(y) \in \mathcal{X} : \quad b(y; u, v) = f(v) \quad \forall v = (v_0, v_1, v_2) \in \mathcal{Y}. \tag{3.88}$$

Note that here y-dependent data g, g_1 and g_2 would be equally admissible. This pointwise in y parametric variational formulation is well-posed *uniformly with respect to the parameter vector* y. More precisely, we have the following result (see Hoang and Schwab (2010*a*)).

Proposition 3.20. Under Assumption 3.10 and under conditions (3.82), for every $y \in \Gamma$, the problem (3.88) admits a unique weak solution $u(y) \in \mathcal{X}$. The parametric weak solutions $\{u(y)\,;\, y \in \Gamma\} \subset \mathcal{X}$ satisfy the *a priori* estimates

$$\forall y \in \Gamma : \quad \|u(\cdot,\cdot,y)\|_{\mathcal{X}} \leq C(\|g\|_{L^2(I;H)} + \|g_1\|_V + \|g_2\|_H), \tag{3.89}$$

where the constant C is independent of y. The map $u : \Gamma \to \mathcal{X}$ is strongly measurable as an $\mathcal{X}$-valued function.

An analogous result also holds for the Galerkin formulation of the parametric deterministic problem. To state this, we introduce the Bochner spaces $\underline{\mathcal{X}} = L^2(\Gamma, \mu; \mathcal{X})$ and $\underline{\mathcal{Y}} = L^2(\Gamma, \mu; \mathcal{Y})$ and define on these spaces the bilinear form $B(\cdot, \cdot) : \underline{\mathcal{X}} \times \underline{\mathcal{Y}} \to \mathbb{R}$ and the linear form $F(\cdot) : \underline{\mathcal{Y}} \to \mathbb{R}$ as

$$B(u, v) = \int_\Gamma b(y; u, v)\mu(\mathrm{d}y), \quad F(v) = \int_\Gamma f(v)\mu(\mathrm{d}y).$$

We may then consider the variational problem: find

$$u \in \underline{\mathcal{X}} \quad \text{such that} \quad B(u, v) = F(v) \quad \forall v \in \underline{\mathcal{Y}}. \tag{3.90}$$

Proposition 3.21. Under Assumptions 3.10 and 3.11, problem (3.90) admits a unique solution $u \in \underline{\mathcal{X}}$, *i.e.*, the parametric solution map belongs to to the Bochner space $L^2(\Gamma, \mu; \mathcal{X})$. Moreover, in terms of the orthonormal basis of $L^2(\Gamma, \mu)$ given by the tensorized Legendre polynomials $(L_\nu)_{\nu \in \mathfrak{F}}$, each function $u \in \underline{\mathcal{X}}$ can be written as an (unconditionally convergent in $\mathcal{X}$) expansion in tensorized Legendre polynomials:

$$u = \sum_{\nu \in \mathfrak{F}} u_\nu L_\nu, \quad u_\nu \in \mathcal{X}. \tag{3.91}$$

Semidiscrete Galerkin approximation

Spectral approximations of the parametric, deterministic wave equation (3.86) are once again obtained by projection onto finite linear combinations of tensorized Legendre polynomials. We briefly present the corresponding results from Hoang and Schwab (2010*a*).

For any set $\Lambda \subset \mathfrak{F}$ of finite cardinality, we define polynomial subspaces of $\underline{\mathcal{X}}$ and $\underline{\mathcal{Y}}$:

$$\underline{\mathcal{X}}_\Lambda = \left\{ u_\Lambda(t, x, y) = \sum_{\nu \in \Lambda} u_\nu(t, x) L_\nu(y) \,;\, u_\nu \in \mathcal{X} \right\} \subset \underline{\mathcal{X}},$$

and

$$\underline{\mathcal{Y}}_\Lambda = \left\{ v_\Lambda(t, x, y) = \sum_{\nu \in \Lambda} v_\nu(t, x) L_\nu(y) \,;\, v_\nu \in \mathcal{Y} \right\} \subset \underline{\mathcal{Y}}.$$

Denoting $v_\nu = (v_{0\nu}, v_{1\nu}, v_{2\nu})$, we may write the test functions $v \in \underline{\mathcal{Y}}_\Lambda$ componentwise as Legendre GPC expansions:

$$v_{0\Lambda}(t, x, y) = \sum_{\nu \in \Lambda} v_{0\nu}(t, x) L_\nu(y), \quad v_{i\Lambda}(x, y) = \sum_{\nu \in \Lambda} v_{i\nu}(x) L_\nu(y) \quad i = 1, 2.$$

We consider the following semidiscrete Galerkin projection of u onto $\underline{\mathcal{X}}_\Lambda$.

Find

$$u_\Lambda \in \underline{\mathcal{X}}_\Lambda \quad \text{such that} \quad B(u_\Lambda, v_\Lambda) = F(v_\Lambda) \quad \forall v_\Lambda \in \underline{\mathcal{Y}}_\Lambda. \tag{3.92}$$

Theorem 3.22. Under Assumptions 3.10 and 3.11, for every subset $\Lambda \subset \mathfrak{F}$ of finite cardinality, there exists a unique solution $u_\Lambda \in \underline{\mathcal{X}}_\Lambda$ to the Galerkin equations (3.92).

It is to be expected that the semidiscrete Galerkin approximations are quasi-optimal, *i.e.*, their error is controlled by the best approximation error. This is indeed once more the case. However, due to the lack of a space–time variational formulation, additional regularity of solutions, in particular point values with respect to t, is required. To this end, we introduce the space

$$\mathcal{Z} := H^1(I;V) \cap H^2(I;H) \subset C^0(\overline{I};V) \cap C^1(\overline{I};H). \tag{3.93}$$

Note that $\mathcal{Z} \subset \mathcal{X}$. The following error estimate for semidiscrete approximations of parametric solutions in $\mathcal{Z}$ holds.

Proposition 3.23. Assume that $u \in L^2(\Gamma,\mu;\mathcal{Z})$. Then, for all $\nu \in \mathfrak{F}$ the coefficient u_ν in (3.91) belongs to $\mathcal{Z}$. Assume further that for a subset $\Lambda \subset \mathfrak{F}$, $u_\Lambda \in L^2(\Gamma,\mu;\mathcal{Z})$. Then we have the error bound

$$\|u - u_\Lambda\|_{L^2(\Gamma,\mu;\mathcal{X})} \le c \left\| \sum_{\nu\in\mathfrak{F}\setminus\Lambda} u_\nu L_\nu \right\|_{L^2(\Gamma,\mu;\mathcal{Z})} = c \left(\sum_{\nu\in\mathfrak{F}\setminus\Lambda} \|u_\nu\|_{\mathcal{Z}}^2 \right)^{1/2}. \tag{3.94}$$

Here, the constant $c > 0$ depends only on the coefficient bounds a_- and a_+ in Assumption 3.10.

For a proof, we refer to Hoang and Schwab (2010*a*).

Proposition 3.23 once more implies *quasi-optimality* of the $L^2(\Gamma,\mu;\mathcal{X})$ projection $u_\Lambda \in \underline{\mathcal{X}}_\Lambda$ defined in (3.92). We note, however, that in its proof, the extra regularity $u \in L^2(\Gamma,\mu;\mathcal{Z})$ was required. It is therefore of interest to establish a regularity result for u_Λ which ensures $u \in L^2(\Gamma,\mu;\mathcal{Z})$ and, hence, implies the semidiscrete error bound (3.94). To this end, we recall the smoothness space $W \subset V$ defined in (3.52) and (3.53). On W, we define the W-seminorm and the W-norm as in (3.54). The role of the space W for the regularity of the solution of the parametric wave equation becomes clear from the following result.

Proposition 3.24. If Assumption 3.11 holds and if, moreover,

$$g \in H^1(I;H), \quad g_1 \in W, \quad g_2 \in V, \tag{3.95}$$

then for every $y \in \Gamma$ we have $u(\cdot,\cdot,y) \in \mathcal{Z}$, and its $\mathcal{Z}$-norm is bounded uniformly for all $y \in \Gamma$.

The Galerkin projections require the corresponding space–time solution space

$$\mathcal{W} = L^2(I;W) \cap H^1(I;V) \cap H^2(I;H), \tag{3.96}$$

where W is defined in (3.53). Note that $\mathcal{W} \subset \mathcal{Z}$ for $\mathcal{Z}$ defined in (3.93). We have the following regularity result for the Galerkin-projected GPC approximations of the parametric, deterministic wave equation (3.86).

Proposition 3.25. Under Assumptions 3.10 and 3.11, and if, in addition, $a(\cdot,\cdot) \in L^\infty(\Gamma; W^{1,\infty}(D))$, $g \in H^1(I;H)$, $g_1 \in W$ and $g_2 \in V$, then, for every subset $\Lambda \subset \mathfrak{F}$ of finite cardinality,

$$u_\Lambda \in L^2(\Gamma,\mu;\mathcal{W}) \subset L^2(\Gamma,\mu;\mathcal{Z}).$$

For the parametric wave equations, the extra regularity $u_\Lambda \in L^2(\Gamma,\mu;\mathcal{W})$ is therefore already required to ensure quasi-optimality of the Galerkin projections. It is also necessary in order to prove best N-term convergence rates analogous to those for the parametric elliptic and the parabolic problems. To state these results, we therefore require the following.

Assumption 3.26. We assume in (3.66) that $\bar{a} \in W^{1,\infty}(D)$ and $\psi_j \in W^{1,\infty}(D)$ are such that

$$\sum_{j=1}^{\infty} \|\psi_j\|_{W^{1,\infty}(D)} < \infty.$$

Moreover, we assume that, for some $0 < p < 1$,

$$\sum_{j=1}^{\infty} (\|\psi_j\|_{L^\infty(D)}^p + \|\nabla\psi_j\|_{L^\infty(D)}^p) < \infty.$$

Note that Assumption 3.26 implies Assumption 3.11. Under these assumptions we then have the following regularity and best N-term convergence result (see Hoang and Schwab (2010*a*) for a proof).

Proposition 3.27. Under Assumptions 3.10, 3.11 and 3.26, $\sum_{\nu\in\mathcal{F}} \|u_\nu\|_{\mathcal{Z}}^p$ is finite for the same value of $0 < p < 1$ as in these assumptions.

If, moreover, the compatibility condition (3.95) holds, there exists a sequence $(\Lambda_N) \subset \mathcal{F}$ of index sets with cardinality not exceeding N such that the solutions u_{Λ_N} of the Galerkin semidiscretized problem (3.92) satisfy, as $N \to \infty$, the error bound

$$\|u - u_{\Lambda_N}\|_{\mathcal{X}} \le CN^{-\sigma}, \quad \sigma = \frac{1}{p} - \frac{1}{2}.$$

Here, the constant C depends only on $\sum_{\nu\in\mathcal{F}} \|u_\nu\|_{\mathcal{Z}}^p$.

4. Sparse tensor discretizations

4.1. Sparse tensor stochastic Galerkin discretizations

We consider the diffusion equation with a stochastic diffusion coefficient, as in Section 2.3, discretized by the Legendre chaos basis; see Section 2.3. Foundations for subsequent Galerkin approximation were laid in Section 2.3. In Section 3.1, we proved convergence rates of optimal approximations by non-constructive methods. In this section, we consider Galerkin approximations in problem-adapted subspaces of $L^2(\Gamma,\mu;V)$ with a sparse tensor product structure. As in Bieri, Andreev and Schwab (2009), we show that these computable approximations achieve the optimal convergence rates up to logarithmic factors. We refer to Gittelson (2011*c*, 2011*d*) for an analysis of adaptive solvers with similar convergence properties.

For a bounded Lipschitz domain $D \subset \mathbb{R}^d$, we set $V = H^1_0(D)$. The dual space of V is $V' = H^{-1}(D)$. We denote the parameter domain by $\Gamma := [-1,1]^\infty$; μ is the product of uniform measures on Γ, as in Section 2.2.

Wavelet finite element discretization on the physical domain D

Finite element wavelets provide a stable, hierarchical multi-level basis of $V = H^1_0(D)$. We give a brief overview of the construction of such bases and refer to Cohen (2003) and Nguyen and Stevenson (2009) for details.

Let $D \subset \mathbb{R}^d$ be a bounded Lipschitz polyhedron with plane faces, and let $\mathcal{T}_0$ be a regular simplicial mesh of D. For all $\ell \in \mathbb{N}$, let $\mathcal{T}_\ell$ be the mesh of D constructed by ℓ regular refinements of the initial triangulation $\mathcal{T}_0$. We denote by $\mathcal{I}_\ell$ the interior nodes of $\mathcal{T}_\ell$ and by $\mathcal{N}_\ell := \mathcal{I}_\ell \setminus \mathcal{I}_{\ell-1}$ the new nodes on discretization level $\ell \in \mathbb{N}$. On the coarsest level, we have $\mathcal{N}_0 = \mathcal{I}_0$. Also, denote by $\mathcal{E}_\ell$ the set of edges of $\mathcal{T}_\ell$. Let V_ℓ^D be the space of continuous piecewise linear functions on the $\mathcal{T}_\ell$. We recall the standard approximation result

$$\|v - P_\ell^D v\|_{H^1_0(D)} \le C_\tau^D 2^{-\ell\tau} \|v\|_{H^{1+\tau}(D)} \quad \forall v \in H^{1+\tau}(D), \quad 0 \le \tau \le 1, \tag{4.1}$$

where P_ℓ^D is the orthogonal projection in $H^1_0(D)$ onto V_ℓ^D and $C_\tau^D > 0$ is independent of ℓ. The dimension of V_ℓ^D satisfies

$$\dim V_\ell^D \lesssim 2^{\ell d} \tag{4.2}$$

as $\ell \to \infty$.

The standard, so-called 'one-scale' basis of V_ℓ^D consists of the piecewise linear, nodal basis functions $(\lambda_n^\ell)_{n\in\mathcal{I}_\ell}$ determined by

$$\lambda_n^\ell(m) = \delta_{nm} \quad \forall m \in \mathcal{I}_\ell. \tag{4.3}$$

Following Nguyen and Stevenson (2009), we construct an alternative, hierarchical basis of V_ℓ^D.

We define auxiliary functions $(\eta_n^\ell)_{n\in\mathcal{I}_{\ell-1}}$ in V_ℓ^D satisfying

$$\langle \eta_n^\ell, \lambda_m^{\ell-1}\rangle_{L^2(D)} \simeq \delta_{nm}\|\eta_n^\ell\|_{L^2(D)}\|\lambda_m^{\ell-1}\|_{L^2(D)} \tag{4.4}$$

for all $n, m \in \mathcal{I}_{\ell-1}$ and all $\ell \in \mathbb{N}$. For $d = 1$,

$$\eta_n^\ell(m) := \begin{cases} 3 & m = n, \\ -1/2 & [m,n] \in \mathcal{E}_\ell, \\ 0 & \text{all other } m \in \mathcal{I}_\ell, \end{cases} \qquad n \in \mathcal{I}_{\ell-1}, \tag{4.5}$$

and for $d = 2$,

$$\eta_n^\ell(m) := \begin{cases} 14 & m = n, \\ -1 & [m,n] \in \mathcal{E}_\ell, \\ 0 & \text{all other } m \in \mathcal{I}_\ell, \end{cases} \qquad n \in \mathcal{I}_{\ell-1}. \tag{4.6}$$

Wavelets are given by $\psi_m^0 := \lambda_m^0$ for $m \in \mathcal{I}_0$ and

$$\psi_m^\ell := \lambda_m^\ell - \sum_{n\in\mathcal{I}\ell-1} \frac{\langle \lambda_m^\ell, \lambda_n^{\ell-1}\rangle_{L^2(D)}}{\langle \eta_n^\ell, \lambda_n^{\ell-1}\rangle_{L^2(D)}}\,\eta_n^\ell, \quad m \in \mathcal{N}_\ell. \tag{4.7}$$

We define the detail spaces

$$W_\ell^D := \operatorname{span}\{\psi_m^\ell\,;\ m \in \mathcal{N}_\ell\}, \quad \ell \in \mathbb{N}_0. \tag{4.8}$$

Then $W_0^D = V_0^D$, and V_ℓ^D is the direct sum of W_i^D over all $i \le \ell$. The wavelets whose construction was outlined above satisfy properties (W1)–(W5).

Hierarchical polynomial spaces

By Corollary 2.26, the solution u of (2.44) can be expanded in the Legendre chaos basis as

$$u(y) = \sum_{\nu\in\mathfrak{F}} u_\nu L_\nu(y) \in V, \quad y \in \Gamma. \tag{4.9}$$

For any subset $\Lambda \subset \mathfrak{F}$, we consider the truncated series

$$S_\Lambda u(y) := \sum_{\nu\in\Lambda} u_\nu L_\nu(y) \in V, \quad y \in \Gamma. \tag{4.10}$$

Since $(L_\nu)_{\nu\in\mathfrak{F}}$ is an orthonormal basis of $L^2(\Gamma,\mu;V)$ (see Corollary 2.16), $S_\Lambda u$ is the orthogonal projection of u onto the span of $(L_\nu)_{\nu\in\Lambda}$ in $L^2(\Gamma,\mu;V)$. We denote this space by

$$\langle \Lambda; V\rangle := \left\{ v(y) = \sum_{\nu\in\Lambda} v_\nu L_\nu(y)\,;\ (v_\nu)_{\nu\in\Lambda} \in \ell^2(\Lambda;V)\right\}. \tag{4.11}$$

For any $\Lambda \subset \mathfrak{F}$, $\langle \Lambda; V\rangle$ is a closed subspace of $L^2(\Gamma,\mu;V)$ since it is the kernel of $S_{\mathfrak{F}\setminus\Lambda}$, and thus it is a Hilbert space with the same norm. We abbreviate $\langle\Lambda\rangle$ for the subspace $\langle\Lambda;\mathbb{R}\rangle$ of $L^2(\Gamma,\mu)$.

Let $0 \le \tau \le 1$ and $\gamma > 0$. For all $k \in \mathbb{N}_0$, we define Λ_k^γ as the set of the first $\lfloor 2^{\gamma k}\rfloor$ indices $\nu \in \mathfrak{F}$ in a decreasing rearrangement of $(\|u_\nu\|_{H^{1+\tau}(D)})_{\nu\in\mathfrak{F}}$, *i.e.*, Λ_k^γ consists of the $\lfloor 2^{\gamma k}\rfloor$ indices ν for which $\|u_\nu\|_{H^{1+\tau}(D)}$ is largest. We assume that the same decreasing rearrangement is used for all $k \in \mathbb{N}_0$, such that $\Lambda_k^\gamma \subset \Lambda_{k+1}^\gamma$. The approximation spaces in $L^2(\Gamma,\mu)$ induced by the sets Λ_k^γ are

$$V_k^\Gamma := \langle \Lambda_k^\gamma\rangle, \quad k \in \mathbb{N}_0, \tag{4.12}$$

and the detail spaces are given by $W_0^\Gamma := V_0^\Gamma$ and

$$W_k^\Gamma := V_k^\Gamma \ominus V_{k-1}^\Gamma = \langle \Lambda_k^\gamma \setminus \Lambda_{k-1}^\gamma\rangle, \quad k \in \mathbb{N}. \tag{4.13}$$

Then $V_k^\Gamma \otimes V$ and $W_k^\Gamma \otimes V$ serve as approximation and detail spaces in $L^2(\Gamma,\mu;V)$.

Remark 4.1. The sets Λ_k^γ are not computationally accessible. We present a problem-adapted construction for an alternative sequence of index sets in Section 4.2.

Sparse tensor product spaces

The approximation spaces $V_k^\Gamma \subset L^2(\Gamma,\mu)$ and $V_\ell^D \subset V = H_0^1(D)$ can be combined to define finite-dimensional subspaces of $L^2(\Gamma,\mu;V)$. We recall that the Lebesgue–Bochner space $L^2(\Gamma,\mu;V)$ is isometrically isomorphic to the Hilbert tensor product $L^2(\Gamma,\mu)\otimes V$.

For any $L \in \mathbb{N}_0$, the tensor product of V_L^Γ and V_L^D can be expanded as

$$V_L^\Gamma \otimes V_L^D = \bigoplus_{k=0}^{L} W_k^\Gamma \otimes V_L^D = \bigoplus_{0\le\ell,k\le L} W_k^\Gamma \otimes W_\ell^D. \tag{4.14}$$

The sparse tensor product space of level $L \in \mathbb{N}_0$ is defined by restricting the last sum in (4.14) to only the most important component spaces,

$$V_L^\Gamma \widehat{\otimes} V_L^D := \bigoplus_{0\le\ell+k\le L} W_k^\Gamma \otimes W_\ell^D = \bigoplus_{k=0}^{L} W_k^\Gamma \otimes V_{L-k}^D. \tag{4.15}$$

Thus $V_L^\Gamma \widehat{\otimes} V_L^D$ is equal to $\mathcal{V}_N$ from (2.90), with $V_{N,\nu}$ equal to V_{L-k}^D if $\nu \in \Lambda_k^\gamma \setminus \Lambda_{k-1}^\gamma$, where $\Lambda_{-1}^\gamma := \emptyset$. Any element v of $V_L^\Gamma \widehat{\otimes} V_L^D$ is of the form

$$v(y,x) = \sum_{k=0}^{L} \sum_{\nu\in\Lambda_k^\gamma\setminus\Lambda_{k-1}^\gamma} \sum_{\ell=0}^{L-k} \sum_{n\in\mathcal{N}_\ell} c_{\nu,n}\psi_n^\ell(x)L_\nu(y), \quad x \in D, \quad y \in \Gamma. \tag{4.16}$$

We denote the Galerkin projection of u onto $V_L^\Gamma \widehat{\otimes} V_L^D$ by $\widehat{u}_L$. It is determined by (2.91) for $\mathcal{V}_N = V_L^\Gamma \widehat{\otimes} V_L^D$. By Proposition 2.27, $\widehat{u}_L$ is a quasi-optimal approximation of u in $V_L^\Gamma \widehat{\otimes} V_L^D$.

Convergence estimate

Due to the quasi-optimality of the Galerkin solution $\widehat{u}_L$ in $V_L^\Gamma \widehat{\otimes} V_L^D$, the convergence of $\widehat{u}_L$ to u is equivalent to that of the best approximation. Let $\widehat{P}_L$ denote the orthogonal projection in $L^2(\Gamma, \mu; V)$ onto $V_L^\Gamma \widehat{\otimes} V_L^D$. By (4.15), it has the form

$$\widehat{P}_L v = \sum_{k=0}^{L} \sum_{\nu \in \Lambda_k^\gamma \setminus \Lambda_{k-1}^\gamma} (P_{L-k}^D v_\nu) L_\nu, \quad v = \sum_{\nu \in \mathfrak{F}} v_\nu L_\nu \in L^2(\Gamma, \mu; V), \tag{4.17}$$

where $\Lambda_{-1}^\gamma := \emptyset$.

Theorem 4.2. Let $0 < \tau \le 1$, $\gamma > 0$, $0 < p \le 2$ and $s := 1/p - 1/2$. Furthermore, let a_+ and a_- be the bounds from (2.50). Then

$$\|u - \widehat{u}_L\|_{L^2(\Gamma,\mu;V)} \le C_t^D \lambda_{s\gamma,\tau}(L) \sqrt{\frac{a_+}{a_-}} \left(\sum_{\nu \in \mathfrak{F}} \|u_\nu\|_{H^{1+\tau}(D)}^p \right)^{1/p} 2^{-\min(s\gamma,\tau)L} \tag{4.18}$$

with

$$\lambda_{s\gamma,\tau}(L) = \begin{cases} \sqrt{2 + 2^{2\min(s\gamma,\tau)} L} & \text{always,} \\ \sqrt{2 + |2^{-2s\gamma} - 2^{-2\tau}|^{-1}} & \text{if } s\gamma \ne \tau. \end{cases} \tag{4.19}$$

Proof. By (4.17) and Parseval's identity in $L^2(\Gamma, \mu)$, using (4.1),

$$\begin{aligned} \|u - \widehat{P}_L u\|_{L^2(\Gamma,\mu;V)}^2 &= \sum_{k=0}^{L} \sum_{\nu \in \Lambda_k^\gamma \setminus \Lambda_{k-1}^\gamma} \|u_\nu - P_{L-k}^D u_\nu\|_V^2 + \sum_{\nu \in \mathfrak{F} \setminus \Lambda_L^\gamma} \|u_\nu\|_V^2 \\ &\le (C_\tau^D)^2 \left(\sum_{k=0}^{L} 2^{-2(L-k)\tau} \sum_{\nu \in \Lambda_k^\gamma \setminus \Lambda_{k-1}^\gamma} \|u_\nu\|_{H^{1+\tau}(D)}^2 + \sum_{\nu \in \mathfrak{F} \setminus \Lambda_L^\gamma} \|u_\nu\|_{H^{1+\tau}(D)}^2 \right). \end{aligned}$$

Applying Stechkin's lemma (3.13), we have for any $0 < p \le 2$ and $s := 1/p - 1/2$,

$$\left(\sum_{\nu \in \mathfrak{F} \setminus \Lambda_{k-1}^\gamma} \|u_\nu\|_{H^{1+\tau}(D)}^2 \right)^{1/2} \le (\#\Lambda_{k-1}^\gamma + 1)^{-s} \left(\sum_{\nu \in \mathfrak{F}} \|u_\nu\|_{H^{1+\tau}(D)}^p \right)^{1/p}$$

for $k = 1, \dots, L+1$. Since also

$$\left(\sum_{\nu \in \mathfrak{F}} \|u_\nu\|_{H^{1+\tau}(D)}^2 \right)^{1/2} \le \left(\sum_{\nu \in \mathfrak{F}} \|u_\nu\|_{H^{1+\tau}(D)}^p \right)^{1/p},$$

using $\#\Lambda_{k-1}^{\gamma} + 1 \geq 2^{\gamma(k-1)}$, it follows that

$$\|u - \widehat{P}_L u\|_{L^2(\Gamma,\mu;V)} \leq C_\tau^D \Sigma_L^{1/2} \left(\sum_{\nu \in \mathfrak{F}} \|u_\nu\|_{H^{1+\tau}(D)}^p \right)^{1/p}$$

with

$$\Sigma_L = 2^{-2L\tau} + 2^{-2s\gamma L} + \sum_{k=1}^{L} 2^{-2(L-k)\tau} 2^{-2s\gamma(k-1)}.$$

If $s\gamma = \tau$, then Σ_L simplifies to

$$\Sigma_L = (2^{2s\gamma} L + 2) 2^{-2s\gamma L} = (2^{2\tau} L + 2) 2^{-2\tau L}.$$

More generally, if $s\gamma \neq \tau$, estimating the sum by L times the maximal summand,

$$\Sigma_L \leq 2^{-2L\tau} + 2^{-2s\gamma L} + 2^{-2\min(s\gamma,\tau)(L-1)} L.$$

Alternatively, if $s\gamma \neq \tau$, summing the geometric series leads to

$$\Sigma_L = 2^{-2L\tau} + 2^{-2s\gamma L} + \frac{2^{-2s\gamma L} - 2^{-2\tau L}}{2^{-2s\gamma} - 2^{-2\tau}}.$$

If $s\gamma < \tau$, we have

$$\Sigma_L \leq 2^{-2s\gamma L} \left(2 + \frac{1}{2^{-2s\gamma} - 2^{-2\tau}} \right),$$

and similarly, if $s\gamma > \tau$,

$$\Sigma_L \leq 2^{-2\tau L} \left(2 + \frac{1}{2^{-2\tau} - 2^{-2s\gamma}} \right).$$

Then the claim follows using the quasi-optimality property from Proposition 2.27 with $\widehat{c} = a_+$ and $\check{c} = a_-^{-1}$. □

We express the convergence estimates in Theorem 4.2 with respect to the total number of degrees of freedom used to approximate u.

Lemma 4.3. The dimension of the sparse tensor product $V_L^\Gamma \widehat{\otimes} V_L^D$ scales as

$$\widehat{N}_L := \dim V_L^\Gamma \widehat{\otimes} V_L^D \lesssim \begin{cases} (L+1)2^{dL} & \text{if } \gamma = d, \\ 2^{\max(\gamma,d)L} & \text{if } \gamma \neq d, \end{cases} \quad \forall L \in \mathbb{N}_0. \tag{4.20}$$

Proof. Using the last expression in (4.15),

$$\widehat{N}_L = \dim V_L^\Gamma \widehat{\otimes} V_L^D = \sum_{k=0}^{L} \#(\Lambda_k^\gamma \setminus \Lambda_{k-1}^\gamma) \dim V_{L-k}^D.$$

By definition, $\#(\Lambda_k^\gamma \setminus \Lambda_{k-1}^\gamma) \le \#\Lambda_k^\gamma \le 2^{\gamma k}$, and by (4.2), $\dim V_{L-k}^D \lesssim 2^{d(L-k)}$. Therefore,

$$\widehat{N}_L \lesssim \sum_{k=0}^{L} 2^{\gamma k} 2^{d(L-k)} = 2^{dL} \sum_{k=0}^{L} 2^{(\gamma-d)k}.$$

If $\gamma = d$, then all the summands are one, and we arrive at

$$\widehat{N}_L \lesssim (L+1)2^{dL}.$$

Otherwise, we sum the geometric series to find

$$\widehat{N}_L \lesssim 2^{dL} \frac{2^{(\gamma-d)L} - 1}{2^{\gamma-d} - 1} \lesssim 2^{\max(\gamma,d)L}$$

with constants independent of L. □

Corollary 4.4. In the setting of Theorem 4.2, if $\gamma \neq d$,

$$\begin{aligned} &\|u - \widehat{u}_L\|_{L^2(\Gamma,\mu;V)} \\ &\quad \le C_\tau^D \lambda_{s\gamma,\tau}(L) \sqrt{\frac{a_+}{a_-}} \left(\sum_{\nu\in\mathfrak{F}} \|u_\nu\|_{H^{1+\tau}(D)}^p \right)^{1/p} \widehat{N}_L^{-\min(s,s\gamma/d,\tau/\gamma,\tau/d)}. \end{aligned} \tag{4.21}$$

If $\gamma = d$, then

$$\begin{aligned} &\|u - \widehat{u}_L\|_{L^2(\Gamma,\mu;V)} \\ &\quad \le C_\tau^D \lambda_{sd,\tau}(L) \sqrt{\frac{a_+}{a_-}} \left(\sum_{\nu\in\mathfrak{F}} \|u_\nu\|_{H^{1+\tau}(D)}^p \right)^{1/p} (L+1)^{\min(s,\tau/d)} \widehat{N}_L^{-\min(s,\tau/d)}. \end{aligned} \tag{4.22}$$

Proof. The claim follows from Theorem 4.2 using Lemma 4.3 to estimate 2^L from below by a power of $\widehat{N}_L$. □

Remark 4.5. Up to logarithmic factors, the convergence rates in Corollary 4.2 with respect to the total number of degrees of freedom $\widehat{N}_L$ reaches the optimum $\min(s, \tau/d)$ from Theorem 3.9 for two choices of γ. Therefore, assuming a sparse tensor product structure of the Galerkin subspace of $L^2(\Gamma, \mu; V)$ does not significantly deteriorate the convergence behaviour. If $\gamma = \tau/s \neq d$, then $s\gamma/d = \tau/d$ and $\tau/\gamma = s$, so (4.21) becomes

$$\|u - \widehat{u}_L\|_{L^2(\Gamma,\mu;V)} \le C_\tau^D \sqrt{2 + 2^{2\tau} L} \sqrt{\frac{a_+}{a_-}} \left(\sum_{\nu\in\mathfrak{F}} \|u_\nu\|_{H^{1+\tau}(D)}^p \right)^{1/p} \widehat{N}_L^{-\min(s,\tau/d)}. \tag{4.23}$$

As already stated in the corollary, $\gamma = d$ also leads to this convergence rate, albeit with an additional factor that is logarithmic in $\widehat{N}_L$. Nevertheless, this choice has the advantage of being independent of the regularity parameters s and τ.

Remark 4.6. The above derivation of Theorem 4.2 and Corollary 4.4 corrects some mathematical inaccuracies in the proof of Proposition 3.5 of Bieri *et al.* (2009). In particular, it is apparent from the above result that the sets Λ_k^γ for $k = 0, \dots, L-1$ should be defined to contain the indices $\nu \in \mathfrak{F}$ for which $\|u_\nu\|_{H^{1+\tau}(D)}$ is maximal, and not $\|u_\nu\|_V$ as claimed in Bieri *et al.* (2009). For Λ_L^γ, it is also reasonable to use $\|u_\nu\|_V$; doing so sacrifices the sparse tensor product structure, but may reduce the total number of degrees of freedom.

Remark 4.7. Corollary 4.4 should be compared with the convergence of full tensor product discretizations. In this case, the dimension of the Galerkin subspace of $L^2(\Gamma, \mu; V)$ scales as $2^{(\gamma+d)L}$, and the error satisfies Theorem 4.2 without the logarithmic terms. Therefore, the convergence rate of the error with respect to the dimension is $\min(s\gamma, \tau)/(\gamma + d)$. For example, if $\gamma = d$, this is $\min(s, \tau/d)/2$, which is just half of the rate from (4.22). If $\gamma = \tau/s$, then the convergence rate $(1/s + d/\tau)^{-1}$ is half of the harmonic mean of s and τ/d, which is only a minor improvement.

Remark 4.8. The convergence rates in Corollary 4.4 and Remark 4.5 are limited by the order of convergence $\tau/d \le 1/d$ of linear finite elements on D. This can be overcome by using higher-order finite elements. For example, if $u \in \ell^p(\mathfrak{F}; H^{1+\tau}(D))$ with $s = 1/p - 1/2$ and $\tau = sd$, using piecewise polynomial finite elements of degree $\lceil \tau \rceil$ leads to a convergence rate of s with respect to the total number $\widehat{N}_L$ of degrees of freedom.

4.2. Algorithmic aspects of polynomial chaos

Hierarchical index sets

The index sets Λ_k^γ from Section 4.1 are not computationally accessible. We consider a different family of index sets that can be constructed explicitly. For any sequence $\eta = (\eta_m)_{m=1}^\infty \in \mathbb{R}^\infty$, since $\eta_m^0 = 1$ for all $m \in \mathbb{N}$,

$$\eta^\nu := \prod_{m=1}^\infty \eta_m^{\nu_m} = \prod_{m \in \operatorname{supp} \nu} \eta_m^{\nu_m}, \quad \nu \in \mathfrak{F}. \tag{4.24}$$

We assume that for some $\eta \in \mathbb{R}^\infty$ with $\eta_m \in (0, 1)$, $\eta_m \to 0$, up to a constant factor, (4.24) is an estimate for $\|u_\nu\|_{H^{1+t}(D)}$ for all $\nu \in \mathfrak{F}$; see Section 4.2. Then, for any $\epsilon > 0$,

$$\Lambda_\epsilon(\eta) := \{\nu \in \mathfrak{F} \,;\, \eta^\nu \ge \epsilon\}. \tag{4.25}$$

Due to the assumptions $0 < \eta_m < 1$ and $\eta_m \to 0$, $\Lambda_\epsilon(\eta)$ is a finite subset of $\mathfrak{F}$ for all $\epsilon > 0$. The construction of the sets (4.25) does not require exact knowledge of $\|u_\nu\|_{H^{1+t}(D)}$, only estimates with a certain structure. Also, the sets $\Lambda_\epsilon(\eta)$ differ from Λ_k^γ in that they are defined by a thresholding tolerance instead of a prescribed cardinality.

It is clear from the definition that the sets $\Lambda_\epsilon(\eta)$ are monotonic in both parameters ϵ and η. If $\epsilon \le \bar{\epsilon}$, then $\Lambda_\epsilon(\eta) \supseteq \Lambda_{\bar{\epsilon}}(\eta)$. Similarly, if $\eta \in \mathbb{R}^\infty$ with $\eta_m \le \bar{\eta}_m < 1$ for all $m \in \mathbb{N}$, then $\Lambda_\epsilon(\eta) \subseteq \Lambda_\epsilon(\bar{\eta})$.

For a $\sigma > 0$, let $\eta^\sigma \in \mathbb{R}^\infty$ be defined by $\eta^\sigma_m := (m+1)^{-\sigma}$, $m \in \mathbb{N}$. For η^σ, sharp asymptotics on the cardinality of the index sets $\Lambda_\epsilon(\eta^\sigma)$ follow from results on integer factorization (Bieri *et al.* 2009, Proposition 4.5).

Proposition 4.9. For any $\sigma > 1/2$, as $\epsilon \to 0$,

$$\#\Lambda_\epsilon(\eta^\sigma) \simeq \epsilon^{-1/\sigma} \frac{\mathrm{e}^{2\sqrt{\sigma^{-1}\log\epsilon^{-1}}}}{2\sqrt{\pi}(\sigma^{-1}\log\epsilon^{-1})^{3/4}}. \tag{4.26}$$

In particular, $\{(\eta^\sigma)^\nu\}_{\nu\in\mathfrak{F}} \in \ell^p(\mathfrak{F})$ for any $p > 1/\sigma$.

Proof. We observe that for all $\epsilon > 0$ and $\sigma > 0$, $\Lambda_\epsilon(\eta^\sigma) = \Lambda_{\epsilon^{1/\sigma}}(\eta^1)$. Let $f(n)$ denote the number of multiplicative partitions of $n \in \mathbb{N}$, disregarding the order of the factors. For example, $f(12) = 4$ since 12, $2\cdot 6$, $2\cdot 2\cdot 3$ and $3\cdot 4$ are the only factorizations of 12. Sharp asymptotics for

$$F_\Sigma(x) := \sum_{n=1}^{\lfloor x\rfloor} f(n)$$

as $x \to \infty$ were obtained in Canfield, Erdős and Pomerance (1983), based on earlier work (Oppenheim 1927, Szekeres and Turán 1933). The result required here is

$$F_\Sigma(x) = F(x)(1 + \mathcal{O}(1/\log n)) \quad \text{with} \quad F(x) = \frac{x\,\mathrm{e}^{2\sqrt{\log x}}}{2\sqrt{\pi}(\log x)^{3/4}}.$$

Let $\epsilon > 0$ and $\nu \in \Lambda_\epsilon(\eta^1)$. By definition,

$$n := \prod_{m=1}^{\infty} (m+1)^{\nu_m} \le \frac{1}{\epsilon}.$$

Since n is an integer, the index ν represents a multiplicative partition of an integer $n \le 1/\epsilon$. Conversely, for any integer $n \le 1/\epsilon$, any multiplicative partition of n is of the form $n = \prod_m (m+1)^{\nu_m}$ for a $\nu \in \mathfrak{F}$, and therefore $\nu \in \Lambda_\epsilon(\eta^1)$. Consequently,

$$\#\Lambda_\epsilon(\eta^\sigma) = \#\Lambda_{\epsilon^{1/\sigma}}(\eta^1) = F_\Sigma(\epsilon^{-1/\sigma}) \simeq F(\epsilon^{-1/\sigma})$$

as $\epsilon \to 0$, which implies (4.26).

Note that $\epsilon^{-1/\sigma}$ is the dominating term in (4.26). Let $z = \sqrt{\sigma^{-1}\log\epsilon^{-1}}$. Then (4.26) is bounded by a constant times

$$\mathrm{e}^{z^2}\mathrm{e}^{2z}\mathrm{e} = \mathrm{e}^{(z+1)^2}.$$

For any $\kappa > 1$, we have $z + 1 \le \sqrt{\kappa}\, z$ for all $z \ge (\sqrt{\kappa} - 1)^{-1}$, and thus

$$\mathrm{e}^{(z-1)^2} \le \mathrm{e}^{\kappa z^2} = \epsilon^{-\kappa/\sigma} \quad \forall z \ge (\sqrt{\kappa} - 1)^{-1}.$$

Consequently, by (4.26), for all $\kappa > 1$,

$$\#\Lambda_\epsilon(\eta^\sigma) = \#\{\nu \in \mathfrak{F}\,;\ (\eta^\sigma)^\nu \ge \epsilon\} \lesssim \epsilon^{-\kappa/\sigma}.$$

This is a characterizing property of the weak Lebesgue sequence space $\ell^q_w(\mathfrak{F})$ with $q = \kappa/\sigma$ (DeVore 1998). Since $\ell^q_w(\mathfrak{F}) \subset \ell^p(\mathfrak{F})$ for all $p > q$, and since $\kappa > 1$ is arbitrary, it follows that $\{(\eta^\sigma)^\nu\}_{\nu\in\mathfrak{F}} \in \ell^p(\mathfrak{F})$ for all $p > 1/\sigma$. □

As demonstrated in Bieri *et al.* (2009, Figure 4.1), (4.26) provides a good approximation for thresholds ϵ as large as 10^{-2}.

Proposition 4.9 gives bounds on the size of index sets $\Lambda_\epsilon(\eta)$. In order to control the complexity of the numerical construction of these index sets, it is also important to bound the length $\#\operatorname{supp}\nu$ of indices $\nu \in \Lambda_\epsilon(\eta)$. To this end, we define the maximal dimension reached by $\Lambda_\epsilon(\eta)$,

$$M_\epsilon(\eta) := \max\{m \in \mathbb{N}\,;\ \eta_m \ge \epsilon\} = \max\{m \in \mathbb{N}\,;\ \epsilon_m \in \Lambda_\epsilon(\eta)\}, \tag{4.27}$$

where $\epsilon_m \in \mathfrak{F}$ denotes the Kronecker sequence $(\epsilon_m)_n = \delta_{mn}$.

Proposition 4.10. Let $\eta = (\eta_m)_{m=1}^\infty$ satisfy $0 < \eta_{m+1} \le \eta_m < 1$, $\eta_m \to 0$, and

$$c_1 m^{-\sigma_1} \le \eta_m \le c_2 m^{-\sigma_2} \quad \forall m \in \mathbb{N}, \tag{4.28}$$

with $c_1, c_2 > 0$ and $\sigma_1 \ge \sigma_2 > 0$. Then there is a constant $C > 0$ such that

$$\#\operatorname{supp}\nu \le C \log^+ M_\epsilon(\eta) \le C \log \#\Lambda_\epsilon(\eta) \quad \forall \nu \in \Lambda_\epsilon(\eta) \tag{4.29}$$

for all $0 < \epsilon \le 1$, *i.e.*, whenever $\Lambda_\epsilon(\eta) \ne \emptyset$.

Proof. By (4.28) and (4.27), due to the definition of $\Lambda_\epsilon(\eta)$,

$$\epsilon > \eta_{M_\epsilon(\eta)+1} \ge c_1 (M_\epsilon(\eta) + 1)^{-\sigma_1},$$

and therefore $M_\epsilon(\eta) + 1 \le (\epsilon/c_1)^{-1/\sigma_1}$. Since $\eta_m \le \eta_1 < 1$ for all $m \in \mathbb{N}$, we can assume without loss of generality that $c_2 = 1$, possibly at the cost of decreasing σ_2. Then, for all $\nu \in \mathfrak{F}$,

$$\eta^\nu = \prod_{m\in\operatorname{supp}\nu} \eta_m^{\nu_m} \le \prod_{m\in\operatorname{supp}\nu} m^{-\nu_m\sigma_2} \le [(\#\operatorname{supp}\nu)!]^{-\sigma_2}.$$

We note that, by Stirling's approximation, $(x+1)^\tau = o(\lceil \tau \log x \rceil!)$ for any $\tau > 0$ as $x \to \infty$. Indeed, abbreviating $n := \lceil \tau \log x \rceil$, since $n! \ge \sqrt{2\pi} n^n \mathrm{e}^{-n}$,

$$\frac{(x+1)^\tau}{\lceil \tau \log x \rceil!} \le \frac{1}{\sqrt{2\pi}} \mathrm{e}^{\tau \log x + n - n \log n} \to 0, \quad x \to \infty.$$

Consequently, also $(\lceil \tau \log x \rceil!)^{-1} = o((x+1)^{-\tau})$.

Suppose that $\nu \in \mathfrak{F}$ with $\#\operatorname{supp}\nu \geq (1+\sigma_1/\sigma_2)\log M_\epsilon(\eta)$. Then, by the above estimates with $x = M_\epsilon(\eta)$ and $\tau = 1+\sigma_1/\sigma_2$,

$$\begin{aligned}\eta^\nu &\leq \left(\left\lceil\left(1+\frac{\sigma_1}{\sigma_2}\right)\log M_\epsilon(\eta)\right\rceil!\right)^{-\sigma_2}\\ &= o\big((M_\epsilon(\mu)+1)^{-\sigma_1-\sigma_2}\big) = o(\epsilon^{(\sigma_1+\sigma_2)/\sigma_1}) = o(\epsilon).\end{aligned}$$

The $o(\cdot)$ is with respect to $M_\epsilon(\eta)\to\infty$, which by (4.27) is equivalent to $\epsilon\to 0$. Therefore, there is an $\epsilon_0 > 0$ such that, for all $0<\epsilon\leq\epsilon_0$, $\#\operatorname{supp}\nu \geq (1+\sigma_1/\sigma_2)\log M_\epsilon(\eta)$ implies $\eta^\nu < \epsilon$. Equivalently,

$$\#\operatorname{supp}\nu \leq \left(1+\frac{\sigma_1}{\sigma_2}\right)\log M_\epsilon(\eta) \quad \forall \nu \in \Lambda_\epsilon(\eta).$$

As there are only finitely many distinct sets $\Lambda_\epsilon(\eta)$ with $\epsilon > \epsilon_0$, this estimate holds for all $\epsilon > 0$, with a larger constant and $\log^+(n) := \max(\log(n), 0)$ in place of $\log(n)$ to accommodate $M_\epsilon(\eta) = 0$. The second part of (4.29) follows using the observation that $M_\epsilon(\eta) \leq \#\Lambda_\epsilon(\eta)$ since the Kronecker sequences $(\epsilon_m)_n = \delta_{mn}$ are in $\Lambda_\epsilon(\eta)$ for $m = 1,\dots,M_\epsilon(\eta)$. □

Numerical construction

By Proposition 4.10, any $\nu \in \Lambda_\epsilon(\eta)$ for $\eta = (\eta_m)_{m=1}^\infty$ satisfying (4.28) can be stored in $\mathcal{O}(\log\#\Lambda_\epsilon(\eta))$ memory in the form

$$\{(m,\nu_m)\,;\ m\in\operatorname{supp}\nu\}, \tag{4.30}$$

assuming that an arbitrary integer can be stored in a single storage location. Therefore, the total memory required to store the full set $\Lambda_\epsilon(\eta)$ is of the order $\#\Lambda_\epsilon(\eta)\log\#\Lambda_\epsilon(\eta)$. We show how $\Lambda_\epsilon(\eta)$ can be constructed in $\mathcal{O}(\#\Lambda_\epsilon(\eta)\log\#\Lambda_\epsilon(\eta))$ time.

For any sequence $c = (c_m)_{m=1}^\infty$, we define the translation

$$\tau_1 c := (c_{m+1})_{m=1}^\infty. \tag{4.31}$$

Furthermore, we define the concatenation of integers with subsets of $\mathfrak{F}$. For any $n\in\mathbb{N}_0$ and $\Lambda\subset\mathfrak{F}$,

$$[n,\Lambda] := \{\nu\in\mathfrak{F}\,;\ \nu_1 = n,\ \tau_1\nu\in\Lambda\} = \{[n,\nu]\,;\ \nu\in\Lambda\}, \tag{4.32}$$

where $[n,\nu] := (n,\nu_1,\nu_2,\dots)$. We observe that

$$\Lambda_\epsilon(\eta) = \bigsqcup_{n=0}^{N_\epsilon(\eta)} [n, \Lambda_{\epsilon\eta_1^{-n}}(\tau_1\eta)] \tag{4.33}$$

with

$$N_\epsilon(\eta) := \max\{m\in\mathbb{N}_0\,;\ \eta_1^n \geq \epsilon\} = \left\lfloor\frac{\log\epsilon}{\log\eta_1}\right\rfloor. \tag{4.34}$$

This suggests a recursive construction of $\Lambda_\epsilon(\eta)$. The precise algorithm is given in $\texttt{Construct}(\eta, \epsilon)$.

We store indices $\nu \in \mathfrak{F}$ in the sparse form (4.30). We append to each $\nu \in \Lambda_\epsilon(\eta)$ the value η^ν, which we compute during the construction of $\Lambda_\epsilon(\eta)$. In the concatenation $[n, \Lambda_\delta(\tau_1\eta)]$, these values are updated by

$$\eta^{[n,\nu]} = \eta_1^n(\tau_1\eta)^\nu, \quad \nu \in \Lambda_\delta(\tau_1\eta), \tag{4.35}$$

where $(\tau_1\eta)^\nu$ is known from the construction of $\Lambda_\delta(\tau_1\eta)$ for all $\nu \in \Lambda_\delta(\tau_1\eta)$.

$\texttt{Construct}(\eta, \epsilon) \mapsto \Lambda_\epsilon(\eta)$

---

if $\eta_1 < \epsilon$ **then**
  if $\epsilon > 1$ **then**
    return $\emptyset$
  else
    return $\{0\}$
  end
end
$N \longleftarrow \left\lfloor \dfrac{\log \epsilon}{\log \eta_1} \right\rfloor$
for $n = 0, 1, \ldots, N$ **do**
  $\Lambda_n \longleftarrow \texttt{Construct}(\tau_1\eta, \epsilon\eta_1^{-n})$
  $\Lambda_n \longleftarrow [n, \Lambda_n]$
end
$\Lambda \longleftarrow \bigsqcup_{n=0}^{N} \Lambda_n$
return Λ

Lemma 4.11. For any $\epsilon > 0$ and any η satisfying the assumptions of Proposition 4.10, $\texttt{Construct}(\eta, \epsilon)$ constructs $\Lambda_\epsilon(\eta)$ at a computational cost of $\mathcal{O}(\#\Lambda_\epsilon(\eta) \log \#\Lambda_\epsilon(\eta))$.

Proof. It is clear from (4.33) that $\texttt{Construct}(\eta, \epsilon)$ does construct $\Lambda_\epsilon(\eta)$. Due to the sparse storage format (4.30), concatenation $[0, \nu]$ of $\nu \in \mathfrak{F}$ with zero does not involve any work. Therefore, each index $\nu \in \Lambda_\epsilon(\eta)$ is constructed in $\#\operatorname{supp}\nu$ identical steps. The union operations can be performed in constant time, for example, with a linked list data structure. The claim follows since $\#\operatorname{supp}\nu = \mathcal{O}(\log \#\Lambda_\epsilon(\eta))$ by Proposition 4.10. □

Remark 4.12. For the construction of a sparse tensor product similar to (4.15), we need not only one index set $\Lambda_\epsilon(\eta)$, but a hierarchical sequence of such sets. For a sequence of thresholds

$$\epsilon_0 \geq \epsilon_1 \geq \cdots \geq \epsilon_L = \epsilon > 0, \tag{4.36}$$

we use the index sets $\Lambda_k := \Lambda_{\epsilon_k}(\eta)$, which satisfy

$$\emptyset =: \Lambda_{-1} \subseteq \Lambda_0 \subseteq \Lambda_1 \subseteq \cdots \subseteq \Lambda_L = \Lambda_\epsilon(\eta). \tag{4.37}$$

The detail spaces (4.13) induced by (4.37) are $W_k^\Gamma = \langle \Lambda_k \setminus \Lambda_{k-1} \rangle$. Their construction requires the partitioning of $\Lambda_\epsilon(\eta)$ into

$$\Lambda_\epsilon(\eta) = \bigsqcup_{k=0}^{L} \Lambda_k \setminus \Lambda_{k-1}. \tag{4.38}$$

This can be done after constructing $\Lambda_\epsilon(\eta)$ by sorting the indices $\nu \in \Lambda_\epsilon(\eta)$ with respect to the values η^ν, which are computed during the construction of the index set using the recursion (4.35). Once sorted, the indices ν are easily assigned to the partitions $\Lambda_k \setminus \Lambda_{k-1}$ by comparing η^ν to ϵ_k. Alternatively, for a parameter $\gamma > 0$ as in Section 4.1, the thresholds $\epsilon_0, \dots, \epsilon_{L-1}$ can be chosen such that $\#\Lambda_k = \lfloor 2^{\gamma k} \rfloor$ for $k = 0, \dots, L-1$. The total work is of order $\mathcal{O}(\#\Lambda_\epsilon(\eta) \log \#\Lambda_\epsilon(\eta) + L)$ as $\epsilon \to 0$ or $L \to \infty$.

Remark 4.13. The stochastic Galerkin matrix has the form (2.81), *i.e.*, the Legendre coefficients of the Galerkin projection $\widehat{u}_L$ satisfy a discretized version of the system of equations (2.88), which has direct dependencies between indices $\nu, \bar{\nu} \in \Lambda_\epsilon(\nu)$ that are identical in all dimensions but one, and differ by one in this dimension. We call ν and $\bar{\nu}$ *neighbours* if there is an $\bar{m} \in \mathbb{N}$ such that $|\nu_m - \bar{\nu}_m| = \delta_{m\bar{m}}$. Fast access to the neighbours in $\Lambda_\epsilon(\eta)$ of $\nu \in \Lambda_\epsilon(\eta)$ is crucial to solving the Galerkin system. The neighbourhood data are stored efficiently in a directed graph, in which there is an edge from ν to $\bar{\nu}$ if there is an $\bar{m} \in \operatorname{supp} \nu$ such that $\bar{\nu} = \nu - \epsilon_{\bar{m}}$, where $\epsilon_{\bar{m}}$ is the Kronecker sequence $(\epsilon_{\bar{m}})_m = \delta_{m\bar{m}}$. Then there are exactly $\#\operatorname{supp} \nu$ edges starting at $\nu \in \Lambda_\epsilon(\eta)$. It is useful to label each edge by the appropriate $\bar{m} \in \mathbb{N}$. The routine `Construct`(η, ϵ) can be extended to compute the neighbourhood relations during the construction of $\Lambda_\epsilon(\eta)$ using the observation that ν and $\bar{\nu}$ are neighbours as above if and only if, at a call of `Construct` at recursion depth $\bar{m}$, ν and $\bar{\nu}$ are constructed by appending n and $n-1$, respectively, to the same index $\widetilde{\nu} \in \mathfrak{F}$, for some $n \in \mathbb{N}$, and all subsequent additions are the same.

Choice of parameters

The preceding sections leave open the choice of $\eta = (\eta_m)_{m=1}^\infty$. Due to Theorem 4.2, η^ν should approximate $\|u_\nu\|_{H^{1+\tau}(D)}$ for the unknown u and some $0 \le \tau \le 1$; see Remark 4.6. A reasonable choice is

$$\eta_m = \frac{\alpha_m}{\bar{a}_-} \|\varphi_m\|_{W^{\tau,\infty}(D)}, \quad m \in \mathbb{N}, \tag{4.39}$$

using the notation from Section 2.3. We refer to Bieri *et al.* (2009) for numerical experiments with this and other choices of η_m and $\tau = 0$. Better

a priori choices of η_m may be obtainable from sharper *a priori* bounds on $\|u_\nu\|_{H^{1+\tau}(D)}$.

As mentioned in Remark 4.12, the thresholds $\epsilon_0, \dots, \epsilon_{L-1}$ can be chosen such that $\#\Lambda_k = \lfloor 2^{\gamma k} \rfloor$ for $k = 0, \dots, L-1$ with a parameter $\gamma > 0$ as in Section 4.1.

4.3. Sparse tensor stochastic collocation

Stochastic collocation is an alternative to the stochastic Galerkin method. It is also based on a polynomial approximation in the parameter domain. However, the Galerkin projection is replaced by a suitable interpolation operator. A sparse tensor product construction similar to that presented in Section 4.1 is possible in this setting.

We consider the diffusion equation with a stochastic diffusion coefficient, as in Section 2.3. The following discussion is based primarily on Bieri (2009*a*, 2009*b*).

Stochastic collocation

We recall the spaces of polynomials on $\Gamma = [-1,1]^\infty$ from Section 4.1,

$$\langle \Lambda \rangle = \left\{ v(y) = \sum_{\nu \in \Lambda} v_\nu L_\nu(y) \,;\, (v_\nu)_{\nu \in \Lambda} \in \ell^2(\Lambda) \right\} \tag{4.40}$$

for $\Lambda \subset \mathfrak{F}$, where L_ν is the tensor product Legendre polynomial from Section 2.2. If Λ is finite and *monotonic* in the sense that if $\mu \in \Lambda$, then Λ also contains all $\nu \in \mathfrak{F}$ with $\nu_m \le \mu_m$ for all $m \in \mathbb{N}$, then

$$\langle \Lambda \rangle = \left\{ v(y) = \sum_{\nu \in \Lambda} v_\nu y^\nu \,;\, (v_\nu)_{\nu \in \Lambda} \in \mathbb{R}^\Lambda \right\}. \tag{4.41}$$

Let $V_k^\Gamma := \langle \Lambda_k \rangle$, $k \in \mathbb{N}_0$, for a nested sequence of finite monotonic index sets

$$\Lambda_0 \subset \Lambda_1 \subset \cdots \subset \Lambda_k \subset \Lambda_{k+1} \subset \cdots \subset \mathfrak{F}. \tag{4.42}$$

We assume that for each $k \in \mathbb{N}_0$, there is a finite set

$$\mathcal{Y}_k = \{ y_i^k \,;\, i = 1, \dots, N_k^\Gamma \} \subset \Gamma$$

and an interpolation operator

$$\mathcal{I}_k : \mathbb{R}^{N_k^\Gamma} \to V_k^\Gamma, \quad (\mathcal{I}_k (a_i)_{i=1}^{N_k^\Gamma})(y_i^k) = a_i \quad \forall i = 1, \dots, N_k^\Gamma, \quad \forall (a_i)_{i=1}^{N_k^\Gamma} \in \mathbb{R}^{N_k^\Gamma}. \tag{4.43}$$

These interpolation operators extend to maps

$$\mathcal{I}_k : C(\Gamma) \to V_k^\Gamma, \quad (\mathcal{I}_k f)(y_i^k) = f(y_i^k) \quad \forall i = 1, \dots, N_k^\Gamma, \quad \forall f \in C(\Gamma), \tag{4.44}$$

which we assume to be the identity on V_k^Γ, *i.e.*, $\mathcal{I}_k$ is a projection of $C(\Gamma)$ onto V_k^Γ.

Example 4.14. Let $\boldsymbol{p} \in \mathfrak{F}$ and

$$\Lambda_{\boldsymbol{p}} := \{\nu \in \mathfrak{F}\,;\ \nu_m \le p_m\ \forall m \in \mathbb{N}\}. \tag{4.45}$$

For each $m \in \mathbb{N}$, let $(y_i^{(m)})_{i=0}^{p_m} \subset [-1,1]$ be an arbitrary set of nodes. The corresponding Lagrange polynomials are

$$\ell_i^{(m)}(z) = \frac{\prod_{j\neq i}(z-y_j)}{\prod_{j\neq i}(y_i-y_j)}, \quad z \in [-1,1], \quad i = 1,\dots,p_m+1. \tag{4.46}$$

In particular, $\ell_1^{(m)} = 1$ if $p_m = 0$. Nodes on $\Gamma = [-1,1]^\infty$ are given by

$$\mathcal{Y}_{\boldsymbol{p}} := \big\{(y_{i_m}^{(m)})_{m=1}^\infty\,;\ 0 \le i_m \le p_m\ \forall m \in \mathbb{N}\big\} \subset \Gamma. \tag{4.47}$$

The tensor product Lagrange polynomials are

$$\ell_y = \bigotimes_{m=1}^\infty \ell_{i_m}^{(m)} \in \langle \Lambda_{\boldsymbol{p}}\rangle, \quad y = (y_{i_m}^{(m)})_{m=1}^\infty \in \mathcal{Y}_{\boldsymbol{p}}. \tag{4.48}$$

By construction, they satisfy

$$\ell_y(z) = \delta_{yz} \quad \forall y,z \in \mathcal{Y}_{\boldsymbol{p}}. \tag{4.49}$$

Consequently, the Lagrange polynomials are linearly independent, and since

$$\#\mathcal{Y}_{\boldsymbol{p}} = \#\Lambda_{\boldsymbol{p}} = \prod_{m=1}^\infty (p_m+1) = \prod_{m\in\operatorname{supp}\boldsymbol{p}} (p_m+1), \tag{4.50}$$

$(\ell_y)_{y\in\mathcal{Y}_{\boldsymbol{p}}}$ spans $\langle\Lambda_{\boldsymbol{p}}\rangle$. Therefore, the interpolation operator (4.44) has the form

$$(\mathcal{I}_{\boldsymbol{p}} f)(z) = \sum_{y\in\mathcal{Y}_{\boldsymbol{p}}} f(y)\ell_y(z), \quad f \in C(\Gamma), \tag{4.51}$$

and it is a projection of $C(\Gamma)$ onto $\langle\Lambda_{\boldsymbol{p}}\rangle$.

The product grids from Example 4.14 are prohibitive for high-dimensional parameter domains since by (4.50), $\#\mathcal{Y}_{\boldsymbol{p}}$ grows exponentially in $\#\operatorname{supp}\boldsymbol{p}$.

For each dimension $m \in \mathbb{N}$, we consider a non-decreasing sequence of polynomial degrees $(q_k^{(m)})_{k=0}^\infty \subset \mathbb{N}_0$, $q_{k+1}^{(m)} \ge q_k^{(m)}$. For every $m \in \mathbb{N}$ and $k \in \mathbb{N}_0$, let $(y_{k,j}^{(m)})_{j=0}^{q_k^{(m)}}$ be a set of nodes in $[-1,1]$, and define the interpolation operator

$$(\mathcal{I}_k^{(m)} f)(\xi) = \sum_{j=0}^{q_k^{(m)}} f(y_{k,j}^{(m)})\, \frac{\prod_{i\neq j}(\xi - y_{k,i}^{(m)})}{\prod_{i\neq j}(y_{k,j}^{(m)} - y_{k,i}^{(m)})}, \quad f \in C([-1,1]). \tag{4.52}$$

Furthermore, we define the univariate differences

$$\Delta_0^{(m)} := \mathcal{I}_0^{(m)}, \quad \Delta_k^{(m)} := \mathcal{I}_k^{(m)} - \mathcal{I}_{k-1}^{(m)}, \quad k \in \mathbb{N}. \tag{4.53}$$

Then the interpolation operator $\mathcal{I}_{\boldsymbol{p}}$ from (4.51) in Example 4.14 with $p_m = q_{k_m}^{(m)}$ can be expanded as

$$\mathcal{I}_{\boldsymbol{p}} = \bigotimes_{m=1}^{\infty} \mathcal{I}_{k_m}^{(m)} = \sum_{0 \le n_m \le k_m} \bigotimes_{m=1}^{\infty} \Delta_{n_m}^{(m)}. \tag{4.54}$$

We approximate $\mathcal{I}_{\boldsymbol{p}}$ by truncating the sum in the last term of (4.54).

We illustrate this construction for $p_m = q_k^{(m)}$ if $m \le M$ and $p_m = 0$ if $m \ge M+1$. The parameter M truncates the dimensions of the parameter domain Γ, and k determines the order of interpolation in the dimensions $m \le M$. Note that $q_k^{(m)}$ may still depend on m, even though k is fixed. For this $\boldsymbol{p}$, (4.54) can be written as

$$\mathcal{I}_{M,k} := \mathcal{I}_{\boldsymbol{p}} = \sum_{0 \le n_1, \dots, n_m \le k} \Delta_{n_1}^{(1)} \otimes \cdots \otimes \Delta_{n_M}^{(M)} \otimes \mathcal{I}_0^{(M+1)} \otimes \mathcal{I}_0^{(M+2)} \otimes \cdots. \tag{4.55}$$

The corresponding Smolyak interpolation operator is

$$\widehat{\mathcal{I}}_{M,k} := \sum_{0 \le n_1 + \cdots + n_M \le k} \Delta_{n_1}^{(1)} \otimes \cdots \otimes \Delta_{n_M}^{(M)} \otimes \mathcal{I}_0^{(M+1)} \otimes \mathcal{I}_0^{(M+2)} \otimes \cdots. \tag{4.56}$$

Inserting (4.53), we arrive at the representation

$$\widehat{\mathcal{I}}_{M,k} = \sum_{0 \le |n| \le k} (-1)^{k-|n|} \binom{M-1}{k-|n|} \mathcal{I}_{n_1}^{(1)} \otimes \cdots \otimes \mathcal{I}_{n_M}^{(M)} \otimes \mathcal{I}_0^{(M+1)} \otimes \mathcal{I}_0^{(M+2)} \otimes \cdots, \tag{4.57}$$

where $|n| = n_1 + \cdots + n_M$.

Remark 4.15. By definition, the Smolyak interpolation operator maps into $\langle \widehat{\Lambda}_{M,k} \rangle$ for

$$\widehat{\Lambda}_{M,k} := \bigcup_{0 \le |n| \le k} \{\nu \in \mathfrak{F}\,;\ \nu_m \le q_{n_m}^{(m)},\ 1 \le m \le M,\ \nu_m = 0,\ m \ge M\}. \tag{4.58}$$

Let $\nu \in \widehat{\Lambda}_{M,k}$. For any $n = (n_1, \dots, n_M)$ with $|n| \ge k+1$, there is an $m \le M$ for which $\nu_m \le q_{n_m - 1}^{(m)}$, and consequently $\big(\Delta_{n_1}^{(1)} \otimes \cdots \otimes \Delta_{n_M}^{(M)}\big) y^{\nu} = 0$ for all $y \in [-1,1]^M$ since $\mathcal{I}_{n_m}^{(m)} y_m^{\nu_m} = \mathcal{I}_{n_m - 1}^{(m)} y_m^{\nu_m} = y_m^{\nu_m}$. As the difference of $\mathcal{I}_{M,k}$ and $\widehat{\mathcal{I}}_{M,k}$ is merely a sum of such operators, $\widehat{\mathcal{I}}_{M,k}$ coincides with $\mathcal{I}_{M,k}$ on $\langle \widehat{\Lambda}_{M,k} \rangle$, and the latter acts as the identity on this space. Therefore, the Smolyak interpolation operator $\widehat{\mathcal{I}}_{M,k}$ is a projection of $C(\Gamma)$ onto $\langle \widehat{\Lambda}_{M,k} \rangle$,

though it is generally not actually an interpolation operator in the sense of (4.51).

As in (2.26), we consider product distributions

$$\mu := \bigotimes_{m=1}^{\infty} \mu_m \tag{4.59}$$

on $\Gamma = [-1,1]^\infty$, where each μ_m is a probability measure on $[-1,1]$.

For every $m \in \mathbb{N}$ and any $q \in \mathbb{N}_0$, the $(q+1)$-point Gaussian quadrature rule for the distribution μ_m consists of abscissae $(\xi_j)_{j=0}^q$ in $[-1,1]$ and weights $(w_j)_{j=0}^q$ such that

$$\int_{-1}^{1} f(\xi)\mu_m(\mathrm{d}\xi) \approx \sum_{j=0}^{q} w_j f(\xi_j), \tag{4.60}$$

and (4.60) is exact if f is a polynomial of degree at most $2q+1$. The points $(\xi_j)_{j=0}^q$ are the roots of the orthonormal polynomial of degree $q+1$ with respect to the measure μ_m.

For a parameter γ to be specified below, we define $(y_{k,j}^{(m)})_{j=0}^{\lceil \gamma k \rceil}$ as the abscissae of the $(\lceil \gamma k \rceil + 1)$-point Gaussian quadrature rule for the measure μ_m. In this case, we have $q_k^{(m)} = \lceil \gamma k \rceil$ in the construction of Smolyak interpolation operators $\widehat{\mathcal{I}}_{k,M}$ from (4.56). By Bieri (2009*b*, Lemma 6.2.2), the operator $\widehat{\mathcal{I}}_{k,M}$ based on these nodes uses

$$N_k^\Gamma = \binom{\lceil \gamma k \rceil + 2M}{2M} \tag{4.61}$$

collocation points in Γ.

Sparse tensorization

Let $V_\ell^D \subset V = H_0^1(D)$ denote the finite element spaces from Section 4.1, and let P_ℓ^D be the orthogonal projection in $H_0^1(D)$ onto V_ℓ^D. We recall that $N_\ell^D := \dim V_\ell^D \lesssim 2^{\ell d}$, and these spaces satisfy the approximation property (4.1).

For parameters M and L, the stochastic collocation solution for the Smolyak interpolation operator $\widehat{\mathcal{I}}_{M,L}$ and the finite element space V_L^D is

$$u_{M,L} := (\widehat{\mathcal{I}}_{M,L} \otimes P_L^D)u. \tag{4.62}$$

It can be expanded as

$$u_{M,L} = \sum_{0 \le k,\ell \le L} (\widehat{\mathcal{I}}_{M,k} - \widehat{\mathcal{I}}_{M,k-1}) \otimes (P_\ell^D - P_{\ell-1}^D)u. \tag{4.63}$$

This approximation can be computed by solving a linear system in V_L^D for

each collocation point of $\widehat{\mathcal{I}}_{M,L}$. Therefore, a measure for the computational cost of computing $u_{M,L}$ is the product $N_L := N_L^\Gamma N_L^D$ of the number of collocation points and the dimension of the finite element space.

The sparse tensor approximation can be derived by truncating the sum in (4.63), analogously to the sparse tensor product construction in Section 4.1, as

$$\widehat{u}_{M,L} := \sum_{0\le k+\ell\le L} (\widehat{\mathcal{I}}_{M,k} - \widehat{\mathcal{I}}_{M,k-1}) \otimes (P_\ell^D - P_{\ell-1}^D)u. \tag{4.64}$$

A (rough) measure for the computational cost of obtaining $\widehat{u}_{M,L}$ is the total number of degrees of freedom, *i.e.*,

$$\widehat{N}_L := \sum_{0\le k+\ell\le L} N_k^\Gamma N_\ell^D, \tag{4.65}$$

which is significantly smaller than N_L due to the geometric growth of N_k^Γ and N_ℓ^D. We turn next to the error analysis of this sparse collocation approximation. We assume that $y_0^{(m)} = 0$ for all $m \ge M+1$. This is satisfied by Gaussian abscissae if the distributions μ_m are symmetric. All collocation points in (4.62) and (4.65) are of the form $y = (y', y_0^{(M+1)}, y_0^{(M+2)}, \ldots)$ with $y' = (y_m)_m \in [-1,1]^M$, and thus the diffusion coefficient

$$\begin{aligned} a((y', y_0^{(M+1)}, y_0^{(M+2)}, \ldots), x) &= \bar{a}(x) + \sum_{m=1}^{M} y_m a_m(x) + \sum_{m=M+1}^{\infty} y_0^{(m)} a_m(x) \\ &= \bar{a}(x) + \sum_{m=1}^{M} y_m a_m(x), \end{aligned} \tag{4.66}$$

$a_m(x) = \alpha_m \varphi_m(x)$, depends only on $y' \in [-1,1]^M$.

Now let $u_M \in C(\Gamma; V)$ denote the solution of

$$\int_D a_M(y,x) \nabla u_M(y,x) \cdot \nabla v(x)\,\mathrm{d}x = \int_D f(x) v(x)\,\mathrm{d}x \quad \forall v \in V, \quad \forall y \in \Gamma, \tag{4.67}$$

for the truncated series

$$a_M(y,x) = \bar{a}(x) + \sum_{m=1}^{M} y_m a_m(x), \quad a_m(x) = \alpha_m \varphi_m(x). \tag{4.68}$$

Since the collocation approximations (4.62) and (4.64) depend only on the first M dimensions of Γ, we can only expect convergence to u_M, not to the solution u of (4.67) with the exact diffusion coefficient $a(y,x)$. The following statement is Proposition 6.3.1 from Bieri (2009*b*).

Theorem 4.16. For any $0 \le \tau \le 1$, if $\gamma = \frac{\tau \log 2}{r_{\min}-1}$, then

$$\|u_M - \widehat{u}_{M,L}\|_{L^2(\Gamma,\mu;V)} \le C\|u_M\|_{C(\Sigma(\Gamma,\varrho);H^{1+\tau}(D))}\widehat{N}_L^{-\min\left(\frac{r_{\min}-1}{1+\log 2M}, \frac{\tau\log 2}{d}\right)}, \tag{4.69}$$

where $r_{\min} := \min\{r_m\,;\ m = 1,\dots,M\}$,

$$r_m := \log\bigl(\varrho_m + \sqrt{1+\varrho_m^2}\bigr), \quad 1 < \varrho_m < \frac{a_-}{2\|a_m\|_{L^\infty(D)}},$$

and

$$\Sigma(\Gamma,\varrho) := \prod_{m=1}^{M} \{c \in \mathbb{C}\,;\ \operatorname{dist}(z,[-1,1]) \le \varrho_m\} \subset \mathbb{C}^M.$$

Remark 4.17. For comparison, by Bieri (2009*b*, Remark 6.3.2),

$$\|u_M - u_{M,L}\|_{L^2(\Gamma,\mu;V)} \le C\|u_M\|_{C(\Sigma(\Gamma,\varrho);H^{1+\tau}(D))}N_L^{-\left(\frac{d}{\tau}+\frac{1+\log 2M}{r_{\min}-1}\right)^{-1}}. \tag{4.70}$$

Therefore, the sparse tensor approximation $\widehat{u}_{M,L}$ converges to u_M significantly faster than the full tensor approximation $u_{M,L}$ as $L \to \infty$.

4.4. The multi-level Monte Carlo finite element method

We regard the multi-level Monte Carlo finite element method as a sparse tensorization of a Monte Carlo method and a standard finite element method. It is the third example of a class of sparse tensor product discretizations after stochastic Galerkin in Section 4.1 and stochastic collocation in Section 4.3.

Preliminaries

For a bounded Lipschitz domain $D \subset \mathbb{R}^d$ and a probability space $(\Omega,\Sigma,\mathbb{P})$, we consider the stochastic isotropic diffusion equation

$$\begin{aligned} -\nabla\cdot\bigl(a(\omega,x)\nabla u(\omega,x)\bigr) &= f(\omega,x), \quad x \in D, \quad \omega \in \Omega, \\ u(\omega,x) &= 0, \quad x \in \partial D, \quad \omega \in \Omega \end{aligned} \tag{4.71}$$

as in Section 2.3. The differential operators in (4.71) are meant with respect to the physical variable $x \in D$. Here, f is a stochastic source term, and a is a stochastic diffusion coefficient. We assume there are constants a_- and a_+ such that

$$0 < a_- \le a(\omega,x) \le a_+ < \infty \quad \forall x \in D, \ \ \forall \omega \in \Omega, \tag{4.72}$$

and a is a strongly measurable map from Ω into $L^\infty(D)$. We assume homogeneous Dirichlet boundary conditions in (4.71) only for simplicity. All of the following can be generalized, for instance to inhomogeneous boundary conditions, or to mixed Dirichlet and Neumann boundary conditions (Barth, Schwab and Zollinger 2010).

For any fixed $\omega \in \Omega$, the weak formulation in space of (4.71) is to find $u(\omega) \in H_0^1(D)$ such that

$$\int_D a(\omega,x)\nabla u(\omega,x)\cdot\nabla v(x)\,\mathrm{d}x = \int_D f(\omega,x)v(x)\,\mathrm{d}x \quad \forall v \in H_0^1(D). \tag{4.73}$$

The solution $u(\omega)$ is well-defined by (4.73) if $f(\omega) \in L^2(D)$. We abbreviate

$$V := H_0^1(D), \quad \|v\|_V := \left(\int_D |\nabla v(x)|^2\,\mathrm{d}x\right)^{1/2}. \tag{4.74}$$

Then the Lax–Milgram lemma implies existence and uniqueness of $u(\omega) \in V$, and

$$\|u(\omega)\|_V \le \Big(\operatorname*{ess\,inf}_{x\in D} a(\omega,x)\Big)^{-1}\|f(\omega)\|_{V'} \le \frac{1}{a_-}\|f(\omega)\|_{V'}. \tag{4.75}$$

Consequently, if $f \in L^r(\Omega;V')$ for any $1 \le r \le \infty$, then $u \in L^r(\Omega;V)$ and

$$\|u\|_{L^r(\Omega,V)} \le \frac{1}{a_-}\|f\|_{L^r(\Omega,V')}. \tag{4.76}$$

We next turn to finite element discretizations of (4.71). Let $\mathcal{T}_0$ be a regular partition of D into simplices K, and let $\{\mathcal{T}_\ell\}_{\ell=0}^\infty$ be the sequence of partitions obtained by uniform mesh refinement. We set

$$V_\ell = S^p(D,\mathcal{T}_\ell) = \{u \in C^0(\bar D)\,;\ u|_K \in \mathcal{P}_p(K) \quad \forall K \in \mathcal{T}_\ell\}, \tag{4.77}$$

where $\mathcal{P}_p(K)$ denote the space of polynomials of degree at most p on K. We denote the mesh width by

$$h_\ell := \max\{\operatorname{diam} K\,;\ K \in \mathcal{T}_\ell\} = 2^{-\ell}h_0. \tag{4.78}$$

The dimension of V_ℓ is

$$N_\ell := \dim V_\ell = \mathcal{O}(h_\ell^{-d}) = \mathcal{O}(2^{\ell d}). \tag{4.79}$$

For any $\omega \in \Omega$ and any $\ell \in \mathbb{N}_0$, let $u_\ell(\omega)$ be the Galerkin projection of $u(\omega)$ onto V_ℓ. By (4.73), $u_\ell(\omega)$ is the unique element of V_ℓ such that

$$\int_D a(\omega,x)\nabla u_\ell(\omega,x)\cdot\nabla v_\ell(x)\,\mathrm{d}x = \int_D f(\omega,x)v_\ell(x)\,\mathrm{d}x \quad \forall v_\ell \in V_\ell. \tag{4.80}$$

The Lax–Milgram lemma implies existence and uniqueness of $u_\ell(\omega)$, and as in (4.75),

$$\|u_\ell(\omega)\|_V \le \frac{1}{a_-}\|f(\omega)\|_{L^2(D)} \quad \forall \omega \in \Omega. \tag{4.81}$$

Furthermore, $u_\ell(\omega)$ is a quasi-optimal approximation of $u(\omega)$ in V_ℓ,

$$\|u(\omega) - u_\ell(\omega)\|_V \le C_a \inf_{v_\ell\in V_\ell}\|u(\omega) - v_\ell\|_V \quad \forall \omega \in \Omega, \tag{4.82}$$

where $C_a = \sqrt{\frac{a_+}{a_-}}$.

We define the scales of Hilbert spaces

$$X_s := V \cap H^{1+s}(D), \quad Y_s := V' \cap H^{-1+s}(D), \quad s \geq 0. \tag{4.83}$$

Then $X_s \supset X_t$ and $Y_s \supset Y_t$ whenever $s < t$. Let $s^* \geq 0$ such that $f(\omega) \in Y_s$ and assume that $a(\omega) \in W^{s,\infty}(D)$ for $\mathbb{P}$-a.e. $\omega \in \Omega$. We also assume

$$\|u(\omega)\|_{X_s} \leq C_s(a)\|f(\omega)\|_{Y_s} \quad \forall s \in [0, s^*] \tag{4.84}$$

for a constant $C_s(a)$ depending continuously on a_-, a_+ and $\|a(\omega)\|_{W^{s,\infty}(D)}$. If $0 < s \leq p$, we have the approximation property

$$\inf_{v_\ell \in V_\ell} \|w - v_\ell\|_V \leq C_I 2^{-s\ell} h_0 \|w\|_{X_s} \quad \forall w \in X_s, \tag{4.85}$$

with a constant C_I independent of ℓ. Consequently, if in addition $s \leq s^*$,

$$\|u(\omega) - u_\ell(\omega)\|_V \leq C_a C_I h_0 2^{-s\ell} \|u(\omega)\|_{X_s}. \tag{4.86}$$

Computation of the mean field

Let $(u^i)_{i=1}^\infty$ be independent copies of the solution u of (4.71). The sample mean with M samples is the V-valued random variable

$$E_M[u] := \frac{1}{M} \sum_{i=1}^{M} u^i. \tag{4.87}$$

As in Theorem 1.12, the sample mean $E_M[u]$ converges to the mean $\mathbb{E}[u]$ in probability, with a rate of $M^{-1/2}$.

Proposition 4.18. If $f \in L^2(\Omega; V')$, then for all $\eta > 0$ and all $M \in \mathbb{N}$,

$$\mathbb{P}\big(\|\mathbb{E}[u] - E_M[u]\|_V \geq \eta\big) \leq \frac{1}{\eta^2 M} \|u\|^2_{L^2(\Omega;V)}. \tag{4.88}$$

Proof. By Chebyshev's inequality, using that $(u^i)_{i=1}^M$ are uncorrelated,

$$\mathbb{P}\big(\|\mathbb{E}[u] - E_M[u]\|_V \geq \eta\big) \leq \frac{1}{\eta^2} \operatorname{Var}(E_M[u]) = \frac{1}{\eta^2 M} \operatorname{Var}(u).$$

The claim follows since $\operatorname{Var}(u) \leq \|u\|^2_{L^2(\Omega;V)}$. □

Equation (4.88) is equivalent to

$$\mathbb{P}\left(\|\mathbb{E}[u] - E_M[u]\|_V \leq \frac{\|u\|_{L^2(\Omega;V)}}{\sqrt{\epsilon M}}\right) \geq 1 - \epsilon \quad \forall \epsilon > 0. \tag{4.89}$$

In this sense, $E_M[u]$ converges to $\mathbb{E}[u]$ in V at rate $M^{-1/2}$, with probability $1 - \epsilon$, which can be chosen arbitrarily close to one.

For all $i \in \mathbb{N}$ and any $\ell \in \mathbb{N}_0$, let u^i_ℓ denote the Galerkin projection of u^i onto V_ℓ. The Monte Carlo finite element (MC–FE) method consists of

approximating $\mathbb{E}[u]$ by

$$E_M[u_\ell] = \frac{1}{M}\sum_{i=1}^{M} u_\ell^i \tag{4.90}$$

for a single discretization level ℓ.

Theorem 4.19. If $f \in L^2(\Omega; Y_s)$ and $a \in L^\infty(\Omega; W^{s,\infty}(D))$ with $0 < s \leq \min(p, s^*)$, then for any $\epsilon > 0$,

$$\mathbb{P}\left[\|\mathbb{E}[u] - E_M[u_\ell]\|_V \leq \left(\frac{1}{\sqrt{\epsilon M}} + C_s(a)C_a C_I h_0 2^{-s\ell}\right)\|f\|_{L^2(\Omega;Y_s)}\right] \geq 1-\epsilon. \tag{4.91}$$

Proof. The statement follows from Proposition 4.18, (4.86) and (4.84) by splitting the error as

$$\|\mathbb{E}[u] - E_M[u_\ell]\|_V \leq \|\mathbb{E}[u] - E_M[u]\|_V + \|E_M[u] - E_M[u_\ell]\|_V$$

and using linearity of E_M. By the assumption $a \in L^\infty(\Omega; W^{s,\infty}(D))$, the constant $C_s(a)$ in (4.84) is independent of ω. □

Remark 4.20. The optimal choice of sample size M versus discretization level ℓ is reached when the statistical and discretization errors are equilibrated. This is the case when $\epsilon^{-1/2}M^{-1/2} = 2^{-s\ell}$, or equivalently, $M = 2^{2s\ell}\epsilon^{-1}$, with some rounding strategy. Since $N_\ell = \mathcal{O}(2^{\ell d})$, if the computational cost of computing a sample of u_ℓ^i is estimated as N_ℓ, then the total cost of reaching a tolerance $C2^{-s\ell}$ with probability $1-\epsilon$ by the MC–FE method is $\mathcal{O}(MN_\ell)$, which is equal to $\mathcal{O}(2^{\ell(2s+d)}\epsilon^{-1})$.

In the multi-level Monte Carlo finite element (MLMC–FE) method, the multi-level splitting of the finite element space V_L is used to obtain a hierarchy of discretizations which are sampled with level-dependent MC sample sizes M_ℓ. Specifically, for any $L \in \mathbb{N}_0$, setting $u_{-1} := 0$, we write, using the linearity of the expectation operator $\mathbb{E}[\cdot]$,

$$\mathbb{E}[u_L] = \mathbb{E}\left[\sum_{\ell=0}^{L}(u_\ell - u_{\ell-1})\right] = \sum_{\ell=0}^{L}\mathbb{E}[u_\ell - u_{\ell-1}]. \tag{4.92}$$

Each of the remaining expectations can be approximated by a different number of samples. This leads to the approximation

$$E^L[u] := \sum_{\ell=0}^{L} E_{M_\ell}[u_\ell - u_{\ell-1}] \tag{4.93}$$

of $\mathbb{E}[u]$, with E_{M_ℓ} defined as in (4.87).

Theorem 4.21. If $f \in L^2(\Omega; Y_s)$ and $a \in L^\infty(\Omega; W^{s,\infty}(D))$ with $0 < s \le \min(p, s^*)$, then there is a constant C depending only on a, D and s such that, for any $\epsilon > 0$ and any $L \in \mathbb{N}_0$,

$$\mathbb{P}\left[\|\mathbb{E}[u] - E^L[u]\|_V \le C\left(h_L^s + \sqrt{\frac{L+1}{\epsilon}}\sum_{\ell=0}^{L} M_\ell^{-1/2} h_\ell^s\right)\|f\|_{L^2(\Omega;Y_s)}\right] \ge 1 - \epsilon.$$

Proof. Adding and subtracting $\mathbb{E}[u_L]$, we have

$$\|\mathbb{E}[u] - E^L[u]\|_V \le \|\mathbb{E}[u] - \mathbb{E}[u_L]\|_V + \sum_{\ell=0}^{L}\|\mathbb{E}[u_\ell - u_{\ell-1}] - E^L[u_\ell - u_{\ell-1}]\|_V.$$

Using (4.86), the first term is bounded by

$$\|\mathbb{E}[u] - \mathbb{E}[u_L]\|_V \le \mathbb{E}[\|u - u_L\|_V] \le C_a C_I h_L^s \mathbb{E}[\|u\|_{X_s}] \le C_a C_I h_L^s \|u\|_{L^2(\Omega;X_s)}.$$

We apply (4.89) with $\epsilon_\ell > 0$ to each of the remaining terms. Thus, with probability $1 - \epsilon_0 - \cdots - \epsilon_L$,

$$\sum_{\ell=0}^{L}\|\mathbb{E}[u_\ell - u_{\ell-1}] - E^L[u_\ell - u_{\ell-1}]\|_V \le \sum_{\ell=0}^{L}\frac{1}{\sqrt{\epsilon_\ell M_\ell}}\|u_\ell - u_{\ell-1}\|_{L^2(\Omega;V)}.$$

Furthermore, due to (4.86),

$$\|u_\ell - u_{\ell-1}\|_{L^2(\Omega;V)} \le 3C_a C_I h_\ell^s \|u\|_{L^2(\Omega;X_s)}.$$

The claim follows with the choice $\epsilon_\ell := (L+1)^{-1}\epsilon$, using (4.84). □

Remark 4.22. Theorem 4.21 holds for arbitrary choices of M_ℓ in (4.93). For any $\epsilon > 0$, we consider

$$M_\ell := \lceil 2^{2s(L-\ell)}(L+1)^3\epsilon^{-1}\rceil, \quad \ell = 0, 1, \ldots, L. \tag{4.94}$$

Then Theorem 4.21 states that, with probability $1 - \epsilon$,

$$\|\mathbb{E}[u] - E^L[u]\|_V \le 2Ch_L^s\|f\|_{L^2(\Omega;Y_s)}. \tag{4.95}$$

Assuming the availability of an optimal finite element solver such as a full multigrid method or, for $d = 1$, a direct solver, the total work required to compute $E^L[u]$ is

$$\sum_{\ell=0}^{L} M_\ell N_\ell \sim 2^{2sL}(L+1)^3\epsilon^{-1}\sum_{\ell=0}^{L} 2^{\ell(d-2s)}. \tag{4.96}$$

In terms of $N_L = \mathcal{O}(2^{Ld})$, the computational cost is

$$\sum_{\ell=0}^{L} M_\ell N_\ell \lesssim \begin{cases} N_L(\log N_L)^3\epsilon^{-1} & \text{if } 2s < d, \\ N_L(\log N_L)^4\epsilon^{-1} & \text{if } 2s = d, \\ N_L^{2s/d}(\log N_L)^3\epsilon^{-1} & \text{if } 2s > d. \end{cases} \tag{4.97}$$

Remark 4.23. If $2s \leq d$, the cost (4.97) of the MLMC–FE method is equivalent to that of the finite element method for solving a single deterministic problem to the same tolerance, up to a logarithmic factor. For example, if $d = 2$, the MLMC–FE method exhibits this optimal behaviour for $p = 1$ and $s = 1$. If $d = 3$, linear complexity is retained up to order $s = 3/2$. In the case $d = 1$, this is only true up to $s = 1/2$. Already for $p = 1$ and $s = 1$, the cost of MLMC–FE in $d = 1$ is $\mathcal{O}(N_L^2(\log N_L)^3\epsilon^{-1})$. This still compares favourably to the MC–FE method, which requires $\mathcal{O}(N_L^3\epsilon^{-1})$ work to achieve the same accuracy. Generally, if $2s > d$, the cost of the MLMC–FE method is equivalent to the total number $\sum_\ell M_\ell$ of Monte Carlo samples, independently of the dimension of the finite element spaces. Thus the efficiency of the MLMC–FE method is dominated by the weaker of its two constituent methods.

The above error analysis of the MLMC–FE method is with respect to convergence in probability. Analogous results also hold when the convergence in $L^2(\Omega; V)$ of the MLMC–FE method is analysed, as in Theorem 1.11. We refer to Barth *et al.* (2010) for proofs and numerical experiments.

Approximation of higher moments

We recall some notation from Section 1. For any $k \in \mathbb{N}$, we denote the k-fold Hilbert tensor product of the Hilbert space V by

$$V^{(k)} := \underbrace{V \otimes \cdots \otimes V}_{k \text{ times}}. \tag{4.98}$$

Let $u^{(k)}(\omega)$ denote the k-fold tensor product $u(\omega)\otimes\cdots\otimes u(\omega) \in V^{(k)}$. Then, if $u \in L^k(\Omega; V)$,

$$\begin{aligned} \|u^{(k)}\|_{L^1(\Omega;V^{(k)})} &= \int_\Omega \|u(\omega) \otimes \cdots \otimes u(\omega)\|_{V^{(k)}} \mathbb{P}(\mathrm{d}\omega) \\ &= \int_\Omega \|u(\omega)\|_V^k \mathbb{P}(\mathrm{d}\omega) = \|u\|_{L^k(\Omega;V)}^k. \end{aligned} \tag{4.99}$$

The kth moment of u is

$$\mathcal{M}^k u := \mathbb{E}[u^{(k)}] = \int_\Omega \underbrace{u(\omega) \otimes \cdots \otimes u(\omega)}_{k \text{ times}} \mathbb{P}(\mathrm{d}\omega) \in V^{(k)}. \tag{4.100}$$

We use analogous notation for other Hilbert spaces V.

We assume that $f \in L^{r^*}(\Omega; Y_{s^*})$ for some $r^* \geq 2$ and $s^* > 0$ such that (4.84) holds. Also, we assume $a \in L^\infty(\Omega; W^{s,\infty}(D))$ for all $s \in [0, s^*]$. The following regularity property generalizes Theorem 1.6 and is shown in Barth *et al.* (2010, Theorem 5.3).

Theorem 4.24. Under the above assumptions, for all $2 \le k \le r^*$, all $1 \le r \le r^*/k$ and every $0 \le s \le s^*$,

$$\|u^{(k)}\|_{L^r(\Omega;X_s^{(k)})} \le C\|f^{(k)}\|_{L^r(\Omega;Y_s^{(k)})} \le C\|f\|^k_{L^{rk}(\Omega;Y_s)}. \tag{4.101}$$

We recall the sparse tensor product construction from Section 1.2. Let $(\psi_j^\ell)_{(\ell,j)\in\nabla}$ be a wavelet basis of V such that V_L is the span of all ψ_j^ℓ with $\ell \le L$. Then the operator P_L on V defined as the restriction to the coordinates with $\ell \le L$ is a projection onto V_L.

Let $Q_\ell := P_\ell - P_{\ell-1}$ for $\ell \in \mathbb{N}_0$, with $P_{-1} := 0$, and let W_ℓ denote the range of Q_ℓ. Then W_ℓ is the span of ψ_j^ℓ for all indices j. It complements $V_{\ell-1}$ to V_ℓ in that $V_\ell = V_{\ell-1} \oplus W_\ell$, with a direct, but generally not orthogonal sum.

We recall from (1.36) that the k-fold sparse tensor product with level $L \in \mathbb{N}_0$ is given by

$$\widehat{V}_L^{(k)} := \sum_{0\le\ell_1+\cdots+\ell_k\le L} V_{\ell_1}\otimes\cdots\otimes V_{\ell_k} = \bigoplus_{0\le\ell_1+\cdots+\ell_k\le L} W_{\ell_1}\otimes\cdots\otimes W_{\ell_k} \subset V^{(k)}. \tag{4.102}$$

The dimension of $\widehat{V}_L^{(k)}$ is only $\mathcal{O}(N_L(\log N_L)^{k-1})$, compared to N_L^k for the full tensor product $V_L^{(k)}$.

If a hierarchical basis of the subspaces $\{V_\ell\}_{\ell=0}^\infty \subset V$ satisfying (W1)–(W5) is explicitly given, the mappings P_ℓ can be realized by truncating the corresponding expansions: see (1.38) and (1.39). These truncations provide numerically computable, stable and quasi-optimally accurate projections onto $\widehat{V}_L^{(k)}$. They are defined by

$$\widehat{P}_L^{(k)} := \sum_{0\le\ell_1+\cdots+\ell_k\le L} Q_{\ell_1}\otimes\cdots\otimes Q_{\ell_k}. \tag{4.103}$$

This is simply the restriction of an element of $V^{(k)}$, expanded in the tensor product hierarchical basis $(\psi^{\ell_1\cdots\ell_k}_{j_1\cdots j_k})_{(\ell_i,j_i)\in\nabla}$ (assumed to be V-orthogonal between different levels), to the indices with $\ell_1+\cdots+\ell_k \le L$.

By Lemma 1.9, the projection $\widehat{P}_L^{(k)}$ is stable on $V^{(k)}$ in the sense that (1.42) holds. Furthermore, the quasi-interpolant $\widehat{P}_L^{(k)}$ is quasi-optimal (provided that the basis (ψ_j^ℓ) is V-orthogonal between different levels ℓ). Using Remark 1.8, for all $0 \le s \le \min(p, s^*)$, there is a constant $C(k,s) > 0$ such that, for all $L \in \mathbb{N}_0$ and every $U \in X_s^{(k)}$,

$$\|U - \widehat{P}_L^{(k)}U\|_{V^{(k)}} \le C(k,s)N_L^{-s/d}(\log N_L)^{(k-1)/2}\|U\|_{X_s^{(k)}}. \tag{4.104}$$

By Proposition 1.7, the approximation rate in (4.104) is optimal. In terms

of the projection $\widehat{P}_L^{(k)}$, the sparse tensor multi-level Monte Carlo approximation of the kth moment $\mathcal{M}^k u = \mathbb{E}[u^{(k)}]$ reads

$$\widehat{E}^L[u^{(k)}] := \sum_{\ell=0}^{L} E_{M_\ell}\big[\widehat{P}_\ell^{(k)} u_\ell^{(k)} - \widehat{P}_{\ell-1}^{(k)} u_{\ell-1}^{(k)}\big], \tag{4.105}$$

where E_{M_ℓ} is defined as in (4.87), and $u_{-1} := 0$. In the case $k = 1$, (4.105) reduces to (4.93). The proof of the following statement is analogous to the proof of Theorem 4.21.

Theorem 4.25. If $f \in L^{2k}(\Omega; Y_s)$ with $0 < s \le \min(p, s^*)$, then there exists a constant C depending only on a, D, s and k such that, for any $\epsilon > 0$ and any $L \in \mathbb{N}_0$, with probability $1 - \epsilon$,

$$\begin{aligned} &\|\mathcal{M}^k u - \widehat{E}^L[u^{(k)}]\|_{V^{(k)}} \\ &\le C\left(h_L^s |\log h_L|^{(k-1)/2} + \sqrt{\frac{L+1}{\epsilon}} \sum_{\ell=0}^{L} M_\ell^{-1/2} h_\ell^s |\log h_\ell|^{(k-1)/2}\right) \|f\|_{L^{2k}(\Omega;Y_s)}^k. \end{aligned} \tag{4.106}$$

Remark 4.26. We consider the same choice of M_ℓ as in Remark 4.22,

$$M_\ell := \big\lceil 2^{2s(L-\ell)} (L+1)^3 \epsilon^{-1} \big\rceil, \quad \ell = 0, 1, \ldots, L. \tag{4.107}$$

Then the error bound (4.106) in Theorem 4.25 becomes

$$\|\mathcal{M}^k u - \widehat{E}^L[u^{(k)}]\|_{V^{(k)}} \le 2C h_L^s |\log h_L|^{(k-1)/2} \|f\|_{L^{2k}(\Omega;Y_s)}^k \tag{4.108}$$

with probability $1-\epsilon$. Assuming the availability of an optimal finite element solver, such as a full multigrid method or, for $d = 1$, a direct solver, since the dimension of $\widehat{V}_\ell^{(k)}$ is on the order of $N_\ell (\log N_\ell)^{k-1}$, the total work required to compute $\widehat{E}^L[u^{(k)}]$ is

$$\sum_{\ell=0}^{L} M_\ell N_\ell (\log N_\ell)^{k-1} \lesssim 2^{2sL} (L+1)^{k+2} \epsilon^{-1} \sum_{\ell=0}^{L} 2^{\ell(d-2s)}. \tag{4.109}$$

In terms of $N_L = \mathcal{O}(2^{Ld})$, the computational cost is

$$\sum_{\ell=0}^{L} M_\ell N_\ell (\log N_\ell)^{k-1} \lesssim \begin{cases} N_L (\log N_L)^{k+2} \epsilon^{-1} & \text{if } 2s < d, \\ N_L (\log N_L)^{k+3} \epsilon^{-1} & \text{if } 2s = d, \\ N_L^{2s/d} (\log N_L)^{k+2} \epsilon^{-1} & \text{if } 2s > d. \end{cases} \tag{4.110}$$

We refer to Barth *et al.* (2010) for a proof and further details. They also provide an error analysis of MLMC methods in the $L^2(\Omega, V^{(k)})$-norm.

APPENDIX

A. Review of probability

A.1. Basic notation

Definition A.1. (probability space) A triple $(\Omega, \Sigma, \mathbb{P})$ is called a *probability space* if Ω is a set, Σ is a σ-algebra on Ω, and $\mathbb{P}$ is a positive measure on Σ with $\mathbb{P}(\Omega) = 1$. Measurable sets $E \in \Sigma$ are *events*; $\mathbb{P}(E)$ is the *probability* of the event E. $\mathbb{P}$ is called the *probability measure*.

Remark A.2.

(i) We shall always have $\{\omega\} \in \Sigma$ for all $\omega \in \Omega$. This is *not* implied by Definition A.1.

(ii) For all $E \in \Sigma$, $\mathbb{P}(E) \in [0, 1]$.

(iii) Instead of '$\mathbb{P}$-a.e.', we write '$\mathbb{P}$-a.s.', which stands for '$\mathbb{P}$ almost surely'.

Consider a finite number n of experiments $\mathcal{E}_i$ with random outcomes E_i, $i = 1, \ldots, n$. To describe them, we assume we are given probability spaces $(\Omega_i, \Sigma_i, \mathbb{P}_i)$, $i = 1, \ldots, n$. Imagine next a new experiment $\mathcal{E}$ consisting of 'mutually independent parallel' experiments $\mathcal{E}_1, \ldots, \mathcal{E}_n$. What is a suitable probability space $(\Omega, \Sigma, \mathbb{P})$ for the mathematical description of $\mathcal{E}$?

Any realization is of the form $\mathcal{E} \ni (\omega_1, \ldots, \omega_n) \in \Omega_1 \times \cdots \times \Omega_n$. If $A_i \in \Sigma_i$, $i = 1, \ldots, n$ is the outcome of $\mathcal{E}_i$, then we consider the outcomes $A_1, \ldots, A_n$ of $\mathcal{E}_1, \ldots, \mathcal{E}_n$ (in this order). Hence, $A := A_1 \times \cdots \times A_n \subset \Omega_1 \times \cdots \times \Omega_n$ is the outcome of $\mathcal{E}$. 'Independence' suggests that we *choose* $\mathbb{P}(A) := \mathbb{P}_1(A_1) \cdots \mathbb{P}_n(A_n)$. The set of all events $\{A_1 \times \cdots \times A_n \,;\, A_i \in \Sigma_i\}$ is a generator of the product sigma algebra $\Sigma := \bigotimes_{i=1}^n \Sigma_i$. On Σ, the product measure $\mathbb{P} = \mathbb{P}_1 \otimes \cdots \otimes \mathbb{P}_n = \bigotimes_{i=1}^n \mathbb{P}_i$ is the only measure satisfying the consistency condition

$$\mathbb{P}(A_1 \times \cdots \times A_n) = \prod_{i=1}^n \mathbb{P}_i(A_i) \quad \forall A_i \in \Sigma_i.$$

Obviously, $\mathbb{P}$ is a probability measure on Σ. Hence

$$(\Omega, \Sigma, \mathbb{P}) = \bigotimes_{i=1}^n (\Omega_i, \Sigma_i, \mathbb{P}_i)$$

is a probability space for $\mathcal{E}$.

A.2. Random variables, distributions, moments

Definition A.3. (random variable) Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, and let (Ω', Σ') be any measurable space. A (Ω', Σ')-valued *random variable* is any (Σ, Σ')-measurable map $X : \Omega \to \Omega'$.

Remark A.4.

(i) If the measurable space (Ω', Σ') is clear from the context, we omit it.

(ii) Images of elementary events $\omega \in \Omega$, *i.e.*, $\omega \mapsto X(\omega) \in \Omega$, are referred to as *samples* (of X), *draws* (of X), or *realizations* (of X).

(iii) Images $X(A)$ of 'complex' events $A \in \Sigma$ are called *ensembles*.

Example A.5.

(1) If $(\Omega', \Sigma') = (\mathbb{R}, \mathcal{B}^1, \mathbb{R}^1)$, we call X a random number resp. a random variable (RV).

(2) If $(\Omega', \Sigma') = (\mathbb{R}^n, \mathcal{B}^n)$, $n > 1$, we call X a random vector. We always assume for $\Omega' = \mathbb{R}^d$ that $\Sigma' = \mathcal{B}^n$.

(3) If $I = [0, T]$ is an interval in $\mathbb{R}$ and $\Omega' = C^0(I)$, we call $\Omega \ni \omega \mapsto X_t(\omega) \in C^0(I)$ a *stochastic process*; for given fixed $\omega \in \Omega$, the realization $X_t(\omega) : I \ni t \mapsto X_t(\omega)$ is called a (continuous) *sample path* (of X).

(4) We shall be interested in random variables mapping into a function space Ω' over a domain $D \subset \mathbb{R}^d$, for example the Sobolev space $\Omega' = H_0^1(D)$. Then a 'sample' $\Omega \ni \omega \mapsto u(\omega) \in H_0^1(D)$ is a *random function* or *random field*. The construction of a σ-algebra Σ' on such an Ω' will be explained below.

Notation A.6. Let X be a (Ω', Σ')-valued random variable. For any $A' \in \Sigma'$,

$$\{X \in A'\} := X^{-1}(A') \in \Sigma, \tag{A.1}$$

$$\mathbb{P}\{X \in A'\} := \mathbb{P}(X^{-1}(A')). \tag{A.2}$$

The set $\{X \in A'\}$ is called the 'event that X lies in A'', and $\mathbb{P}\{X \in A'\}$ is the probability of this event. Note that

$$A' \mapsto \mathbb{P}\{X \in A'\}, \quad A' \in \Sigma',$$

is the image measure of $\mathbb{P}$ under X on (Ω', Σ'). Since $\mathbb{P}\{X \in \Omega'\} = \mathbb{P}(\Omega) = 1$, it is a probability measure on (Ω', Σ').

Random variables are measurable maps between measurable spaces.

Proposition A.7. Let (Ω, Σ), (Ω', Σ') be measurable spaces.

(a) A map $T : \Omega \to \Omega'$ is measurable if and only if

$$\forall A' \in \mathcal{E}' : \quad T^{-1}(A') \in \Sigma \tag{A.3}$$

for some generator $\mathcal{E}'$ of Σ'.

(b) If $T_1 : (\Omega_1, \Sigma_1) \to (\Omega_2, \Sigma_2)$ and $T_2 : (\Omega_2, \Sigma_2) \to (\Omega_3, \Sigma_3)$ are measurable, then $T_2 \circ T_1 : (\Omega_1, \Sigma_1) \to (\Omega_3, \Sigma_3)$ is measurable.

(c) If $T : (\Omega, \Sigma) \to (\Omega', \Sigma')$ is measurable. Then, for every measure μ on Σ the map

$$A' \longmapsto \mu(T^{-1}(A')) =: \mu'(A') \tag{A.4}$$

is a measure μ' on (Ω', Σ').

Definition A.8. (image measure) The measure μ' in (A.4) is the *image measure* of μ under T, denoted by $\mu' = T_{\#}(\mu)$, *i.e.*,

$$T_{\#}(\mu)(A') := \mu(T^{-1}(A')) \quad \forall A' \in \Sigma'. \tag{A.5}$$

Note that

$$(T_2 \circ T_1)_{\#}(\mu) = T_{2\#}(T_{1\#}(\mu)). \tag{A.6}$$

Consider now a family $((\Omega_i, \Sigma_i))_{i\in I}$ of measurable spaces and a family $(T_i)_{i\in I}$ of maps $T_i : \Omega \to \Omega_i$ into Ω_i. Then the σ-algebra Σ generated by $\bigcup_{i\in I} T_i^{-1}(\Sigma_i)$ in Ω is the smallest σ-algebra such that each T_i is (Σ_i, Σ)-measurable. We write

$$\sigma(T_i\,;\ i \in I) := \sigma\left(\bigcup_{i\in I} T_i^{-1}(\Sigma_i)\right). \tag{A.7}$$

Definition A.9. (distribution, law) Let X be a (Ω', Σ')-valued random variable on a probability space $(\Omega, \Sigma, \mathbb{P})$. Then

$$\mathbb{P}_X := X_{\#}(\mathbb{P}) = \mathbb{P} \circ X^{-1} \tag{A.8}$$

is called the *distribution* of X (with respect to $\mathbb{P}$) or the *law* of X.

Hence

$$\mathbb{P}_X(A') = \mathbb{P}\{X \in A'\}, \quad A' \in \Sigma'. \tag{A.9}$$

Definition A.10. (expectation, mean field) Let $X \in \mathbb{R}^n$ be a random variable on a probability space $(\Omega, \Sigma, \mathbb{P})$. Then

$$\mathbb{E}(X) = \mathbb{E}_{\mathbb{P}}(X) := \int_\Omega X\, \mathbb{P}(\mathrm{d}\omega) \in \mathbb{R}^n \tag{A.10}$$

is called the *expected value* or *expectation* of X.

If $X \in \Omega'$ is a random field in a separable Banach space Ω', then

$$\mathbb{E}(X) = \mathbb{E}_{\mathbb{P}}(X) = \int_\Omega X\, \mathbb{P}(\mathrm{d}\omega) \in \Omega' \tag{A.10$'$}$$

is called the *mean field* or *ensemble average* of X and is sometimes denoted by $\langle X \rangle$ when $\mathbb{P}$ is clear from the context.

Remark A.11. The Bochner integral in (A.10$'$) is well-defined if $\mathbb{E}(\|X\|) < \infty$, where $\|\cdot\|$ denotes the norm on Ω'.

Remark A.12. Let $(\Omega', \Sigma') = (\mathbb{R}^n, \mathcal{B}^n)$. Then, for any Borel-measurable function f on $\mathbb{R}^n$ which is $\mathbb{P}_X$-integrable, we have

$$\mathbb{E}(f \circ X) = \int_{\mathbb{R}^n} f \, \mathrm{d}\mathbb{P}_X \tag{A.11}$$

or

$$\mathbb{E}_{\mathbb{P}}(f \circ X) = \mathbb{E}_{\mathbb{P}_X}(f). \tag{A.11$'$}$$

In particular, if X is integrable, $f(x) := x$ gives

$$\mathbb{E}(X) = \int_{\mathbb{R}^n} x \mathbb{P}_X(\mathrm{d}x). \tag{A.12}$$

Definition A.13. (covariance) Let X be an integrable $(\mathbb{R}^n, \mathcal{B}^n)$-valued random variable on a probability space $(\Omega, \Sigma, \mathbb{P})$. Then

$$\mathrm{Cov}(X) := \mathbb{E}\big((X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top\big) \in \mathbb{R}^n \otimes \mathbb{R}^n = \mathbb{R}^{n \times n} \tag{A.13}$$

is called the *covariance* of X.

Note that

$$\mathrm{Cov}(X) := \int_{\mathbb{R}^n} (x - \mathbb{E}[X])(x - \mathbb{E}[X])^\top \, \mathrm{d}x \tag{A.14}$$

is finite if and only if X is square-integrable.

Proposition A.14. A real-valued random variable X on a probability space $(\Omega, \Sigma, \mathbb{P})$ is square-integrable if and only if X is integrable and

$$\mathrm{Cov}(X) = \mathrm{Var}(X) < \infty.$$

Then

$$\begin{aligned} \mathrm{Var}(X) &= \mathbb{E}\big((X - \mathbb{E}(X))^2\big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= \int_{\mathbb{R}} x^2 \, \mathbb{P}_X(\mathrm{d}x) - \left(\int_{\mathbb{R}} x \mathbb{P}_X(\mathrm{d}x) \right)^2. \end{aligned} \tag{A.15}$$

A.3. Independence

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and let I be a set of indices.

Definition A.15. (independent events) A family $(A_i)_{i \in I}$ of events in Σ is called *independent* (with respect to $\mathbb{P}$) if, for every non-empty, finite index set $\{i_1, \ldots, i_n\} \subset I$,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_n}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_n}). \tag{A.16}$$

Example A.16. Let $(\Omega_i, \Sigma_i, \mathbb{P}_i)$, $i = 1, \ldots, n$, be probability spaces, and $(\Omega, \Sigma, \mathbb{P}) = \bigotimes_{i=1}^n (\Omega_i, \Sigma_i, \mathbb{P}_i)$. For each i, let $A_i' \in \Sigma_i$. Then the events

$$A_i := \Omega_1 \times \cdots \times \Omega_{i-1} \times A_i' \times \Omega_{i+1} \times \cdots \times \Omega_n, \quad i = 1, \ldots, n,$$

in Ω are independent.

Definition A.17. (independent families of events) Let $(\mathcal{E}_i)_{i\in I}$ be a *family of sets* in Σ. It is called *independent* if (A.16) holds for every non-empty, finite index set $\{i_1,\dots,i_n\}\subset I$ and every possible choice of $A_{i_\nu}\in\mathcal{E}_{i_\nu}$, $\nu=1,\dots,n$.

It is clear from the definition that independence is preserved if each $\mathcal{E}_i$ is *reduced*.

Remark A.18.

(i) Independence is preserved if each $\mathcal{E}_i$ is increased to its Dynkin system $\delta(\mathcal{E}_i)\subset\Sigma$.

(ii) If $(\mathcal{E}_i)_{i\in I}\subset\Sigma$ is any *independent* family of $\cap$-*stable* subsets $\mathcal{E}_i$ of Σ, then the family $(\sigma(\mathcal{E}_i))_{i\in I}$ is independent.

(iii) If $(\mathcal{E}_i)_{i\in I}\subset\Sigma$ is as in (ii), and $(I_j)_{j\in\mathcal{J}}$ is a partition of I into mutually disjoint $I_j\subset I$, then the system

$$\Sigma_j := \sigma\left(\bigcup_{i\in I_j}\mathcal{E}_i\right),\quad j\in\mathcal{J},$$

is independent.

A.4. Independent random variables

Let $(\Omega,\Sigma,\mathbb{P})$ be a probability space. By Remark A.18(ii), the family $(A_i)_{i\in I}$ is independent if and only if the family $(\Sigma_i)_{i\in I}$ of σ-algebras is independent, where $\Sigma_i=\{\emptyset,A_i,A_i^c,\Omega\}$.

Definition A.19. (independent random variables) A family $(X_i)_{i\in I}$ of random variables (with i-dependent ranges) is *independent* if $(\sigma(X_i))_{i\in I}$ is independent.

Theorem A.20. Let $(X_i)_{i=1,\dots,n}$ be (Ω_i,Σ_i)-valued random variables, and let $\mathcal{G}_i$ be a $\cap$-stable generator of Σ_i, $\Omega_i\in\mathcal{G}_i$, $i=1,\dots,n$. Then $(X_i)_{i=1,\dots,n}$ is independent if and only if, for all $Q_i\in\mathcal{G}_i$, $i=1,\dots,n$,

$$\mathbb{P}\left(\bigcap_{i=1}^n X_i^{-1}(Q_i)\right)=\prod_{i=1}^n\mathbb{P}(X_i^{-1}(Q_i)). \tag{A.17}$$

Proof. Put

$$\mathcal{E}_i := \{X_i^{-1}(Q_i)\,;\ Q_i\in\mathcal{G}_i\}.$$

Then $\mathcal{E}_i$ is a generator of $\sigma(X_i)$, and, by $\cap$-stability of $\mathcal{G}_i$, $\mathcal{E}_i$ is $\cap$-stable and $\Omega\in\mathcal{E}_i$. By Remark A.18(ii), we must show that independence of $(\mathcal{E}_i)_{i\in I}$ is equivalent to

$$\forall E_i\in\mathcal{E}_i:\quad \mathbb{P}(E_1\cap\cdots\cap E_n)=\mathbb{P}(E_1)\cap\cdots\cap\mathbb{P}(E_n).$$

This is evident. □

Let X_i be (Ω_i, Σ_i)-valued random variables, $i = 1, \dots, n$, on a *single* probability space $(\Omega, \Sigma, \mathbb{P})$, and define the *product map*

$$Y := X_1 \otimes \cdots \otimes X_n : \Omega \to \Omega_1 \times \cdots \times \Omega_n$$

by

$$Y(\omega) := (X_1(\omega), \dots, X_n(\omega)), \quad \omega \in \Omega. \tag{A.18}$$

Then, for each $A_1 \times \cdots \times A_n$, $A_i \in \Sigma_i$, $i = 1, \dots, n$,

$$Y^{-1}(A_1 \times \cdots \times A_n) = X_1^{-1}(A_1) \cap \cdots \cap X_n^{-1}(A_n). \tag{A.19}$$

Hence Y is a $(\prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \Sigma_i)$-valued random variable on $(\Omega, \Sigma, \mathbb{P})$, and the distributions $\mathbb{P}_{X_i}$, $i = 1, \dots, n$, and $\mathbb{P}_Y$ are well-defined. We call

$$\mathbb{P}_Y = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$$

the *joint distribution* of $X_1, \dots, X_n$. It is a probability measure on the product-measurable space

$$\left(\prod_{i=1}^n \Omega_i, \bigotimes_{i=1}^n \Sigma_i \right) = \prod_{i=1}^n (\Omega_i, \Sigma_i).$$

Theorem A.21. Finitely many random variables X_i, $i = 1, \dots, n$, are *independent* if and only if their joint distribution is the product of their marginal distributions, *i.e.*, if

$$\mathbb{P}_{X_1 \otimes \cdots \otimes X_n} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}. \tag{A.20}$$

Proof. For each $i = 1, \dots, n$, let $A_i \in \Sigma_i$ be an event. By (A.19),

$$\mathbb{P}_Y \left(\prod_{i=1}^n A_i \right) = \mathbb{P}\left[Y^{-1} \left(\prod_{i=1}^n A_i \right) \right] = \mathbb{P}\left(\bigcap_{i=1}^n X_i^{-1}(A_i) \right),$$

and

$$\mathbb{P}_{X_i}(A_i) = \mathbb{P}(X_i^{-1}(A_i)), \quad i = 1, \dots, n.$$

Hence, $\mathbb{P}_Y$ is the measure of the $\mathbb{P}_{X_i}$ if and only if, for any $A_i \in \Sigma_i$,

$$\mathbb{P}_Y(A_1 \times \cdots \times A_n) = \mathbb{P}_{X_1}(A_1) \cdots \mathbb{P}_{X_n}(A_n).$$

This is equivalent to

$$\mathbb{P}\left(\bigcap_{i=1}^n X_i^{-1}(A_i) \right) = \prod_{i=1}^n \mathbb{P}(X_i^{-1}(A_i)) \quad \forall A_i \in \Sigma_i, \quad i = 1, \dots, n.$$

By Theorem A.20, (A.17), $X_1, \dots, X_n$ are independent. □

A.5. Infinite products of probability spaces

We saw that the proper model for the description of $n < \infty$ independent experiments with random outcome is the product probability space $(\Omega, \Sigma, \mathbb{P}) = \bigotimes_{i=1}^{n}(\Omega_i, \Sigma_i, \mathbb{P}_i)$.

We now consider the case where we have an infinite family $\mathcal{E} = (\mathcal{E}_n)_{n=1}^{\infty}$ of 'independent', 'random' experiments. Each experiment is described by a probability space $(\Omega_n, \Sigma_n, \mathbb{P}_n)$, $n = 1, 2, \ldots$. We build a probability space to describe $\mathcal{E}$. It should satisfy the following conditions.

(1) Each elementary event $\omega \in \Omega$ is a *sequence* $(\omega_n)_{n=1}^{\infty}$ of elementary events $\omega_n \in \Omega_n$, *i.e.*,

$$\Omega = \prod_{n=1}^{\infty} \Omega_n = \Omega_1 \times \Omega_2 \times \cdots .$$

(2) If $A_1 \in \Sigma_1, \ldots, A_n \in \Sigma_n$ are possible outcomes at the first n experiments, we view the set

$$A = A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \times \cdots , \quad n = 1, 2, \ldots \tag{A.21}$$

as that outcome of $\mathcal{E}$ which gives $A_1, \ldots, A_n$ in the first n experiments of the (infinite) sequence $(\mathcal{E}_i)_{i=1}^{\infty}$ of experiments. Therefore, we *require* that A defined in (A.21) satisfies $A \in \Sigma$ and that

$$\mathbb{P}(A) = \mathbb{P}_1(A_1) \cdots \mathbb{P}_n(A_n). \tag{A.22}$$

The requirements (A.21) and (A.22) and a certain minimum property define the measure $\mathbb{P}$ uniquely. Given an *index set* $I \neq \emptyset$ and a family $((\Omega_i, \Sigma_i, \mathbb{P}_i))_{i \in I}$ of probability spaces, for each $K \subseteq I$ we define

$$\Omega_K := \prod_{i \in K} \Omega_i, \tag{A.23}$$

and we set

$$\Omega := \Omega_I = \prod_{i \in I} \Omega_i. \tag{A.24}$$

Note that Ω_K is the set of all maps

$$\omega_K : K \to \bigcup_{i \in K} \Omega_i \quad \text{such that} \quad \omega_K(i) \in \Omega_i \quad \forall i \in K.$$

Restricting ω_K to $J \subset K$, we get the *projection map*

$$p_J^K := \Omega_K \to \Omega_J. \tag{A.25}$$

If $K = I$, we write $p_J := p_J^I$; if $j = \{i\}$, $p_i^K := p_{\{i\}}^K$. Then

$$p_J^L = p_J^K \circ p_K^L, \quad J \subset K \subset I. \tag{A.26}$$

By $\mathcal{F} = \mathcal{F}(I)$ we denote the set of all finite subsets of I. For each $J \in \mathcal{F}$, we have

$$\Sigma_J := \bigotimes_{i\in J} \Sigma_i, \quad \mathbb{P}_J := \bigotimes_{i\in J} \mathbb{P}_i. \tag{A.27}$$

Definition A.22. (infinite product of σ-algebras) We call the *product* $\bigotimes_{i\in I} \Sigma_i$ of the family $\{\Sigma_i\,;\ i \in I\}$ of σ-algebras the smallest σ-algebra Σ_0 in Ω for which each projection p_i is (Σ_0, Σ_i)-measurable, *i.e.*,

$$\Sigma_0 = \bigotimes_{i\in I} \Sigma_i := \sigma(p_i\,;\ i \in I). \tag{A.28}$$

For all $J \in \mathcal{F}$, p_J is (Σ_0, Σ_J)-measurable, since, by (A.26), we have $p_i = p_i^J \circ p_J$ for all $i \in J$. This allows us to extend (A.28) to

$$\bigotimes_{i\in I} \Sigma_i = \sigma(p_i\,;\ i \in I) = \sigma(p_J\,;\ J \in \mathcal{F}(I)). \tag{A.28$'$}$$

We now wish to find a probability measure $\mathbb{P}$ on Σ_0 such that

$$\mathbb{P}\left(p_J^{-1}\left(\prod_{i\in J} A_i\right)\right) = \prod_{i\in J} \mathbb{P}_k(A_i) \quad \forall J \in \mathcal{F}, \quad \forall A_i \in \Sigma_i, \quad \forall i \in J.$$

By definition of the image of a measure under a mapping, each $p_J(\mathbb{P})$ of a set $\prod_{i\in J} A_i$ has the value

$$(p_J)_{\#}(\mathbb{P})\left(\prod_{i\in J} A_i\right) = \prod_{i\in J} \mathbb{P}_i(A_i). \tag{A.29}$$

For all $J \in \mathcal{F}$, the finite product measure $\mathbb{P}_J$ in (A.27) is the unique measure such that (A.29) holds. Does there exist a probability measure $\mathbb{P}$ on Σ_0 such that its image under *any* projection p_J, $J \in \mathcal{F}$, equals $\mathbb{P}_J$?

Theorem A.23. There exists a unique measure $\mathbb{P}$ on $\Sigma_0 = \bigotimes_{i\in I} \Sigma_i$ such that

$$(p_J)_{\#}(\mathbb{P}) = \mathbb{P}_J \quad \forall J \in \mathcal{F}(I). \tag{A.30}$$

$\mathbb{P}$ is a probability measure on (Ω, Σ_0).

We refer to Bauer (1996) for the proof.
If $|I| < \infty$, $\mathbb{P} = \mathbb{P}_I$ by (A.30).

Definition A.24. (infinite product measure) The unique probability measure $\mathbb{P}$ on Σ_0 from Theorem A.23 is called the *product of the measures* $(\mathbb{P}_i)_{i\in I}$ and is denoted by $\bigotimes_{i\in I} \mathbb{P}_i$. The probability space

$$(\Omega, \Sigma_0, \mathbb{P}) = \left(\prod_{i\in I} \Omega_i, \bigotimes_{i\in I} \Sigma_i, \bigotimes_{i\in I} \mathbb{P}_i\right)$$

is called the *product of the probability spaces* $((\Omega_i, \Sigma_i, \mathbb{P}_i)_{i\in I})$ and is denoted by

$$(\Omega, \Sigma_0, \mathbb{P}) =: \bigotimes_{i\in I} (\Omega_i, \Sigma_i, \mathbb{P}_i).$$

Now we can extend Theorem A.21 on independence of random variables $X_1, \dots, X_n$.

Theorem A.25. A family $(X_i)_{i\in I}$ of random variables is independent if and only if their joint distribution is the product of the distributions $\mathbb{P}_{X_i}$, *i.e.*,

$$\mathbb{P}_{\bigotimes_{i\in I} X_i} = \bigotimes_{i\in I} \mathbb{P}_{X_i}. \tag{A.31}$$

Proof. For every $\emptyset \neq J \subset I$, $J \in \mathcal{F}$, let

$$p_J : \prod_{i\in I} \Omega_i \to \prod_{j\in J} \Omega_j$$

denote the projection, let Y denote the mapping $\bigotimes_{i\in I} X_i$, and let $Y_J : \Omega \to \prod_{j\in J} \Omega_j$ denote $\bigotimes_{j\in J} X_j$.

Then $Y_J = p_J \circ Y$, whence it follows that

$$\mathbb{P}_{Y_J} = (p_J)_{\#}(\mathbb{P}_Y),$$

by transitivity (A.6) of image measures.

Independence of $(X_i)_{i\in I}$ is equivalent to independence of $(X_j)_{j\in J}$ for all $J \in \mathcal{F}$, *i.e.*, by Theorem A.21,

$$\mathbb{P}_{Y_J} = \bigotimes_{j\in J} \mathbb{P}_{X_j} \quad \forall J \in \mathcal{F}.$$

By Theorem A.23, (A.31) is equivalent to

$$(p_J)_{\#}(\mathbb{P}_Y) = \bigotimes_{j\in J} \mathbb{P}_{X_j} \quad \forall J \in \mathcal{F}.$$

The assertion follows. □

Corollary A.26. For *any* family $((\Omega_i, \Sigma_i, \mathbb{P}_i))_{i\in I}$ of probability spaces, there exists an independent family $(X_i)_{i\in I}$ of (Ω_i, Σ_i)-valued random variables X_i on a suitable probability space $(\Omega, \Sigma, \mathbb{P})$ such that

$$\forall i \in I : \quad \mathbb{P}_i = \mathbb{P}_{X_i} \tag{A.32}$$

is the distribution of X_i.

Proof. We choose

$$(\Omega, \Sigma, \mathbb{P}) = \bigotimes_{i\in I} (\Omega_i, \Sigma_i, \mathbb{P}_i)$$

and

$$X_j = p_j : \prod_{i\in I} \Omega_i \to \Omega_j.$$

Then, by the definition of product measure $\mathbb{P} = \bigotimes_{i\in I} \mathbb{P}_i$, $\mathbb{P}_j$ is the distribution of X_j, for all $j \in I$.

The independence of $(X_i)_{i\in I}$ follows from Theorem A.25, since $\bigotimes_{i\in I} X_i$ is the identity map from $\Omega = \prod_{i\in I} \Omega_i$ onto itself, whence

$$\mathbb{P}_{\bigotimes_{i\in I} X_i} = \mathbb{P} = \bigotimes_{i\in I} \mathbb{P}_i = \bigotimes_{i\in I} \mathbb{P}_{X_i}. \qquad \square$$

B. Review of Hilbert spaces

We review several standard notions and definitions of bases in separable Hilbert spaces to the extent necessary in the present work; among the many references for this material, we mention in particular Christensen (2008, 2010).

B.1. Basic properties

By H, we denote a *real*, *separable* Hilbert space, with norm $\|\cdot\|_H$ and inner product $\langle\cdot,\cdot\rangle$.

Definition B.1. (Schauder basis) A sequence $(e_k)_{k=1}^\infty \subset H$ is called a (*Schauder*) *basis* of H if for any $x \in H$ there exists a unique sequence $(c_k)_{k=1}^\infty$ such that

$$\left\| x - \sum_{k=1}^{n} c_k e_k \right\|_H \to 0, \quad n \to \infty, \tag{B.1}$$

or equivalently,

$$x = \sum_{k=1}^{\infty} c_k e_k \quad \text{in } H. \tag{B.1$'$}$$

Proposition B.2. Every separable Hilbert space over $\mathbb{R}$ admits a basis $(e_k)_{k=1}^\infty$. Given any basis $(e_k)_{k=1}^\infty$ of H, there exists a unique sequence $(g_k)_{k=1}^\infty \subset H$ such that

$$\forall f \in H : \quad f = \sum_{k=1}^{\infty} \langle f, g_k\rangle e_k \quad \text{in } H. \tag{B.2}$$

Then $(g_k)_{k=1}^\infty$ is also a basis of H, called the *dual basis*. The sequences $(e_k)_{k=1}^\infty$ and $(g_k)_{k=1}^\infty$ are *biorthogonal*,

$$\langle e_k, g_j\rangle = \delta_{kj}. \tag{B.3}$$

We also say $(g_k)_{k=1}^\infty$ is the biorthogonal system for $(e_k)_{k=1}^\infty$.

Definition B.3. (Bessel sequence) $(f_k)_{k=1}^\infty \subset H$ is a Bessel sequence if

$$\exists B > 0 \quad \forall f \in H: \quad \sum_{k=1}^\infty |\langle f, f_k\rangle|^2 \le B\|f\|_H^2.$$

B is called a *Bessel bound*.

Lemma B.4. $(f_k)_{k=1}^\infty \subset H$ is a Bessel sequence with Bessel bound B if and only if

$$T: (c_k)_{k=1}^\infty \mapsto \sum_{k=1}^\infty c_k f_k$$

is a well-defined, bounded operator from $\ell^2(\mathbb{N})$ into H and $\|T\| \le \sqrt{B}$.

Proof. Assume $(f_k)_k \subset H$ is Bessel and $(c_k)_k \in \ell^2(\mathbb{N})$. To show that $T(c_k)_k$ is well-defined, we show that $\sum_{k=1}^\infty c_k f_k$ converges in H. Let $m, n \in \mathbb{N}$, $n > m$. Then

$$\begin{aligned}
\left\| \sum_{k=1}^n c_k f_k - \sum_{k=1}^m c_k f_k \right\| &= \left\| \sum_{k=m+1}^n c_k f_k \right\| = \sup_{\|g\|=1} \left| \left\langle \sum_{k=m+1}^n c_k f_k, g \right\rangle \right| \\
&\le \sup_{\|g\|=1} \sum_{k=m+1}^n |c_k \langle f_k, g\rangle| \\
&\le \left(\sum_{k=m+1}^n |c_k|^2 \right)^{1/2} \sup_{\|g\|=1} \left(\sum_{k=m+1}^n |\langle f_k, g\rangle|^2 \right)^{1/2} \\
&\le \sqrt{B} \left(\sum_{k=m+1}^n |c_k|^2 \right)^{1/2}.
\end{aligned}$$

Since $(c_k) \in \ell^2(\mathbb{N})$, $(\sum_1^n |c_k|^2)_{n=1}^\infty$ is Cauchy, and thus $(\sum_1^n c_k f_k)_{n=1}^\infty \subset H$ is convergent. Therefore $T: \ell^2(\mathbb{N}) \to H$ is well-defined and bounded with $\|T\| \le \sqrt{B}$. Obviously, T is also linear. Since

$$\sum_{k=1}^\infty |\langle f, f_k\rangle|^2 = \|T^* f\|_{\ell^2}^2 \le \|T\|^2 \|f\|_H^2 \quad \forall f \in H,$$

the claim follows. □

Lemma B.5. Let $(e_k)_{k=1}^\infty$ be a basis of H and $(g_k)_{k=1}^\infty$ the associated biorthogonal system. If $(e_k)_{k=1}^\infty$ is a Bessel sequence with bound B, then

$$\frac{1}{B}\|f\|_H^2 \le \sum_{k=1}^\infty |\langle f, g_k\rangle|^2 \quad \forall f \in H, \tag{B.4}$$

and for all finitely supported sequences $(c_k)_{k=1}^\infty$,

$$\frac{1}{B}\sum_{k=1}^\infty |c_k|^2 \le \left\|\sum_{k=1}^\infty c_k g_k\right\|_H^2. \tag{B.5}$$

Proof. Let $f \in H$. Since $f = \sum_{k=1}^\infty \langle f, g_k\rangle e_k$,

$$\begin{aligned}\|f\|_H^4 &= \left|\sum_{k=1}^\infty \langle f, g_k\rangle\langle e_k, f\rangle\right|^2 \\ &\le \sum_{k=1}^\infty |\langle f, g_k\rangle|^2 \sum_{k=1}^\infty |\langle e_k, f\rangle|^2 \le B\|f\|_H^2 \sum_{k=1}^\infty |\langle f, g_k\rangle|^2.\end{aligned}$$

This shows (B.4).

Let $(c_k)_{k=1}^\infty$ be a finitely supported sequence. Then (B.5) follows from

$$\sum_{k=1}^\infty |c_k|^2 = \sum_{k=1}^\infty \left|\sum_{j=1}^\infty c_j \underbrace{\langle g_j, e_k\rangle}_{\delta_{jk}}\right|^2 = \sum_{k=1}^\infty \left|\left\langle \sum_{j=1}^\infty c_j g_j, e_k\right\rangle\right|^2 \le B\left\|\sum_{j=1}^\infty c_j g_j\right\|_H^2. \quad \square$$

B.2. Bases of Hilbert spaces

As before, we let H denote a separable Hilbert space.

Definition B.6. (orthonormal basis) A sequence $(e_k)_{k=1}^\infty \subset H$ is an *orthonormal system* in H if

$$\langle e_k, e_j\rangle = \delta_{kj}.$$

It is an *orthonormal basis* of H if, in addition, it is a basis of H, *i.e.*

$$\forall x \in H: \quad \left\|x - \sum_{i=1}^n \langle x, e_i\rangle e_i\right\|_H \to 0, \quad n \to \infty, \tag{B.6}$$

or equivalently,

$$\forall x \in H: \quad x = \sum_{i=1}^\infty \langle x, e_i\rangle e_i \quad \text{in } H. \tag{B.6$'$}$$

Remark B.7. Any orthonormal system $(e_k)_k \subset H$ is a Bessel sequence with Bessel constant $B = 1$.

Proof. Let $(c_k)_k \subset \ell^2(\mathbb{N})$, $m, n \in \mathbb{N}$, $n > m$. Then $(\sum_{k=1}^n c_k e_k)_n$ converges in H since

$$\left\|\sum_{k=n+1}^m c_k e_k\right\|_H^2 = \sum_{n+1}^m |c_k|^2 \to 0, \quad m \to \infty,$$

and we find

$$\left\| \sum_{k=1}^{\infty} c_k e_k \right\|_H^2 = \sum_{k=1}^{\infty} |c_k|^2 = \|c\|_{\ell^2(\mathbb{N})}^2.$$

Therefore, for $f \in H$, $(c_k)_k = (\langle f, e_k \rangle)_k \in \ell^2(\mathbb{N})$ and thus $B = 1$. □

Theorem B.8. For an orthonormal system $(e_k)_k \subset H$, the following conditions are equivalent:

(a) $(e_k)_k$ is an orthonormal basis of H,

(b) for all $f \in H$, $f = \sum_{k=1}^{\infty} \langle f, e_k \rangle e_k$,

(c) for all $f, g \in H$, $\langle f, g \rangle = \sum_{k=1}^{\infty} \langle f, e_k \rangle \langle e_k, g \rangle$,

(d) for all $f \in H$, $\sum_{k=1}^{\infty} |\langle f, e_k \rangle|^2 = \|f\|_H^2$ (Parseval),

(e) $\operatorname{span}(e_k)_{k=1}^{\infty}$ is dense in H,

(f) $\langle f, e_k \rangle = 0$ for all $k \in \mathbb{N}$ implies $f = 0$.

Corollary B.9. If $(e_k)_k$ is an orthonormal basis of H, it coincides with its biorthogonal basis, and

$$\forall f \in H: \quad f = \sum_{k=1}^{\infty} \langle f, e_k \rangle e_k.$$

Theorem B.10. Every separable Hilbert space H has an orthonormal basis $(e_k)_k$.

Proof. Since H is separable, there exists a sequence $(f_k)_{k=1}^{\infty}$ for which $\operatorname{span}(f_k)_{k=1}^{\infty}$ is dense in H. We assume, without loss of generality, that for all $n \in \mathbb{N}$, $f_{n+1} \notin \operatorname{span}(f_k)_{k=1}^{n}$. Applying Gram–Schmidt orthogonalization to $(f_k)_{k=1}^{\infty}$, we obtain an orthonormal system $(e_k)_{k=1}^{\infty} \subset H$ such that $\operatorname{span}(e_k)_{k=1}^{n} = \operatorname{span}(f_k)_{k=1}^{n}$ for all $n \in \mathbb{N}$. By Theorem B.8, $(e_k)_k$ is an orthonormal basis of H. □

Example B.11. If $H = \ell^2(\mathbb{N})$, then $(\delta_k) = (0, 0, \ldots, 0, 1, 0, \ldots)$ is an orthonormal basis.

Theorem B.12. Every separable infinite-dimensional Hilbert space H is isometrically isomorphic to $\ell^2(\mathbb{N})$.

Proof. Let $(e_k)_{k=1}^{\infty}$ be an orthonormal basis of H. Then, for all $(c_k) \in \ell^2(\mathbb{N})$, $\sum_{k=1}^{\infty} c_k e_k$ is convergent. Also,

$$\forall f \in H: \quad f = \sum_{k=1}^{\infty} \langle f, e_k \rangle e_k.$$

If $(\delta_k)_{k=1}^\infty$ is the orthonormal basis of $\ell^2(\mathbb{N})$ from Example B.11, we define

$$U : H \to \ell^2(\mathbb{N}) : \quad U\left(\sum_{k=1}^\infty c_k e_k\right) := \sum_{k=1}^\infty c_k \delta_k, \quad (c_k)_k \in \ell^2(\mathbb{N}).$$

Then $U : H \to \ell^2(\mathbb{N})$ is well-defined and bijective. Also, for all $f \in H$, $f = \sum_k \langle f, e_k\rangle e_k$ and

$$\|Uf\|^2_{\ell^2(\mathbb{N})} = \left\|\sum_{k=1}^\infty \langle f, e_k\rangle \delta_k\right\|^2_{\ell^2(\mathbb{N})} = \sum_{k=1}^\infty |\langle f, e_k\rangle|^2 = \|f\|^2_H. \qquad \square$$

Theorem B.13. Let $(e_k)_k$ be an orthonormal basis of a separable Hilbert space H. Then *all orthonormal bases of* H are of the form $(Ue_k)_{k=1}^\infty$ for a unitary map $U : H \to H$.

Proof. Let $(f_k)_k$ be an orthonormal basis of H. Define

$$U : H \to H, \quad U\left(\sum_k c_k e_k\right) = \sum_k c_k f_k, \quad (c_k)_k \in \ell^2(\mathbb{N}). \tag{B.7}$$

Then $U : H \to H$ is bounded and bijective.

For $f, g \in H$, $f = \sum_k \langle f, e_k\rangle e_k$, $g = \sum_k \langle g, e_k\rangle e_k$. Then using (B.7), Theorem B.8 implies

$$\begin{aligned}\langle U^*Uf, g\rangle &= \langle Uf, Ug\rangle \\ &= \left\langle \sum_k \langle f, e_k\rangle f_k, \sum_\ell \langle g, e_\ell\rangle f_\ell \right\rangle = \sum_k \langle f, e_k\rangle \langle g, e_k\rangle = \langle f, g\rangle.\end{aligned}$$

Therefore $U^*U = I$, and since U is surjective by (B.7), U is unitary.

Conversely, if a unitary map U is given, then

$$\langle Ue_k, Ue_j\rangle = \langle U^*Ue_k, e_j\rangle = \langle e_k, e_j\rangle = \delta_{kj},$$

so $(Ue_k)_k$ is an orthonormal system. Since U is surjective, $(Ue_k)_k$ is an orthonormal basis of H. $\square$

B.3. Tensor products of separable Hilbert spaces

Tensor products are useful in the design of sparse approximation schemes. We consider tensor products of separable Hilbert spaces and refer to Light and Cheney (1985), Ryan (2002), Schatten (1943), Grothendieck (1955) and Kalton (2003) for a general theory of tensor products of Banach spaces. We follow the construction in Reed and Simon (1980) for separable Hilbert spaces.

Let H_1, H_2 be two separable Hilbert spaces. For $\varphi_1 \in H_1$, $\varphi_2 \in H_2$, we denote by $\varphi_1 \otimes \varphi_2$ the conjugate bilinear form on $H_1 \times H_2$ defined by

$$(\varphi_1 \otimes \varphi_2)(\varphi_1, \psi_2) := \langle \psi_1, \varphi_1\rangle_{H_1} \langle \psi_2, \varphi_2\rangle_{H_2} \quad \forall \psi_i \in H_i,\ i = 1, 2.$$

Let $\mathcal{E}$ denote the space of all finite linear combinations of such bilinear forms; on $\mathcal{E}$, we define an inner product by

$$\langle \varphi \otimes \psi, \eta \otimes \mu \rangle := \langle \varphi, \eta \rangle_{H_1} \langle \psi, \mu \rangle_{H_2}, \quad \varphi, \eta \in H_1, \quad \psi, \mu \in H_2. \tag{B.8}$$

Proposition B.14. $\langle \cdot, \cdot \rangle$ from (B.8) is well-defined and positive definite.

Proof. To show that $\langle \cdot, \cdot \rangle$ is well-defined, we check that $\langle \lambda, \lambda' \rangle$ is independent of the linear combination of simple tensors used to represent λ and λ' in $\mathcal{E}$. By linearity and symmetry, it suffices to show that if μ is a finite sum in $\mathcal{E}$ equal to the zero form, then

$$\langle \eta, \mu \rangle = 0 \quad \forall \eta \in \mathcal{E}.$$

Let

$$\eta = \sum_{i=1}^{N} c_i (\varphi_i \otimes \psi_i).$$

Then

$$\langle \eta, \mu \rangle = \left\langle \sum_{i=1}^{N} c_i (\varphi_i \otimes \psi_i), \mu \right\rangle = \sum_{i=1}^{N} c_i \mu(\varphi_i, \psi_i) = 0,$$

since μ is the zero form. Hence $\langle \cdot, \cdot \rangle$ is well-defined.

Next, we show that $\langle \cdot, \cdot \rangle$ is positive definite. Let

$$\lambda = \sum_{k=1}^{M} d_k (\eta_k \otimes \mu_k).$$

Then

$$M_1 := \operatorname{span} \{\eta_k\}_{k=1}^{N} \subset H_1, \quad M_2 := \operatorname{span} \{\mu_k\}_{k=1}^{M} \subset H_2$$

are subspaces. Let $\{\varphi_j\}_{j=1}^{N_1}$, $\{\psi_\ell\}_{\ell=1}^{N_2}$ be orthonormal bases of M_1 and M_2. Then we can write each η_k uniquely in terms of the φ_j, and each μ_k uniquely via the ψ_ℓ to get

$$\lambda = \sum_{j=1}^{N_1} \sum_{\ell=1}^{N_2} c_{j\ell} (\varphi_j \otimes \psi_\ell).$$

We compute

$$\begin{aligned} \langle \lambda, \lambda \rangle &= \left\langle \sum_{j,\ell} c_{j\ell} (\varphi_j \otimes \psi_\ell), \sum_{i,m} c_{im} (\varphi_i \otimes \psi_m) \right\rangle \\ &= \sum_{j,\ell,i,m} c_{j\ell} c_{im} \langle \varphi_j, \varphi_i \rangle_{H_1} \langle \psi_\ell, \psi_m \rangle_{H_2} = \sum_{j,\ell} |c_{j\ell}|^2 \geq 0. \end{aligned}$$

Furthermore, if $\langle \lambda, \lambda \rangle = 0$, then $c_{j\ell} = 0$ for all j, ℓ, hence $\lambda = 0$. Thus $\langle \cdot, \cdot \rangle$ is positive definite. □

Therefore, $\mathcal{E}$ is a pre-Hilbert space with inner product $\langle \cdot, \cdot \rangle$.

Definition B.15. (tensor product of separable Hilbert spaces) The *tensor product* $H_1 \otimes H_2$ *of* H_1 and H_2 is the completion of $\mathcal{E}$ under $\langle \cdot, \cdot \rangle$.

Proposition B.16. If $(\varphi_k)_k$ and $(\psi_\ell)_\ell$ are orthonormal bases of H_1 and H_2, respectively, then the set of all dyadic products $(\varphi_k \otimes \psi_\ell)_{k,\ell}$ is an orthonormal basis of $H_1 \otimes H_2$.

Proof. We assume that H_1, H_2 are infinite-dimensional and separable. Then $(\varphi_k \otimes \psi_\ell)_{k\ell}$ is an orthonormal set in $H_1 \otimes H_2$, and hence we must show $\mathcal{E} \subset \mathcal{S} := \operatorname{span}_{H_1 \otimes H_2}(\varphi_k \otimes \psi_\ell)$. Let $\varphi \otimes \psi \in \mathcal{E}$. Since $(\varphi_k)_k$, $(\psi_\ell)_\ell$ are bases, we have

$$\varphi = \sum_k c_k \varphi_k, \quad \sum_k |c_k|^2 < \infty, \quad \psi = \sum_\ell d_\ell \psi_\ell, \quad \sum_\ell |d_\ell|^2 < \infty.$$

Therefore, there exists a vector

$$\mu = \sum_{k,\ell} c_k d_\ell \varphi_k \otimes \psi_\ell \in \mathcal{S},$$

and

$$\left\| \varphi \otimes \psi - \sum_{k=1}^{N_1} \sum_{\ell=1}^{N_2} c_k d_\ell \varphi_k \otimes \psi_\ell \right\| \to 0, \quad N_1, N_2 \to \infty. \qquad \square$$

Let (M_1, μ_1), (M_2, μ_2) denote two measure spaces. We *assume* that $L^2(M_2, \mu_1)$ and $L^2(M_2, \mu_2)$ are separable. Further, let $(\varphi_k(x))_k$, $(\varphi_\ell(y))_\ell$ denote orthonormal bases of $L^2(M_1, \mu_1)$ and of $L^2(M_2, \mu_2)$, respectively. We show that $(\varphi_k(x)\, \psi_\ell(y))_{k\ell}$ is then an orthonormal basis of $L^2(M_1 \times M_2, \mu_1 \otimes \mu_2)$. To see this, we assume $f(x, y) \in L^2(M_1 \times M_2, \mu_1 \otimes \mu_2)$ and

$$\int_{M_1 \times M_2} f(x, y) \varphi_k(x) \psi_\ell(y)\, (\mu_1 \otimes \mu_2)\, \mathrm{d}(x, y) = 0 \quad \forall k, \ell.$$

By Fubini's theorem,

$$\int_{M_2} \left(\int_{M_1} f(x, y) \varphi_k(x)\, \mu_1(\mathrm{d}x) \right) \psi_\ell(y) \mu_2(\mathrm{d}y) = 0.$$

Since (ψ_ℓ) is an orthonormal basis of $L^2(M_2, \mu_2)$, we have

$$\int_{M_1} f(x, y) \varphi_k(x)\, \mu_1(\mathrm{d}x) = 0, \quad \text{for } \mu_2\text{-a.e. } y. \tag{B.9}$$

Let $S_k = \{y \in M_2\,;\ \text{(B.9)} \neq 0\}$. Then $\mu_2(S_k) = 0$, and

$$\forall y \notin \bigcup_k S_k : \quad \int_{M_1} f(x, y) \varphi_k(x)\, \mu_1(\mathrm{d}x) = 0 \quad \forall k.$$

Therefore, $f(x, y) = 0$ for all $y \notin \bigcup_k S_k$, μ_1-a.e. $x \in M_1$, and consequently $f(x, y) = 0$ for $\mu_1 \otimes \mu_2$-a.e. $(x, y) \in M_1 \times M_2$. This implies that $(\varphi_k(x)\, \psi_\ell(y))_{k\ell}$ is a basis of $L^2(M_1 \times M_2, \mu_1 \otimes \mu_2)$.

Let

$$U : \varphi_k \otimes \psi_\ell \mapsto \varphi_k(x)\psi_\ell(y).$$

Then U maps the orthonormal bases $(\varphi_k)_k$ of $L^2(M_1, \mu_1)$ and $(\varphi_\ell)_\ell$ of $L^2(M_2, \mu_2)$ onto the orthonormal basis $(\varphi_k \otimes \psi_\ell)$ of $L^2(M_1 \times M_2, \mu_1 \otimes \mu_2)$. Thus U extends uniquely to a unitary isomorphism

$$U : L^2(M_1, \mu_1) \otimes L^2(M_2, \mu_2) \to L^2(M_1 \times M_2, \mu_1 \otimes \mu_2).$$

Note that for $f \in L^2(M_1, \mu_1)$ and $g \in L^2(M_2, \mu_2)$,

$$\begin{aligned} U(f \otimes g) &= U\left(\sum_k c_k \varphi_k \otimes \sum_\ell d_\ell \psi_\ell\right) = U\left(\sum_{k,\ell} c_k d_\ell \varphi_k \otimes \psi_\ell\right) \\ &= \sum_{k,\ell} c_k d_\ell \varphi_k(x)\psi_\ell(y) = f(x)g(y). \end{aligned}$$

Thus

$$L^2(M_1 \times M_2, \mu_2 \otimes \mu_2) \overset{U}{\cong} L^2(M_1, \mu_1) \otimes L^2(M_2, \mu_2).$$

Also, if (M, μ) is a measure space, and H is a separable Hilbert space with orthonormal basis $(\varphi_k)_k$, then we have in H:

$$\forall g \in L^2(M, \mu; H) : \quad g(x) = \lim_{N\to\infty} \sum_{k=1}^{N} \underbrace{(\varphi_k, g(x))_H}_{=: f_k(x) \in L^2(M, \mathrm{d}\mu)} \varphi_k.$$

Define

$$U : \sum_{k=1}^{N} f_k(x) \otimes \varphi_k \mapsto \sum_{k=1}^{N} f_k(x)\varphi_k.$$

Then U is well-defined on a dense subset of $L^2(M, \mu) \otimes H$ onto a dense set in $L^2(M, \mu; H)$, preserving norms. Thus U extends uniquely to a unitary operator

$$U : L^2(M, \mu) \otimes H \to L^2(M, \mu; H).$$

Theorem B.17. Let (M_1, μ_1) and (M_2, μ_2) be measure spaces such that $L^2(M_1, \mathrm{d}\mu_1)$ and $L^2(M_2, \mathrm{d}\mu_2)$ are separable.

(a) There is a unique unitary isomorphism from $L^2(M_1, \mu_1) \otimes L^2(M_2, \mu_2)$ to $L^2(M_1 \times M_2, \mu_1 \otimes \mu_2)$ such that $f \otimes g \mapsto fg$.

(b) For any separable Hilbert space H, there exists a unique unitary isomorphism from $L^2(M_1, \mu_1) \otimes H$ to $L^2(M_1, \mu_1; H)$ such that $f(x) \otimes \varphi \mapsto f(x)\varphi$.

(c) There exists a unique unitary isomorphism from $L^2(M_1 \times M_2, \mu_1 \otimes \mu_2)$ to $L^2(M_1, \mu_1; L^2(M_2, \mu_2))$ satisfying $f(x, y) \mapsto f(x, \cdot)$.

B.4. Linear operators on Hilbert spaces

Again we denote by H and U *real*, *separable* Hilbert spaces, with norms $\|\cdot\|_H$, $\|\cdot\|_U$, inner products $\langle\cdot,\cdot\rangle_H$, $\langle\cdot,\cdot\rangle_U$ and associated Borel sets $\mathcal{B}(H)$ and $\mathcal{B}(U)$, respectively.

Let $\mathcal{L}(U,H)$ denote the Banach space of all bounded linear maps from U into H. For $U = H$, we write $\mathcal{L}(H) = \mathcal{L}(H,H)$. For $T \in \mathcal{L}(H)$, $T \geq 0$ denotes a non-negative, self-adjoint operator, for which $\langle h, Th\rangle_H \geq 0$ for all $h \in H$. Let $\mathcal{L}^+(H)$ be the set of all such operators,

$$\mathcal{L}^+(H) := \{T \in \mathcal{L}(H)\,;\ \langle Tx, x\rangle \geq 0 \ \wedge\ \langle Tx, y\rangle = \langle x, Ty\rangle \quad \forall x, y \in H\}.$$

By $\mathcal{K}(U,H) \subset \mathcal{L}(U,H)$ we denote the set of compact linear operators from U to H.

In building Gaussian measures on Hilbert spaces, we shall be using two important subsets of $\mathcal{L}(U,H)$: the *nuclear operators*, which are also called *trace-class operators*, and the *Hilbert–Schmidt operators*. We recapitulate basic properties as needed here and refer to the Appendix of Peszat and Zabczyk (2007) for a more detailed overview.

Definition B.18. By $\mathcal{L}_1(U,H) \subset \mathcal{L}(U,H)$ we denote the subset of nuclear operators from U to H: $T \in \mathcal{L}(U,H)$ is nuclear if there exist sequences $\{a_j\}_{j\in\mathbb{N}} \subset H$, $\{b_j\}_{j\in\mathbb{N}} \subset U$ such that $\sum_{j=1}^{\infty} \|a_j\|_H \|b_j\|_U < \infty$ and such that

$$\forall f \in U: \quad Tf = \sum_{j=1}^{\infty} \langle f, b_j\rangle_U a_j. \tag{B.10}$$

The set $\mathcal{L}_1(U,H)$ is a Banach space with norm

$$\|T\|_{\mathcal{L}_1(U,H)} := \inf\left\{\sum_{j=1}^{\infty} \|a_j\|_H \|b_j\|_U\ ;\ Tf = \sum_{j=1}^{\infty} \langle f, b_j\rangle_U a_j, \quad \forall f \in U\right\}. \tag{B.11}$$

Note that $\mathcal{L}_1(U,H) \subset \mathcal{K}(U,H)$ since each $T \in \mathcal{L}_1(U,H)$ can be approximated in operator norm by a sequence of operators of finite rank.

Lemma B.19. Let $T \in \mathcal{L}_1(H)$ and let $\{e_k\}_{k=1}^{\infty}$ be an orthonormal basis of H. Then

$$\operatorname{Tr} T = \sum_{k=1}^{\infty} \langle Te_k, e_k\rangle_H \tag{B.12}$$

exists and is independent of the particular choice of orthonormal basis.

Definition B.20. We call $T \in \mathcal{L}(U,H)$ a *Hilbert–Schmidt operator* (HS operator) if

$$\sum_{k=1}^{\infty} \|Te_k\|_H^2 < \infty \tag{B.13}$$

for *some* orthonormal basis $(e_k)_{k=1}^{\infty}$ of U. We denote the subset of $\mathcal{L}(U,H)$ of HS operators by $\mathcal{L}_2(U,H)$.

The linear space $\mathcal{L}_2(U,H)$ of HS operators from U into H is a separable Hilbert space: its scalar product is defined in terms of the orthonormal basis $(e_k)_k \subset U$ of U,

$$\langle S,T\rangle_{HS} := \sum_{k=1}^{\infty} \langle Se_k, Te_k\rangle_H. \tag{B.14}$$

We denote by $\|\cdot\|_{HS}$ the corresponding Hilbert–Schmidt norm. For $S \in \mathcal{L}_2(U,H)$ and orthonormal bases $(e_k)_k \subset U$, $(f_k)_k \subset H$ of U and of H, respectively, we have

$$\sum_k \|Se_k\|_H^2 = \sum_{kj} \langle Se_k, f_j\rangle_H^2 = \sum_{kj} \langle e_k, S^*f_j\rangle_U^2 = \sum_j \|S^*f_j\|_U^2.$$

Therefore we have the following.

Proposition B.21. The HS operator norm $\|\cdot\|_{HS}$ does not depend on the choice of orthonormal basis for U. Also, $S \in \mathcal{L}_2(U,H)$ if and only if $S^* \in \mathcal{L}_2(H,U)$ and $\|S\|_{\mathcal{L}_2(U,H)} = \|S^*\|_{\mathcal{L}_2(H,U)}$.

Moreover, if $(f_k)_k \subset H$ and $(e_k)_k \subset U$ are orthonormal bases, then the rank-one operators $(f_j \otimes e_k)_{j,k\in\mathbb{N}}$ defined by

$$(f_j \otimes e_k)(u) := f_j\langle e_k, u\rangle_U, \quad u \in U$$

are an orthonormal basis of $\mathcal{L}_2(U,H)$.

We collect further properties of operators in $\mathcal{L}_1$ and $\mathcal{L}_2$.

Proposition B.22.

(a) For $S \in \mathcal{L}_1(U,H)$ and $T \in \mathcal{L}(H,V)$, $TS \in \mathcal{L}_1(U,V)$ and

$$\|TS\|_{\mathcal{L}_1(U,V)} \le \|S\|_{\mathcal{L}_1(U,H)}\|T\|_{\mathcal{L}(H,V)}.$$

(b) If $S \in \mathcal{L}(U,H)$, $T \in \mathcal{L}_1(H,V)$ then $TS \in \mathcal{L}_1(U,V)$ and

$$\|TS\|_{\mathcal{L}_1(U,V)} \le \|S\|_{\mathcal{L}(U,H)}\|T\|_{\mathcal{L}_1(H,V)}.$$

(c) If $S \in \mathcal{L}(U,H)$ and $T \in \mathcal{L}(H,U)$, and if either S or T is of trace class, then $TS \in \mathcal{L}_1(U)$ and $\mathrm{Tr}(TS) = \mathrm{Tr}(ST)$.

(d) if $S \in \mathcal{L}(U)$ and if $T \in \mathcal{L}_2(U,H)$ then $TS \in \mathcal{L}_2(U,H)$ and

$$\|TS\|_{\mathcal{L}_2(U,H)} \le \|T\|_{\mathcal{L}_2(U,H)}\|S\|_{\mathcal{L}(U)}.$$

(e) If $\mathcal{K}(U,H) \subset \mathcal{L}(U,H)$ denotes the subset of compact linear operators from U to H,

$$\mathcal{L}_1(U,H) \subset \mathcal{L}_2(U,H) \subset \mathcal{K}(U,H) \subset \mathcal{L}(U,H).$$

Proof. We show (e): we can write $R \in \mathcal{L}_1(U,H)$ as $R = \sum_k b_k \otimes a_k$, where $(a_k)_k \subset U$ and $(b_k)_k \subset H$ are bases such that $\sum_k \|a_k\|_U \|b_k\|_H < \infty$ and we recall that $b \otimes a(u) = b\langle a, u\rangle_U$. Let $(e_n)_n \subset U$ denote an orthonormal basis of U. Then

$$\begin{aligned}\sum_n \|Re_n\|_H^2 &= \sum_n \left\| \sum_k b_k \langle a_k, e_n\rangle_U \right\|_H^2 \\ &\le \sum_n \sum_{kl} |\langle b_k, b_l\rangle| \, |\langle a_k, e_n\rangle_U| \, |\langle a_l, e_n\rangle_U| \\ &\le \sum_{kl} \|b_k\|_H \|b_l\|_H \left(\sum_n \langle a_k, e_n\rangle_U^2 \right)^{1/2} \left(\sum_n \langle a_l, e_n\rangle_U^2 \right)^{1/2} \\ &\le \left(\sum_k \|b_k\|_H \|a_k\|_U \right)^2.\end{aligned}$$

Taking the infimum over all bases $(a_k)_k \subset U$ and $(b_k)_k \subset H$ such that $\sum_k \|a_k\|_U \|b_k\|_H < \infty$, we obtain the estimate

$$\|R\|_{\mathcal{L}_2(U,H)} \le \|R\|_{\mathcal{L}_1(U,H)},$$

which proves $\mathcal{L}_1(U,H) \subset \mathcal{L}_2(U,H)$.

The next inclusion follows from the fact that the HS norm is a stronger norm than the operator norm, and that $\mathcal{K}(U,H)$ is closed in $\mathcal{L}(U,H)$ and that each HS operator $R \in \mathcal{L}_2(U,H)$ is the limit, in the HS norm, of the sequence of operators of finite rank,

$$T_n = \sum_{k \le n} f_k \otimes T^* f_k, \quad n \in \mathbb{N}. \qquad \square$$

We recall the spectral theorem for compact, self-adjoint operators.

Proposition B.23. For $Q \in \mathcal{K}(H)$ and $Q = Q^*$, there exists an orthonormal basis $(e_k)_{k=1}^\infty$ of eigenfunctions of H such that $Qe_k = \lambda_k e_k$, and a decreasing sequence $(\lambda_k)_k \subset \mathbb{R}$, $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$ of real, non-negative eigenvalues which accumulate only at zero. Moreover,

$$\forall x \in H: \quad Qx = \sum_{k=1}^\infty \lambda_k \langle x, e_k\rangle e_k,$$

and if $Q \in \mathcal{L}_1^+(H)$,

$$\infty > \mathrm{Tr}(Q) = \sum_{k=1}^\infty \lambda_k.$$

We finally add a result which is useful in the context of *Karhunen–Loève* expansion of random fields. It should be compared to Theorem B.17.

Proposition B.24. Let $(E_i, \mathcal{E}_i)$, $i = 1, 2$ be two measurable spaces and let μ_i be σ-finite measures on $(E_i, \mathcal{E}_i)$, $i = 1, 2$. Further, let $E = E_1 \times E_2$ and $\mathcal{E} := \mathcal{E}_1 \otimes \mathcal{E}_2$ denote the product space. Then product measure $\mu = \mu_1 \otimes \mu_2$ is σ-finite on $(E, \mathcal{E})$. Further, let $U = L^2(E_1, \mathcal{E}_1, \mu_1)$ and $H = L^2(E_2, \mathcal{E}_2, \mu_2)$ be separable.

Then an operator $R \in \mathcal{L}(U, H)$ belongs to $\mathcal{L}_2(U, H)$ if and only if there is a *kernel* $K \in L^2(E, \mathcal{E}, \mu)$ such that

$$\forall \psi \in U, \xi \in E_2 : \quad R\psi(\xi) := \int_{E_1} K(\eta, \xi)\psi(\eta)\mu_1(\mathrm{d}\eta). \tag{B.15}$$

Then

$$\|R\|^2_{\mathcal{L}_2(U,H)} = \int_{E_1}\int_{E_2} |K(\eta, \xi)|^2 \mu_1(\mathrm{d}\eta)\mu_2(\mathrm{d}\xi). \tag{B.16}$$

Proof. Let $(e_k)_{k=1}^\infty$ be an orthonormal basis of $L^2(E_1, \mathcal{E}_1, \mu_1)$ and let $(f_k)_{k=1}^\infty$ be an orthonormal basis of $= L^2(E_2, \mathcal{E}_2, \mu_2)$. Further, let the operator R be given by (B.15).

Then the Parseval equality (Theorem B.8(d)) implies

$$\begin{aligned}\sum_{n=1}^\infty \|Re_n\|_H^2 &= \int_{E_2}\left(\int_{E_1} K(\eta, \xi)e_n(\eta)\mu_1(\mathrm{d}\eta)\right)^2 \mu_2(\mathrm{d}\xi)\\ &= \int_{E_2}\int_{E_1} |K(\eta, \xi)|^2 \mu(\mathrm{d}\eta)\mu_2(\mathrm{d}\xi).\end{aligned}$$

Conversely, assume that $R \in \mathcal{L}_2(U, H)$ and let $(e_n)_{n=1}^\infty$ be an orthonormal basis of $L^2(E_1, \mathcal{E}_1, \mu_1)$ and let $(f_k)_{k=1}^\infty$ be an orthonormal basis of $L^2(E_2, \mathcal{E}_2, \mu_2)$. Define the kernel $K(\eta, \xi) : U \times H \to \mathbb{R}$ by

$$K(\eta, \xi) := \sum_{n=1}^\infty \sum_{k=1}^\infty \langle Re_n, f_k\rangle_N e_n(\eta)f_k(\xi). \qquad \square$$

C. Review of Gaussian measures on Hilbert spaces

C.1. Measures on metric spaces

For any complete metric space E, denote by $\mathcal{B}(E)$ the Borel σ-algebra generated by all closed or, equivalently, open sets of E.

A random variable in $(\Omega, \Sigma, \mathbb{P})$ with values in E is a mapping $X : \Omega \to E$ such that

$$\forall I \in \mathcal{B}(E) : \quad X^{-1}(I) \in \Sigma.$$

The *law of* X is the probability measure $X_\#\mathbb{P}$ on $(E, \mathcal{B}(E))$ defined by

$$X_\#\mathbb{P}(I) := \mathbb{P}(X^{-1}(I)) = \mathbb{P}(X \in I) \quad \forall I \in \mathcal{B}(E).$$

Proposition C.1. (change of variables) Let X be a random variable on $(\Omega, \Sigma, \mathbb{P})$ with values in E. Let $\varphi : E \to \mathbb{R}$ be bounded and $\mathcal{B}(\mathbb{R})$-measurable. Then

$$\int_\Omega \varphi(X(\omega))\mathbb{P}(\mathrm{d}\omega) = \int_E \varphi(x)\, X_\# \mathbb{P}(\mathrm{d}x). \tag{C.1}$$

Proof. We show (C.1) for $\varphi = 1_I$, $I \in \mathcal{B}(E)$. In this case,

$$\varphi(X(\omega)) = 1_{\underbrace{X^{-1}(I)}_{\in\Sigma}}(\omega), \quad \omega \in \Omega.$$

Hence,

$$\int_\Omega \varphi(X(\omega))\, \mathbb{P}(\mathrm{d}\omega) = \mathbb{P}(X^{-1}(I)) = X_\#\mathbb{P}(I) = \int_E \varphi(x)\, X_\#\mathbb{P}(\mathrm{d}x).$$

The general case follows by approximating $\varphi(x)$ by simple functions. □

C.2. Gaussian measures on separable Hilbert spaces

We present a concrete construction of Gaussian measures on separable Hilbert spaces as countable products of Gaussian measures on $\mathbb{R}$, generalizing naturally the construction of Gaussian measures in $\mathbb{R}^d$ for finite dimensions $d < \infty$, following Da Prato (2006).

Gaussian measure on $\mathbb{R}$

For a pair (a, λ) of real numbers with $a \in \mathbb{R}$, $\lambda \geq 0$, define a measure $N_{a,\lambda}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ as follows. If $\lambda = 0$,

$$N_{a,0} := \delta_a,$$

where

$$\delta_a(B) := \begin{cases} 1 & \text{if } a \in B, \\ 0 & \text{else,} \end{cases} \quad B \in \mathcal{B}(\mathbb{R}),$$

and if $\lambda > 0$,

$$N_{a,\lambda}(B) := \frac{1}{\sqrt{2\pi\lambda}} \int_B \mathrm{e}^{-\frac{(x-a)^2}{2\lambda}}\, \mathrm{d}x, \quad B \in \mathcal{B}(\mathbb{R}).$$

Note that

$$N_{a,\lambda}(\mathbb{R}) = \frac{1}{\sqrt{2\pi\lambda}} \int_{-\infty}^{\infty} \mathrm{e}^{-\frac{(x-a)^2}{2\lambda}}\, \mathrm{d}x = 1,$$

hence $N_{a,\lambda}$ is a probability measure. As is well known, $N_{a,\lambda}$ is absolutely continuous with respect to the Lebesgue measure $\mathrm{d}x$, with explicit density

$$N_{a,\lambda}(\mathrm{d}x) = \frac{1}{\sqrt{2\pi\lambda}} \mathrm{e}^{-\frac{(x-a)^2}{2\lambda}}\, \mathrm{d}x.$$

Proposition C.2. For any $a \in \mathbb{R}$, $\lambda \geq 0$, the *mean* of $N_{a,\lambda}$ is

$$\int_{\mathbb{R}} x N_{a,\lambda}(\mathrm{d}x) = a,$$

and its *variance* is

$$\int_{\mathbb{R}} (x-a)^2 N_{a,\lambda}(\mathrm{d}x) = \lambda.$$

For all $m \in \mathbb{N}$, the mth moment of the Gaussian measure $N_{a,\lambda}$ is finite, *i.e.*,

$$\int_{\mathbb{R}} x^m N_{a,\lambda}(\mathrm{d}x) < \infty,$$

and its *Fourier transform* is given by

$$\widehat{N}_{a,\lambda}(h) := \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}hx} N_{a,\lambda}(\mathrm{d}x) = \mathrm{e}^{\mathrm{i}ah - \frac{1}{2}\lambda h^2}, \quad h \in \mathbb{R}.$$

Gaussian measures on finite-dimensional Hilbert spaces

We consider a finite-dimensional Hilbert space H, $d := \dim H < \infty$ with scalar product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|_H$. To this end, we recall that for probability space $(\Omega_i, \Sigma_i, \mathbb{P}_i)$, $i = 1, \ldots, d$, the product probability space is given by

$$(\Omega, \Sigma, \mathbb{P}) := \bigotimes_{i=1}^{d} (\Omega_i, \Sigma_i, \mathbb{P}_i)$$

(see Appendix A).

Let $a \in H$, $Q \in \mathcal{L}^+(H)$ (see Appendix B). Then Q can be represented in any orthonormal basis $(e_i)_{i=1}^d$ of H as a $d \times d$ symmetric, positive semidefinite matrix $(\langle e_j, Qe_j \rangle)_{i,j=1}^d$. We now choose a particular orthonormal basis $(e_i)_{i=1}^d$ such that Q becomes diagonal, *i.e.*, such that

$$Qe_k = \lambda_k e_k, \quad k = 1, \ldots, d, \quad \lambda_k \geq 0.$$

Let

$$\forall x \in H: \quad x_k := \langle x, e_k \rangle, \quad k = 1, \ldots, d.$$

Then $a = (a_1, \ldots, a_d)$, $a_k = \langle a, e_k \rangle$, and H is isomorphic to $\mathbb{R}^d$ via the unitary map

$$\gamma : H \to \mathbb{R}^d, \quad x \mapsto \gamma(x) := (x_1, \ldots, x_d).$$

Define a product measure $N_{a,Q}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ by

$$N_{a,Q} := \bigotimes_{k=1}^{d} N_{a_k, \lambda_k}.$$

If $a = 0$, we write N_Q for $N_{a,Q}$.

Proposition C.3. Let H be a Hilbert space with $d := \dim H < \infty$, and let $Q \in \mathcal{L}^+(H)$. Then

$$\int_H x N_{a,Q}(\mathrm{d}x) = a \in H,$$

for all $y, z \in H$,

$$\int_H \langle y, x-a\rangle \langle z, x-a\rangle N_{a,Q}(\mathrm{d}x) = \langle Qy, z\rangle,$$

and for all $h \in H$,

$$\widehat{N}_{a,Q}(h) := \int_H \mathrm{e}^{\mathrm{i}\langle h,x\rangle} N_{a,Q}(\mathrm{d}x) = \mathrm{e}^{\mathrm{i}\langle a,h\rangle - \frac{1}{2}\langle Qh,h\rangle}.$$

If, moreover, $\det Q > 0$, then $N_{a,Q}(\mathrm{d}x)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, and

$$N_{a,Q}(\mathrm{d}x) = \frac{1}{\sqrt{(2\pi)^d \det Q}} \mathrm{e}^{-\frac{1}{2}\langle Q^{-1}(x-a),(x-a)\rangle}\, \mathrm{d}x.$$

The vector $a \in H$ is the *mean* and $Q \in H \otimes H$ is called the *covariance operator* of $N_{a,Q}$.

Proposition C.4. Let H be a finite-dimensional Hilbert space, $a \in H$, $Q \in \mathcal{L}^+(H)$, and let μ be a finite measure on $(H, \mathcal{B}(H))$ such that

$$\forall h \in H: \quad \int_H \mathrm{e}^{\mathrm{i}\langle h,x\rangle} \mu(\mathrm{d}x) = \mathrm{e}^{\mathrm{i}\langle a,h\rangle - \frac{1}{2}\langle Qh,h\rangle}.$$

Then

$$\mu = N_{a,Q}.$$

Measures on separable Hilbert spaces

Let H be a separable Hilbert space with $\dim H = \infty$, scalar product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|_H$. For any $n \in \mathbb{N}$ and any orthonormal basis $(e_k)_{k=1}^\infty$ of H, we define the projection $P_n : H \to P_n(H) = \mathrm{span}\{e_1, \ldots, e_n\} \subset H$ by

$$P_n x := \sum_{k=1}^n \langle x, e_k\rangle e_k, \quad x \in H. \tag{C.2}$$

Then, for any $x \in H$, $\|x - P_n x\|_H \to 0$ as $n \to \infty$. Denote by $\mathcal{M}(H)$ the set of all bounded measures on $(H, \mathcal{B}(H))$.

Proposition C.5. (uniqueness of measures on H) For $\mu, \nu \in \mathcal{M}(H)$, if, for all continuous, bounded $\varphi : H \to \mathbb{R}$,

$$\int_H \varphi(x)\, \mu(\mathrm{d}x) = \int_H \varphi(x)\, \nu(\mathrm{d}x), \tag{C.3}$$

then $\mu = \nu$.

Proof. Let $C \subset H$ be closed, and let $(\varphi_n)_n$ be a sequence of continuous, bounded functions on H such that

$$\forall x \in H : \varphi_n(x) \underset{n\to\infty}{\longrightarrow} 1_C(x), \quad \sup_{x\in H} |\varphi_n(x)| \le 1, \tag{C.4}$$

e.g.,

$$\varphi_n(x) := \begin{cases} 1 & \text{if } x \in C, \\ 1 - n\, d(x, C) & \text{if } x \notin C \wedge d(x, C') < \dfrac{1}{n}, \\ 0 & \text{if } d(x, C) \ge \dfrac{1}{n}. \end{cases}$$

By dominated convergence,

$$\int_H \varphi_n \,\mathrm{d}\mu = \int_H \varphi_n \,\mathrm{d}\nu \underset{n\to\infty}{\longrightarrow} \mu(C) = \nu(C).$$

Since the closed subsets of H are $\cap$-stable and generate $\mathcal{B}(H)$, $\mu = \nu$. □

Proposition C.6. Let μ, ν be finite measures on $(H, \mathcal{B}(H))$ such that for all $n \in \mathbb{N}$, $(P_n)_\# \mu = (P_n)_\# \nu$. Then

$$\mu = \nu \quad \text{in} \quad \mathcal{M}(H).$$

Proof. Let $\varphi : H \to \mathbb{R}$ be continuous and bounded. By dominated convergence,

$$\int_H \varphi(x)\, \mu(\mathrm{d}x) = \lim_{n\to\infty} \int_H \varphi(P_n x)\, \mu(\mathrm{d}x).$$

The change of variables formula (C.1) implies

$$\begin{aligned} \int_H \varphi(x)\, \mu(\mathrm{d}x) &= \lim_{n\to\infty} \int_H \varphi(P_n x)\, \mu(\mathrm{d}x) = \lim_{n\to\infty} \int_{P_n(H)} \varphi(\xi)(P_n)_\# \mu(\mathrm{d}\xi) \\ &= \lim_{n\to\infty} \int_{P_n(H)} \varphi(\xi)(P_n)_\# \nu(\mathrm{d}\xi) = \lim_{n\to\infty} \int_H \varphi(P_n x)\, \nu(\mathrm{d}\xi) \\ &= \int_H \varphi(x)\, \nu(\mathrm{d}x). \end{aligned}$$

Proposition C.5 implies the assertion $\mu = \nu$. □

Define the *Fourier transform* $\widehat{\mu}$ of $\mu \in \mathcal{M}(H)$ as

$$\forall h \in H : \quad \widehat{\mu}(h) := \int_H \mathrm{e}^{\mathrm{i}\langle x, h\rangle} \mu(\mathrm{d}x). \tag{C.5}$$

Proposition C.7. (Fourier characterization) Let μ, ν be finite measures on $(H, \mathcal{B}(H))$. Then

$$\forall h \in H : \quad \widehat{\mu}(h) = \widehat{\nu}(h) \quad \Longrightarrow \quad \mu = \nu \quad \text{in} \quad \mathcal{M}(H).$$

Proof. By (C.1), for all $n \in \mathbb{N}$,

$$\widehat{\mu}(P_n h) = \int_H \mathrm{e}^{\mathrm{i}\langle x, P_n h\rangle}\,\mu(\mathrm{d}x) = \int_{P_n(H)} \mathrm{e}^{\mathrm{i}\langle \xi, h\rangle}\,(P_n)_{\#}\mu(\mathrm{d}\xi) = \widehat{(P_n)_{\#}\mu}(h),$$

and similarly

$$\widehat{\nu}(P_n h) = \int_H \mathrm{e}^{\mathrm{i}\langle x, P_n h\rangle}\,\nu(\mathrm{d}x) = \int_{P_n(H)} \mathrm{e}^{\mathrm{i}\langle \xi, h\rangle}\,(P_n)_{\#}\nu(\mathrm{d}\xi) = \widehat{(P_n)_{\#}\nu}(h).$$

By assumption, $\widehat{\mu}(P_n h) = \widehat{\nu}(P_n h)$ for all $n \in \mathbb{N}$, hence $\widehat{(P_n)_{\#}\mu} = \widehat{(P_n)_{\#}\nu}$, and thus $(P_n)_{\#}\mu = (P_n)_{\#}\nu$ by a generalization of Proposition C.4. Then Proposition C.6 implies $\mu = \nu$ in $\mathcal{M}(H)$. □

Let $\mathcal{P}(H) \subset \mathcal{M}(H)$ be the set of probability measures on $(H, \mathcal{B}(H))$. Let $\mu \in \mathcal{P}(H)$ satisfy

$$\int_H \|x\|_H\,\mu(\mathrm{d}x) < \infty, \quad \int_H \|x\|_H^2\,\mu(\mathrm{d}x) < \infty.$$

Define

$$F : H \to \mathbb{R}, \quad F(h) := \int_H \langle x, h\rangle\,\mu(\mathrm{d}x), \quad h \in H.$$

Then $F \in H'$ since $F(\cdot)$ is linear and by the Cauchy–Schwarz inequality,

$$|F(h)| \le \int_H \|x\|_H\,\mu(\mathrm{d}x)\|h\|_H \quad \forall h \in H,$$

so

$$\|F\|_{H'} = \sup_{0 \ne h \in H} \frac{|F(h)|}{\|h\|_H} \le \int_H \|x\|_H\,\mu(\mathrm{d}x).$$

By the Riesz representation theorem, there is a unique $m \in H$ such that

$$\langle m, h\rangle = \int_H \langle x, h\rangle\,\mu(\mathrm{d}x), \quad \forall h \in H;$$

m is called the *mean* of $\mu \in \mathcal{P}(H)$; we write $\mathrm{mean}(\mu) := m$.

Consider next the bilinear form $G(\cdot,\cdot) : H \times H \to \mathbb{R}$ defined by

$$G(h,k) := \int_H \langle h, x - m\rangle\langle k, x - m\rangle\,\mu(\mathrm{d}x), \quad h, k \in H.$$

Then by the Cauchy–Schwarz inequality,

$$|G(h,k)| \le \int_H \|x - m\|_H^2\,\mu(\mathrm{d}x)\|h\|_H\|k\|_H \quad \forall h, k \in H.$$

By the Riesz representation theorem, there exists a unique $Q \in \mathcal{L}(H)$ such that

$$\langle Qh, k\rangle = \int_H \langle h, x - m\rangle\langle k, x - m\rangle\,\mu(\mathrm{d}x), \quad \forall h, k \in H.$$

The operator $Q \in \mathcal{L}(H)$ is called the *covariance* (*operator*) of $\mu \in \mathcal{P}(H)$; we write $\mathrm{Cov}(\mu) := Q \in \mathcal{L}(H)$.

Proposition C.8. Let $\mu \in \mathcal{P}(H)$ such that $m = \mathrm{mean}(\mu) \in H$ and $Q = \mathrm{Cov}(\mu) \in \mathcal{L}(H)$ exist. Then $Q \in \mathcal{L}_1^+(H)$, *i.e.*, Q is symmetric, positive and of trace class.

Proof. $\langle Qh, k\rangle = \langle h, Qk\rangle$ for all $h, k \in H$ by definition. For any orthonormal basis $(e_k)_k$ of H,

$$\forall k \in \mathbb{N}: \quad \langle Qe_k, e_k\rangle = \int_H |\langle x - m, e_k\rangle|^2 \mu(\mathrm{d}x).$$

By monotone convergence and Parseval's identity, $Q \in \mathcal{L}(H)$ implies

$$\infty > \mathrm{Tr}\, Q = \sum_{k=1}^{\infty} \int_H |\langle x - m, e_k\rangle|^2 \, \mu(\mathrm{d}x) = \int_H \|x - m\|_H^2 \, \mu(\mathrm{d}x). \qquad \square$$

We close with a lemma on kth moments of a measure μ on H.

Lemma C.9. Let $\mu \in \mathcal{M}(H)$ be a probability measure on $(H, \mathcal{B}(H))$ and let $k \in \mathbb{N}$ be such that

$$\forall h \in H: \quad \int_H |\langle h, x\rangle|^k \mu(\mathrm{d}x) < \infty.$$

Then there exists a constant $C(k, \mu)$ such that, for all $h_1, \ldots, h_k \in H$,

$$\int_H |\langle h_1, x\rangle \cdots \langle h_k, x\rangle| \mu(\mathrm{d}x) \le C(k, \mu) \|h_1\|_H \cdots \|h_k\|_H.$$

In particular, the symmetric k-form

$$\underbrace{H \otimes \cdots \otimes H}_{k \text{ times}} \ni (h_1, \ldots, h_k) \mapsto \int_H \langle h_1, x\rangle \cdots \langle h_k, x\rangle \mu(\mathrm{d}x)$$

is continuous. Observing that for

$$H^{(k)} = \underbrace{H \otimes \cdots \otimes H}_{k \text{ times}} \quad \text{we have} \quad (H^{(k)})' \simeq \underbrace{H' \otimes \cdots \otimes H'}_{k \text{ times}},$$

by the Riesz representation theorem there exists a unique $\mathcal{M}^k \mu \in (H^{(k)})'$, the *kth moment of the measure* μ, such that, for all $h_1 \otimes \cdots \otimes h_k \in H^{(k)}$,

$${}_{(H')^{(k)}}\langle \mathcal{M}^k \mu, h_1 \otimes \cdots \otimes h_k\rangle_{H^{(k)}} = \int_H \langle h_1, x\rangle \cdots \langle h_k, x\rangle \mu(\mathrm{d}x).$$

Gaussian measures on separable Hilbert spaces

Definition C.10. (Gaussian measure) Let $a \in H$ and $Q \in \mathcal{L}_1^+(H)$. The *Gaussian measure*

$$\mu := N_{a,Q} \quad \text{on} \quad (H, \mathcal{B}(H))$$

with mean $\mathrm{mean}(\mu) = a$ and covariance $\mathrm{Cov}(\mu) = Q$ is the $\mu \in \mathcal{P}(H)$ with Fourier transform

$$\widehat{\mu}(h) = \widehat{N_{a,Q}}(h) = \exp\left(\mathrm{i}\langle a, h\rangle - \frac{1}{2}\langle Qh, h\rangle\right), \quad h \in H.$$

Theorem C.11. On any separable Hilbert space H, for any $a \in H$ and $Q \in \mathcal{L}_1^+(H)$, there is a unique Gaussian measure $N_{a,Q}$.

Proof. Since H is separable and $Q \in \mathcal{L}_1^+(H)$, the spectral theorem implies that there is an orthonormal basis $(e_k)_k$ of H and a sequence $(\lambda_k)_k \subset \mathbb{R}_{\geq 0}$ such that

$$Qe_k = \lambda_k e_k \quad \forall k \in \mathbb{N}.$$

For $x \in H$, set $x_k := \langle x, e_k\rangle$, $k \in \mathbb{N}$. Then $(x_k)_k \in \ell^2(\mathbb{N})$. Note that $H \cong \ell^2(\mathbb{N})$ via $\gamma : H \to \ell^2$ defined by

$$x \mapsto \gamma(x) := (x_k)_k \in \ell^2(\mathbb{N}).$$

Define, on $H = \ell^2(\mathbb{N})$, the measure

$$\mu := \bigotimes_{k=1}^{\infty} \underbrace{N_{a_k,\lambda_k}}_{\mu_k}, \quad x = (x_1, x_2, \ldots). \tag{C.6}$$

Note that formally, μ is defined on $\mathbb{R}^\infty$ rather than $\ell^2(\mathbb{N})$. □

Proposition C.12. $\ell^2(\mathbb{N}) \in \mathcal{B}(\mathbb{R}^\infty)$, and for μ as in (C.6), $\mu(\ell^2(\mathbb{N})) = 1$.

Proof. We leave the first statement as an exercise. By monotone convergence,

$$\begin{aligned}\int_{\mathbb{R}^\infty} \|x\|_{\ell^2(\mathbb{N})}^2 \,\mu(\mathrm{d}x) &= \int_{\mathbb{R}^\infty} \left(\sum_{k=1}^{\infty} |x_k|^2\right)^2 \mu(\mathrm{d}x) = \sum_{k=1}^{\infty} \int_{\mathbb{R}} |x_k|^2 \, N_{a_k,\lambda_k}(\mathrm{d}x_k) \\ &= \sum_{k=1}^{\infty} \left(\int_{\mathbb{R}} (x_k - a_k)^2 \, N_{a_k,\lambda_k}(\mathrm{d}x_k) + a_k^2\right) \\ &= \sum_{k=1}^{\infty} (\lambda_k + a_k^2) = \mathrm{Tr}\, Q + \|a\|_{\ell^2(\mathbb{N})} < \infty.\end{aligned}$$

Therefore,

$$\mu(\{x \in \mathbb{R}^\infty \,;\, \|x\|_{\ell^2} = \infty\}) = 0,$$

which implies the second assertion. □

We next characterize the Fourier transform of Gaussian measures on infinite-dimensional, separable Hilbert spaces.

Theorem C.13. For any $a \in H$ and $Q \in \mathcal{L}_1^+(H)$, there exists a unique $\mu \in \mathcal{P}(H)$ such that

$$\widehat{\mu}(h) = \exp\left(\mathrm{i}\langle a, h\rangle - \frac{1}{2}\langle Qh, h\rangle\right) \quad \forall h \in H. \tag{C.7}$$

Moreover, if H is identified with $\ell^2(\mathbb{N})$ via the eigenbasis of Q, then μ is the restriction to H of the product measure

$$\bigotimes_{k=1}^{\infty} \mu_k = \bigotimes_{k=1}^{\infty} N_{a_k,\lambda_k}.$$

Proof. Since the characteristic function $\widehat{\mu}$ of μ uniquely determines μ by Proposition C.7, we must only show existence. The sequence $\mu_k := N_{a_k,\lambda_k}$ of Gaussian measure on $\mathbb{R}$, $k = 1, 2, \ldots$ admits a unique product measure

$$\mu = \bigotimes_{k=1}^{\infty} \mu_k$$

on $\mathbb{R}^\infty$, and $\mu \in \mathcal{P}(\mathbb{R}^\infty)$.

By Proposition C.12, μ is concentrated on $\ell^2(\mathbb{N})$, *i.e.*, $\mu(\ell^2(\mathbb{N})) = 1$. We denote the restriction $\mu|_{\ell^2(\mathbb{N})}$ again by μ. Define

$$\forall n \in \mathbb{N}: \quad \nu_n := \bigotimes_{k=1}^{n} \mu_k.$$

Then using Proposition C.3,

$$\begin{aligned}\int_{\ell^2(\mathbb{N})} \mathrm{e}^{\mathrm{i}\langle x,h\rangle}\mu(\mathrm{d}x) &= \lim_{n\to\infty}\int_{\ell^2(\mathbb{N})} \mathrm{e}^{\mathrm{i}\langle P_n h, P_n x\rangle}\mu(\mathrm{d}x) = \lim_{n\to\infty}\int_{\mathbb{R}^n} \mathrm{e}^{\mathrm{i}\langle P_n h,\xi\rangle}\nu_n(\mathrm{d}\xi)\\ &= \lim_{n\to\infty} \mathrm{e}^{\mathrm{i}\langle P_n h, P_n a\rangle - \frac{1}{2}\langle QP_n h, P_n h\rangle} = \mathrm{e}^{\mathrm{i}\langle h,a\rangle - \frac{1}{2}\langle Qh,h\rangle}. \qquad \square\end{aligned}$$

Corollary C.14.

$$\int_H \|x\|_H^2 \, N_{a,Q}(\mathrm{d}x) = \operatorname{Tr} Q + \|a\|_H^2. \tag{C.8}$$

Gaussian random fields

For the mathematical formulation of results on existence and regularity of solutions to stochastic PDEs, we require Bochner spaces of random variables taking values in separable Hilbert and Banach spaces.

Definition C.15. (Lebesgue–Bochner spaces) Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, K a separable Hilbert space, and $X : \Omega \to K$ a random variable.

(i) We say that X is a *K-valued Gaussian random variable* if the distribution of X is a Gaussian measure on K.

(ii) For any $0 < p < \infty$, denote by $L^p(\Omega, \Sigma, \mathbb{P}; K)$ the linear space of all random variables $X : \Omega \to K$ for which

$$\int_\Omega \|X(\omega)\|_K^p \, \mathbb{P}(\mathrm{d}\omega) < \infty,$$

and by $L^\infty(\Omega, \Sigma, \mathbb{P}; K)$ the space of all X with $\operatorname{ess\,sup}_\omega \|X(\omega)\|_K < \infty$.

(iii) The space $L^p(\Omega, \Sigma, \mathbb{P}; K)$ endowed with

$$\|X\|_{L^p(\Omega,\Sigma,\mathbb{P};K)} := \left(\int_\Omega \|X(\omega)\|_K^p \, \mathbb{P}(\mathrm{d}\omega) \right)^{1/p} \tag{C.9}$$

if $p < \infty$, and $\|X\|_{L^\infty(\Omega,\Sigma,\mathbb{P};K)} := \operatorname{ess\,sup}_\omega \|X(\omega)\|_K$, is a quasi-Banach space for $0 < p < 1$ and a Banach space if $p \geq 1$.

(iv) For $p = 2$, $L^2(\Omega, \Sigma, \mathbb{P}; K)$ equipped with the inner product

$$\langle X, Y \rangle_{L^2(\Omega,\Sigma,\mathbb{P};K)} := \int_\Omega \langle X(\omega), Y(\omega) \rangle_K \, \mathbb{P}(\mathrm{d}\omega) \tag{C.10}$$

is a Hilbert space.

Example C.16. Let $X \in L^2(\Omega, \Sigma, \mathbb{P}; K)$. Then X has mean $m_X \in K$,

$$\langle m_X, h \rangle_K = \int_K \langle y, h \rangle_K \, X_\# \mathbb{P}(\mathrm{d}y) = \int_\Omega \langle X(\omega), h \rangle_K \, \mathbb{P}(\mathrm{d}\omega) \quad \forall h \in K, \tag{C.11}$$

and covariance $Q_X \in \mathcal{L}(K)$,

$$\begin{aligned} \langle Q_X h, k \rangle_K &= \int_K \langle y - m_X, h \rangle_K \langle y - m_X, k \rangle_K \, X_\# \mathbb{P}(\mathrm{d}y) \\ &= \int_\Omega \langle X(\omega) - m_X, h \rangle_K \langle X(\omega) - m_X, k \rangle_K \, \mathbb{P}(\mathrm{d}\omega) \end{aligned} \tag{C.12}$$

for all $h, k \in K$. We abbreviate $\operatorname{mean}(X) := m_X = \operatorname{mean}(X_\# \mathbb{P})$ and

$$\operatorname{Cov}(X) := Q = \operatorname{Cov}(X_\# \mathbb{P}).$$

Proposition C.17. (convergence of Gaussian RV) Let $(X_n)_n$ be a sequence of Gaussian random variables in $(\Omega, \Sigma, \mathbb{P})$, taking values in a separable Hilbert space K. Let $a_n := \operatorname{mean} X_n \in K$ and $Q_n := \operatorname{Cov} X_n \in \mathcal{L}_1^+(K)$ for all $n \in \mathbb{N}$, and assume that $X_n \to X$ in $L^2(\Omega, \Sigma, \mathbb{P}; K)$ as $n \to \infty$. Then X is a Gaussian random variable with law $N_{a,Q}$ where, for every $h, k \in K$,

$$\langle a, h \rangle_K = \lim_{n\to\infty} \langle a_n, h \rangle_K, \quad \langle Qh, k \rangle_K = \lim_{n\to\infty} \langle Q_n h, k \rangle_K.$$

Proof. Let $a := \operatorname{mean}(X)$ and $Q := \operatorname{Cov}(X)$. By dominated convergence,

$$\lim_{n\to\infty} \langle a_n, h \rangle = \lim_{n\to\infty} \int_\Omega \langle X_n(\omega), h \rangle \, \mathbb{P}(\mathrm{d}\omega) = \int_\Omega \langle X_n(\omega), h \rangle \, \mathbb{P}(\mathrm{d}\omega) = \langle a, h \rangle$$

for all $h \in K$, and

$$\lim_{n\to\infty} \langle Q_n h, k\rangle = \lim_{n\to\infty} \int_\Omega \langle X_n(\omega) - a_n, h\rangle \langle X_n(\omega) - a_n, k\rangle \, \mathbb{P}(\mathrm{d}\omega)$$
$$= \int_\Omega \langle X(\omega) - a, h\rangle \langle X(\omega) - a, k\rangle \, \mathbb{P}(\omega) = \langle Qh, k\rangle$$

for all $h, k \in K$.

To show that X is Gaussian, by Theorem C.13 it suffices to show that the Fourier transform of $X_\# \mathbb{P}$ is the Fourier transform of a Gaussian measure, *i.e.*,

$$\int_K \mathrm{e}^{\mathrm{i}\langle y,k\rangle} X_\# \mathbb{P}(\mathrm{d}y) = \mathrm{e}^{\mathrm{i}\langle a,k\rangle - \frac{1}{2}\langle Qk,k\rangle} \quad \forall k \in K.$$

This follows from

$$\int_K \mathrm{e}^{\mathrm{i}\langle y,k\rangle} X_\# \mathbb{P}(\mathrm{d}y) = \int_\Omega \mathrm{e}^{\mathrm{i}\langle X(\omega),k\rangle} \mathbb{P}(\mathrm{d}\omega) = \lim_{n\to\infty} \int_\Omega \mathrm{e}^{\mathrm{i}\langle X_n(\omega),k\rangle} \, \mathbb{P}(\mathrm{d}\omega)$$
$$= \lim_{n\to\infty} \mathrm{e}^{\mathrm{i}\langle a_n,k\rangle - \frac{1}{2}\langle Q_n k,k\rangle} = \mathrm{e}^{\mathrm{i}\langle a,k\rangle - \frac{1}{2}\langle Qk,k\rangle}.$$

Hence X is Gaussian, and $X_\# \mathbb{P} = \mathbb{N}_{a,Q}$. □

It is well known that linear transformations of Gaussian random variables are once again Gaussian. For random fields taking values in function spaces, this covariance takes the following form.

Theorem C.18. (affine transformations of Gaussians) Let $\mu = \mathbb{N}_{a,Q}$ be a Gaussian measure on a separable Hilbert space $(H, \mathcal{B}(H))$. Then

(a) For all $b \in H$, $T : H \to H$, $T(x) := x + b$ is Gaussian on $(H, \mathcal{B}(H))$, and

$$T_\# \mu = N_{a+b,Q}. \tag{C.13}$$

(b) If $T \in \mathcal{L}(H, K)$ for a separable Hilbert space K, T is Gaussian and

$$T_\# \mu = N_{Ta,TQT^*}. \tag{C.14}$$

Proof. By (C.1), for all $k \in K$,

$$\int_K \mathrm{e}^{\mathrm{i}\langle k,y\rangle} \, T_\# \mu(\mathrm{d}y) = \int_H \mathrm{e}^{\mathrm{i}\langle k,Tx\rangle} \, \mu(\mathrm{d}x)$$
$$= \int_H \mathrm{e}^{\mathrm{i}\langle T^* k,x\rangle} \, \mu(\mathrm{d}x) = \mathrm{e}^{\mathrm{i}\langle T^* k,a\rangle - \frac{1}{2}\langle TQT^* k,k\rangle}.$$

Then Theorem C.13 implies (C.14), and (C.13) follows similarly. □

Computation of some Gaussian integrals

Let H be a separable Hilbert space. We abbreviate

$$L^2(H, N_{a,Q}) := L^2(H, \mathcal{B}(H), N_{a,Q})$$

for any Gaussian measure $N_{a,Q}$ on $(H, \mathcal{B}(H))$.

Proposition C.19.

$$\int_H x\, N_{a,Q}(\mathrm{d}x) = a, \tag{C.15}$$

$$\int_H \langle x-a, y\rangle \langle x-a, z\rangle\, N_{a,Q}(\mathrm{d}x) = \langle Qy, z\rangle \quad \forall y, z \in H, \tag{C.16}$$

$$\int_H \|x-a\|_H^2\, N_{a,Q}(\mathrm{d}x) = \operatorname{Tr} Q = \sum_{k=1}^{\infty} \lambda_k. \tag{C.17}$$

Proof. Let $(e_k)_k$ be an orthonormal basis of H, and $P_n x := \sum_{k=1}^n \langle e_k, x\rangle e_k$. Then

$$\begin{aligned}\int_H x\, N_{a,Q}(\mathrm{d}x) &= \lim_{n\to\infty} \int_H P_n x\, N_{a,Q}(\mathrm{d}x)\\ &= \lim_{n\to\infty} \sum_{k=1}^{n} \left(\prod_{\ell=1}^{n} \int_{\mathbb{R}} x_k \lambda_\ell^{-1/2} \mathrm{e}^{-\frac{(x_\ell - a_\ell)}{2\lambda_\ell}}\, \mathrm{d}x_\ell \right) e_k = \sum_{k=1}^{\infty} a_k e_k = a.\end{aligned}$$

Equations (C.16) and (C.17) are proved analogously. □

C.3. Elliptic operator equations with Gaussian data

Elliptic operator equations

Let X, Y be separable Hilbert spaces and $A \in \mathcal{L}(X, Y')$, with associated bilinear form

$$a(u, v) = {}_{Y'}\langle Au, v\rangle_Y, \quad u \in X, \quad v \in Y. \tag{C.18}$$

Theorem C.20. Assume that the bilinear form $a(\cdot,\cdot)$ in (C.18) is

$$\text{continuous:} \qquad \forall u \in X,\ v \in Y: \quad |a(u,v)| \le C_1 \|u\|_X \|v\|_Y, \tag{C.19a}$$

$$\text{coercive:} \qquad \inf_{0\neq u\in X} \sup_{0\neq v\in Y} \frac{a(u,v)}{\|u\|_X \|v\|_Y} \ge C_2 > 0, \tag{C.19b}$$

$$\text{injective:} \qquad \forall v \in Y \setminus \{0\}: \quad \sup_{u\in X} |a(u,v)| > 0. \tag{C.19c}$$

Then, for every $f \in Y'$, the problem

$$Au = f, \tag{C.20a}$$

i.e.,

$$u \in X: \quad a(u,v) = f(v) = {}_{Y'}\langle f, v\rangle_Y \quad \forall v \in Ym, \tag{C.20b}$$

admits a unique solution $u \in X$, and

$$\|u\|_X \le \frac{\|f\|_{Y'}}{C_2}. \tag{C.21}$$

Conversely, $A \in \mathcal{L}(X, Y')$ is boundedly invertible if and only if (C.19) holds.

Example C.21. Let $D \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Consider the equation

$$A_{k^2}u := -\nabla \cdot \boldsymbol{A}(x)\nabla u - k^2 u = f \text{ in } D, \quad u|_{\partial D} = 0, \tag{C.22}$$

where $\boldsymbol{A} \in L^\infty(D, \mathbb{R}^{d\times d}_{\mathrm{sym}})$ is symmetric positive definite, *i.e.*, there is an $\alpha > 0$ such that, for all $\underline{\xi} \in \mathbb{R}^d$,

$$\operatorname*{ess\,inf}_{x\in D} \underline{\xi}^\top \boldsymbol{A}(x)\underline{\xi} \geq \alpha \|\underline{\xi}\|_2^2. \tag{C.23}$$

Let $0 < \mu_1 \leq \mu_2 < \mu_3 < \cdots$, $\mu_n \to \infty$, denote the eigenvalues of the Dirichlet problem, $\sigma = \{\mu_1, \mu_2, \ldots\} = (\mu_n)_{n=1}^\infty$, and denote by $(w_n)_n$ the corresponding sequence of eigenfunctions,

$$-\nabla \cdot \boldsymbol{A}\nabla w_n = \mu_n w_n \text{ in } D, \quad w_n|_{\partial D} = 0. \tag{C.24}$$

We assume that $(w_n)_n$ are normalized in $L^2(D)$; then

$$\langle w_m, w_n\rangle_{L^2(D)} = \delta_{mn}, \quad m, n = 1, 2, \ldots. \tag{C.25}$$

Claim C.22. For every value of k in (C.22) such that

$$k^2 \notin \sigma = \{\mu_1, \mu_2, \ldots\} \quad \text{(no resonance condition)}, \tag{C.26}$$

the bilinear form

$$a_k(u, v) := \langle \nabla v, \boldsymbol{A}(x)\nabla u\rangle_{L^2(D)} - k^2\langle u, v\rangle_{L^2(D)}$$

of (C.22) satisfies (C.19) with $X = Y = H_0^1(D)$ and

$$C_2 = \min_\ell \frac{|k^2 - \mu_\ell|}{\mu_\ell}.$$

Then, for every $f \in V' = H^{-1}(D)$, (C.22) has a unique solution $u \in H_0^1(D)$ and

$$\|u\|_{H_0^1(D)} \leq \frac{1}{\min_\ell \mu_\ell^{-1}|k^2 - \mu_\ell|}\, \|f\|_{H^{-1}(D)}. \tag{C.27}$$

Elliptic operator equations with Gaussian data

We consider now the *special case* $X = Y =: V$, $A \in \mathcal{L}(V, V')$ coercive, *i.e.*, there is a $C_2 > 0$ such that

$$a(v, v) \geq C_2\|v\|_V^2 \quad \forall v \in V. \tag{C.28}$$

Then we obtain (C.19c), and the following problem admits a unique solution: given $f \in V'$, find $u \in V$ such that

$$a(u, v) := {}_{V'}\langle Au, v\rangle_V = {}_{V'}\langle f, v\rangle_V \quad \forall v \in V. \tag{C.29}$$

Theorem C.23. Assume that $f \in L^2(\Omega, \Sigma, \mathbb{P}; V')$ is a Gaussian random field such that $a_f = \mathrm{mean}(f) \in V'$ and $Q_f = \mathrm{Cov}(f) \in \mathcal{L}_1^+(V')$ exist. Then

the following problem admits a unique solution: find $u \in L^2(\Omega, \Sigma, \mathbb{P}; V)$ such that

$$Au = f \quad \text{in} \quad L^2(\Omega, \Sigma, \mathbb{P}; V'). \tag{C.30}$$

Moreover,

$$a_u = \text{mean}(u) = A^{-1} a_f, \quad \text{and} \tag{C.31}$$

$$Q_u \in \mathcal{L}_1^+(V) \quad \text{satisfies} \quad AQ_uA^* = Q_f \quad \text{in} \quad \mathcal{L}(V'). \tag{C.32}$$

Proof. The weak form of (C.30) is

$$u \in L^2(\Omega, \Sigma, \mathbb{P}; V): \quad \tilde{a}(u, v) = \tilde{\ell}(v) \quad \forall v \in L^2(\Omega, \Sigma, \mathbb{P}; V'), \tag{C.33}$$

where

$$\tilde{a}(u, v) := \int_\Omega {}_V\langle v(\omega), Au(\omega)\rangle_{V'} \, \mathbb{P}(\mathrm{d}\omega), \quad \tilde{\ell}(v) := \int_\Omega {}_V\langle v(\omega), f(\omega)\rangle_{V'} \, \mathbb{P}(\mathrm{d}\omega).$$

Letting $\mathcal{V} = L^2(\Omega, \Sigma, \mathbb{P}; V)$, we infer from (C.28) that

$$\forall v \in \mathcal{V}: \quad \tilde{a}(v, v) \geq C_2 \|u\|_V^2,$$

hence (C.33) has a unique solution.

Since $A^{-1} \in \mathcal{L}(V', \mathcal{V})$ is 1 to 1 and onto, we get that

$$u(\omega) = A^{-1} f(\omega) \quad \mathbb{P}\text{-a.s.}$$

By Theorem C.18 with $T = A^{-1}$, u is Gaussian on V since $u = A^{-1}f$, and its distribution is the Gaussian measure

$$N_{A^{-1}a_f, A^{-1}Q_f(A^{-1})^*},$$

i.e., it is characterized completely by

$$a_u = \text{mean}(u) = A^{-1} a_f \quad \text{and} \quad Q_u = \text{Cov}(u) = A^{-1} Q_f (A^{-1})^*. \qquad \square$$

Remark C.24. For any separable Hilbert space H, by Theorem B.17,

$$L^2(\Omega, \Sigma, \mathbb{P}; H) \cong \underbrace{L^2(\Omega, \Sigma, \mathbb{P})}_{S} \otimes H.$$

C.4. Covariance kernels and the Karhunen–Loève expansion

From Theorem C.23, (C.32), we infer that, given a covariance operator $Q_f \in \mathcal{L}_1^+(V')$ on the data space V of the boundedly invertible operator $A \in \mathcal{L}(V, V')$, we have that if f is Gaussian, then $u = A^{-1}f$ is Gaussian on V, with mean a_u satisfying

$$Aa_u = a_f, \tag{C.34}$$

and covariance operator Q_u given by

$$AQ_uA^* = Q_f \in \mathcal{L}(V, V') \quad \text{resp.} \quad Q_u = A^{-1} Q_f (A^{-1})^* \in \mathcal{L}_1^+(V). \tag{C.35}$$

For the *computation* of u in terms of f, it suffices, by Theorem C.18, to compute a_u and Q_u in terms of a_f and Q_f as in (C.35). As we shall see, this is done most easily by means of *covariance kernel representations* of Q_f and Q_u.

Let $(H_1, \langle\cdot,\cdot\rangle_{H_1})$ and $(H_2, \langle\cdot,\cdot\rangle_{H_2})$ be separable Hilbert spaces, and let S denote a 'stochastic' space of random variables with finite second moments. For example, we have

$$L^2(\Omega, \Sigma, \mathbb{P}; H_i) \cong L^2(\Omega, \Sigma, \mathbb{P}) \otimes H_i = S \otimes H_i, \quad i = 1, 2, \ldots, \tag{C.36}$$

for $S = L^2(\Omega, \Sigma, \mathbb{P})$.

Let $(s_m)_{m\in\Lambda}$ be an orthonormal basis in S, with a countable index set Λ. Then any $f \in H_1 \otimes S$ can be uniquely represented as

$$f = \sum_{m\in\Lambda} f_m \otimes s_m \quad \text{in} \quad S \otimes H_1, \tag{C.37}$$

with $(f_m)_m \in \ell^2(\Lambda; H_1)$.

Proposition C.25. The mapping

$$S \otimes H_1 \times S \otimes H_2 \ni (f, g) \mapsto C_{fg} := \sum_{m\in\Lambda} f_m \otimes g_m \in H_1 \otimes H_2$$

is well-defined, bilinear, bounded with norm 1, and independent of the choice of basis $(s_m)_m$ of the stochastic space S.

Definition C.26. (correlation kernel) For $f \in S \otimes H_1$, $g \in S \otimes H_2$, we call $C_{fg} \in H_1 \otimes H_2$ defined in Proposition C.25 the *correlation kernel* of the pair (f, g) in $H_1 \times H_2$.

If $H_1 = H_2 = H$, the set $\{C_f := C_{fg}\,;\ f \in S \otimes H\}$ of *auto-correlation kernels* is in one-to-one correspondence with the class $\mathcal{L}_1^+(H)$ of positive definite trace-class operators.

Theorem C.27. If $(H, \langle\cdot,\cdot\rangle_H)$ and $(S, \langle\cdot,\cdot\rangle_S)$ are separable Hilbert spaces of equal dimension, and $(s_m)_{m\in\Lambda}$ is an orthonormal basis of S, then the auto-correlation kernels of elements f in $S \otimes H$ are in one-to-one correspondence with the positive definite trace-class operators on H, via the correspondence

$$\sum_{m\in\Lambda} f_m \otimes f_m = C_f \mapsto \mathcal{C}_f : H \ni x \mapsto \sum_{m\in\Lambda} \langle x, f_m\rangle_H f_m, \tag{C.38}$$

where

$$S \otimes H \ni f = \sum_{m\in\Lambda} f_m \otimes s_m, \quad f_m \in H, \quad m \in \Lambda. \tag{C.39}$$

Proof. The operator $\mathcal{C}_f$ defined in (C.38) is the $\mathcal{L}_1^+(H)$-norm limit of the finite rank operators $\mathcal{C}_f^n := \sum_{m\in\Lambda_n} \langle x, f_m\rangle_H f_m$, for some sequence $(\Lambda_n)_{n=1}^\infty$

of subsets $\Lambda_n \subset \Lambda$ such that $\#\Lambda_n = n$, since

$$\operatorname{Tr} \mathcal{C}_f = \sum_m \langle \mathcal{C}_f e_m, e_m \rangle_H = \sum_m \sum_n |\langle f_m, e_n \rangle_H|^2 = \sum_m \|f_m\|_H^2 = \|f\|_{S\otimes H}^2 < \infty,$$

for any orthonormal basis $(e_m)_m$ of H. Non-negative definiteness of $\mathcal{C}_f$ is obvious.

Since

$$\forall x, y \in H: \quad \langle \mathcal{C}_f x, y \rangle_H = \langle C_f, x \otimes y \rangle_{H\otimes H}, \tag{C.40}$$

the definition (C.38) of $\mathcal{C}_f$ is independent of the choice of bases in S and H. Hence the map (C.38) is well-defined. The mapping (C.38) from covariance kernels to covariance operators, *i.e.*, the correspondence $C_f \mapsto \mathcal{C}_f$, is injective.

To see that $C_f \mapsto \mathcal{C}_f$ is also surjective, we let $\mathcal{C} \in \mathcal{L}_1^+(H)$ be given. Then $\mathcal{C}$ is compact and has a countable eigensequence $(\lambda_m, \varphi_m)_{m\in\mathbb{N}}$, such that

$$\mathcal{C}\varphi_m = \lambda_m \varphi_m, \quad m \in \mathbb{N}, \quad \langle \varphi_m, \varphi_n \rangle_H = \delta_{mn}. \tag{C.41}$$

The eigenvalues $\lambda_m \in \mathbb{R}$ have finite multiplicity, and we assume they are ordered decreasingly, *i.e.*, $\lambda_1 \geq \lambda_2 \geq \cdots$, and they accumulate only in 0. Then, since $\mathcal{C}$ is of trace class,

$$\sum_m \lambda_m < \infty. \tag{C.42}$$

The series $\sum_m \sqrt{\lambda_m}\, \varphi_m \otimes s_m$ converges, by (C.42), to some $f \in S \otimes H$ such that

$$C_f = \sum_m \lambda_m \varphi_m \otimes \varphi_m \quad \text{in} \quad H \otimes H. \tag{C.43}$$

(C.40), (C.41) and (C.43) imply that $\mathcal{C}$ has the same spectral decomposition as $\mathcal{C}_f$, hence $\mathcal{C} = \mathcal{C}_f$. □

Corollary C.28. Let $(H, \langle\cdot,\cdot\rangle_H)$ be a separable Hilbert space and let $C \in H \otimes H$ be a correlation kernel. Then, with the spectrum (C.8) of its operator $\mathcal{C} \in \mathcal{L}_1^+(H)$ defined as in (C.40), the corresponding *covariance kernel* C can be represented as

$$C = \sum_m \lambda_m \varphi_m \otimes \varphi_m \quad \text{in} \quad H \otimes H. \tag{C.44}$$

Theorem C.29. Let $(H, \langle\cdot,\cdot\rangle_H)$ and $(S, \langle\cdot,\cdot\rangle_S)$ be separable Hilbert spaces, and let $C \in H \otimes H$ be a correlation kernel with representation (C.44). Then $f \in S \otimes H$ satisfies $C_f = C$ in $H \otimes H$ if and only if there exists an S-orthonormal family $(X_m)_m \subset S$ such that

$$f = \sum_m \sqrt{\lambda_m}\, X_m \otimes \varphi_m \quad \text{in} \quad S \otimes H. \tag{C.45}$$

Proof. The 'if' part follows as in the proof of Theorem C.27, upon completion of the family $(X_m)_m \subset S$ to an orthonormal basis of S. Conversely, if $C_f = C$, then we may write

$$f = \sum_m Y_m \otimes \varphi_m \quad \text{in} \quad S \otimes H,$$

with $(Y_m)_m \subset S$, which implies with Proposition C.25 that

$$C_f = \sum_{m,m'} \langle Y_{m'}, Y_m \rangle_S \varphi_m \otimes \varphi_{m'}.$$

Comparing this with (C.44), we find (since $(\varphi_m)_m$ is an orthonormal basis of H) that

$$\langle Y_m, Y_{m'} \rangle_S = \lambda_m \delta_{mm'}, \quad m, m' \in \mathbb{N}.$$

This is (C.44) with $X_m := \lambda_m^{-1/2} Y_m$. □

Definition C.30. The expansion (C.45) of $f \in S \otimes H$ in terms of the spectral decomposition of its covariance operator $\mathcal{C}_f$ is called *Karhunen–Loève expansion* of f.

Theorem C.31. Let X, Y be separable Hilbert spaces. Let $A \in \mathcal{L}(X, Y')$ be boundedly invertible (see Theorem C.20), and let $f \in L^2(\Omega, \Sigma, \mathbb{P}; Y') \cong L^2(\Omega, \Sigma, \mathbb{P}) \otimes Y'$ be a given Gaussian random field on Y', with

$$\operatorname{mean}(f) = a_f \in Y', \quad Q_f = \operatorname{Cov}(f) \in \mathcal{L}_1^+(Y'). \tag{C.46}$$

Then $u = A^{-1} f \in L^2(\Omega, \Sigma, \mathbb{P}; X)$ is also Gaussian on X, with

$$\operatorname{mean}(u) = a_u \in X, \quad Q_u = \operatorname{Cov}(u) \in \mathcal{L}_1^+(X), \tag{C.47}$$

satisfying

$$A a_u = a_f \quad \text{in} \quad Y', \tag{C.48}$$

and

$$A Q_u A^* = Q_f \quad \text{in} \quad \mathcal{L}_1^+(Y'), \quad Q_u = A^{-1} Q_f (A^{-1})^* \in \mathcal{L}_1^+(X). \tag{C.49}$$

The kernels C_u of Q_u, resp. C_f of Q_f, satisfy the equation

$$(A \otimes A) C_u = C_f \quad \text{in} \quad Y' \otimes Y' \cong (Y \otimes Y)'. \tag{C.50}$$

C.5. The white noise map

Let H be a separable Hilbert space, $\dim H = \infty$, and $\mu = N_Q$ a *non-degenerate* centred Gaussian measure on H (*i.e.*, $\ker Q = \{0\} \subset H$); furthermore, let $(e_k)_k$ be an orthonormal basis of H such that $Q e_k = \lambda_k e_k$, $k \in \mathbb{N}$. For $x \in H$, set $x_k := \langle x, e_k \rangle$. Then, for all k, $Q^{-1} e_k = \lambda_k^{-1} e_k$,

hence $Qx_0 \in H^\perp$, so $Qx_0 \in \mathcal{L}(H)$ is not boundedly invertible (since $\lambda_k \to 0$ as $k \to \infty$ due to $\operatorname{Tr} Q < \infty$). The range $Q(H)$ of Q does not equal H, $Q(H) \neq H$.

Lemma C.32. *$Q(H)$ is a dense subspace of H.*

Proof. Let $x_0 \in Q(H)^\perp \subset H$. Then, since Q is self-adjoint,

$$\forall x \in H: \quad 0 = \langle x_0, Qx\rangle_H = \langle Qx_0, x\rangle_H,$$

hence $Qx_0 = 0$. But $\ker Q = \{0\}$, so $x_0 = 0$. □

Define the operator $Q^{1/2}$ by

$$Q^{1/2}x := \sum_{k=1}^{\infty} \sqrt{\lambda_k}\, \langle x, e_k\rangle_H, \quad x \in H. \tag{C.51}$$

Its range $Q^{1/2}(H)$ is the *reproducing kernel Hilbert space* or the *Cameron–Martin space* of the measure $\mu = N_Q$ in H. It is a dense subspace of H, and $H \neq Q^{1/2}(H)$.

We introduce an isometry $W : H \to L^2(H, N_Q)$, the *white noise map*. Let $Q^{1/2}(H) \ni f \mapsto W_f \in L^2(H, N_Q)$ be given by

$$W_f(x) = \langle Q^{-1/2}f, x\rangle_H \quad \forall x \in H. \tag{C.52}$$

We have

$$\int_H W_f(x)W_g(x)\, N_Q(\mathrm{d}x) = \langle f, g\rangle_H \quad \forall f, g \in H. \tag{C.53}$$

This map $W : Q^{1/2}(H) \to L^2(H, N_Q)$ is a densely defined isometry which can be extended to all of H.

Lemma C.33. *For any $f \in H$, W_f is a real Gaussian random variable with mean zero and variance $\|f\|_H^2$.*

Proof. Define $\nu_f := (W_f)_{\#}\mu$. We must show

$$\forall \eta \in \mathbb{R} \quad \widehat{\nu}_f(\eta) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\xi\eta}\nu_f(\mathrm{d}\xi) = \int_H \mathrm{e}^{\mathrm{i}\eta W_f(x)}\mu(\mathrm{d}x) = \mathrm{e}^{-\frac{1}{2}\eta^2\|f\|_H^2}.$$

To this end, let $(z_n)_n \subset Q^{1/2}(H)$ be a sequence such that $z_n \to z \in H$. By dominated convergence,

$$\begin{aligned}\int_H \mathrm{e}^{\mathrm{i}\eta W_z(x)}\mu(\mathrm{d}x) &= \lim_{n\to\infty} \int_H \mathrm{e}^{\mathrm{i}\eta\langle Q^{-1/2}z_n, x\rangle}\mu(\mathrm{d}x) \\ &= \lim_{n\to\infty} \mathrm{e}^{-\frac{1}{2}\eta^2\|z_n\|_H^2} = \mathrm{e}^{-\frac{1}{2}\eta^2\|z\|_H^2}.\end{aligned}$$

We pick $z = f$ to conclude. □

Remark C.34. Given $z \in H \setminus Q^{1/2}(H)$, one could try to define W_z by

$$\forall x \in Q^{1/2}(H): \quad W_z(x) = \langle Q^{-1/2}x, z\rangle_H,$$

rather than by (C.52). This is meaningless due to $\mu(Q^{1/2}(H)) = 0$.

Lemma C.35. $\mu(Q^{1/2}(H)) = 0$.

Proof. For all $n, k \in \mathbb{N}$, set

$$U_n := \left\{ y \in H\,;\, \sum_{\ell=1}^{\infty} \lambda_\ell^{-1} Y_\ell^2 < n^2 \right\}, \quad U_{n,k} := \left\{ y \in H\,;\, \sum_{\ell=1}^{2k} \lambda_\ell^{-1} Y_\ell^2 < n^2 \right\},$$

Then $U_n \uparrow Q^{1/2}(H)$, as $n \to \infty$, and, for every fixed $n \in \mathbb{N}$, $U_{n,k} \downarrow U_n$ as $k \to \infty$. Hence we are done, if

$$\forall n: \quad \mu(U_n) = \lim_{k\to\infty} \mu(U_{n,k}) = 0. \tag{C.54}$$

To see (C.54), we use that for all $n, k \in \mathbb{N}$, for $z_\ell := \lambda_\ell^{-1/2} Y_\ell$,

$$\mu(U_{n,k}) = \int_{U_{n,k}} \bigotimes_{\ell=1}^{2k} N_{\lambda_k}(\mathrm{d}y_k) = \int_{\{z\in\mathbb{R}^{2k}\,;\,|z|<n\}} N_{\mathbb{I}_{R^{2k}}}(\mathrm{d}x).$$

We compute

$$\mu(U_{n,k}) = \frac{\mu(U_{n,k})}{\mu(H)} = \frac{\int_0^n \mathrm{e}^{-\frac{r^2}{2}} r^{2k-1}\,\mathrm{d}r}{\int_0^\infty \mathrm{e}^{-\frac{r^2}{2}} r^{2k-1}\,\mathrm{d}r} = \frac{\int_0^{n^2/2} \mathrm{e}^{-\varrho}\varrho^{k-1}\,\mathrm{d}\varrho}{\int_0^\infty \mathrm{e}^{-\varrho}\varrho^{k-1}\,\mathrm{d}\varrho}.$$

Hence,

$$\mu(U_{nk}) = \frac{1}{(k-1)!}\int_0^{n^2/2} \mathrm{e}^{-\varrho}\varrho^{k-1}\,\mathrm{d}\varrho \le \frac{1}{(k-1)!}\int_0^{n^2/2} \varrho^{k-1}\,\mathrm{d}\varrho = \frac{1}{k!}\left(\frac{n^2}{2}\right)^k,$$

whence (C.54). □

Proposition C.36. (properties of the white noise map)

(a) For any $n \in \mathbb{N}$, $z_1, \dots, z_n \in H$, the law of $(W_{z_1}, \dots, W_{z_n}) \in \mathbb{R}^n$ is given by $N_{(\langle z_i, z_j\rangle_H)_{i,j=1}^n}$.

(b) $W_{z_1}, \dots W_{z_n}$ are independent if and only if $z_1, \dots z_n$ are H-orthogonal.

(c) For all $f \in H$,

$$\int_H \mathrm{e}^{W_f(x)}\mu(\mathrm{d}x) = \mathrm{e}^{\frac{1}{2}\|f\|_H^2} \tag{C.55}$$

and

$$\int_H \mathrm{e}^{\mathrm{i}\lambda W_f(x)}\mu(\mathrm{d}x) = \mathrm{e}^{-\frac{1}{2}\lambda^2\|f\|_H^2}. \tag{C.56}$$

(d) The exponential map $H \to L^2(H, N_Q)$, $f \mapsto \mathrm{e}^{W_f}$ is continuous, and

$$\begin{aligned}\int_H \left(\mathrm{e}^{W_f(x)} - \mathrm{e}^{W_g(x)}\right)^2 N_Q(\mathrm{d}x) &= \int_H \left(\mathrm{e}^{2W_f} - 2\mathrm{e}^{W_{f+g}} + \mathrm{e}^{2W_g}\right) N_Q(\mathrm{d}x)\\ &= \mathrm{e}^{2\|f\|_H^2} - 2\mathrm{e}^{\frac{1}{2}\|f+g\|_H^2} + \mathrm{e}^{2\|g\|_H^2}\\ &= \left(\mathrm{e}^{\|f\|_H^2} - \mathrm{e}^{\|g\|_H^2}\right) + 2\mathrm{e}^{\|f\|_H^2+\|g\|_H^2}\left(1 - \mathrm{e}^{-\frac{1}{2}\|f-g\|_H^2}\right).\end{aligned}$$

Proposition C.37. Assume that $M \in \mathcal{L}(H)$ is symmetric such that, for given $Q \in \mathcal{L}_1^+(H)$,

$$Q^{1/2}MQ^{1/2} \le 1 \quad (\iff \forall x \in H : \langle Q^{1/2}MQ^{1/2}x, x\rangle_H \le \|X\|_H^2\ ),$$

and let $b \in H$ be arbitrary. Then

$$\begin{aligned}&\int_H \exp\left(\frac{1}{2}\langle My, y\rangle + \langle b, y\rangle\right) N_Q(\mathrm{d}y) \qquad (C.57)\\ &\quad = \left[\det(1 - Q^{1/2}MQ^{1/2})\right]^{-1/2} \exp\left(\frac{1}{2}|(1 - Q^{1/2}MQ^{1/2})^{-1/2}Q^{1/2}b|^2\right).\end{aligned}$$

C.6. Absolute continuity of Gaussian measures

Let H be a separable Hilbert space, $\dim H = \infty$, $\mu = N_Q$ a non-degenerate centred Gaussian measure on H with covariance operator $Q \in \mathcal{L}_1^+(H)$, $\ker Q = \{0\}$, $(e_k)_k$ orthonormal basis of H with $Qe_k = \lambda_k e_k$, $k = 1, 2, \ldots$ (see Theorem B.23).

Given $a \in H$, when are the two Gaussian measures N_Q and $N_{a,Q}$ singular, resp. equivalent?

Recall that two measures μ, ν on (Ω, Σ) are *equivalent* if $\mu \ll \nu$ and $\nu \ll \mu$. Here, $\mu \ll \nu$ ('μ is absolutely continuous with respect to ν') if, for all $A \in \Sigma$ with $\nu(A) = 0$, also $\mu(A) = 0$.

If $\mu \ll \nu$, the Radon–Nikodym theorem implies that there exists a unique $\varrho \in L^1(\Omega, \Sigma, \nu)$ such that

$$\forall A \in \Sigma : \quad \mu(A) = \int_A \varrho \,\mathrm{d}\nu.$$

Assume for the moment $\dim H < \infty$. Then $\ker Q = \{0\} \subset H$ implies $\det Q > 0$. Hence, for all $a \in H$, $N_Q \sim N_{a,Q}$ and for all $x \in H$,

$$\frac{\mathrm{d}N_{a,Q}}{\mathrm{d}N_Q}(x) = \frac{\mathrm{e}^{-\frac{1}{2}\langle Q^{-1}(x-a), x-a\rangle_H}}{\mathrm{e}^{-\frac{1}{2}\langle Q^{-1}x, x\rangle_H}} = \mathrm{e}^{-\frac{1}{2}\|Q^{-1/2}a\|_H^2 + \langle Q^{-1/2}a, Q^{-1/2}x\rangle_H}. \quad (C.58)$$

We claim:

(1) if $a \in Q^{1/2}(H)$, then $N_{a,Q} \sim N_Q$,

(2) if $a \in H \cap Q^{1/2}(H)^\perp$, then $N_{a,Q} \perp N_Q$.

In the first case, (C.58) still holds if $\dim H = \infty$, but $\langle Q^{-1/2}a, Q^{-1/2}x\rangle_H$ is replaced by $W_{Q^{-1/2}a}(x)$.

Hellinger integral

Let μ and ν be probability measures on (Ω, Σ). Then both μ and ν are absolutely continuous with respect to the probability measure $\zeta = (\mu+\nu)/2$ on (Ω, Σ).

Definition C.38. (Hellinger integral) The Hellinger integral of μ, ν is defined by

$$H(\mu,\nu) := \int_\Omega \sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\zeta}\,\frac{\mathrm{d}\nu}{\mathrm{d}\zeta}}\,\zeta(\mathrm{d}\omega). \tag{C.59}$$

Obviously, $0 \le H(\mu,\nu) \le 1$. By Hölder's inequality, we have

$$0 \le H(\mu,\nu) \le \left(\int_\Omega \frac{\mathrm{d}\mu}{\mathrm{d}\zeta}\,\mathrm{d}\zeta\right)^{1/2}\left(\int_\Omega \frac{\mathrm{d}\nu}{\mathrm{d}\zeta}\,\mathrm{d}\zeta\right)^{1/2} = 1.$$

Remark C.39. If λ is a probability measure on (Ω, Σ) such that $\mu \ll \lambda$ and $\nu \ll \lambda$, then also $\zeta \ll \lambda$ and

$$\frac{\mathrm{d}\mu}{\mathrm{d}\zeta} = \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\,\frac{\mathrm{d}\lambda}{\mathrm{d}\zeta} \quad \wedge \quad \frac{\mathrm{d}\nu}{\mathrm{d}\zeta} = \frac{\mathrm{d}\nu}{\mathrm{d}\lambda}\,\frac{\mathrm{d}\lambda}{\mathrm{d}\zeta},$$

and we find

$$H(\mu,\nu) = \int_\Omega \sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\,\frac{\mathrm{d}\nu}{\mathrm{d}\lambda}}\,\mathrm{d}\lambda.$$

Remark C.40. Assume $\mu \sim \nu$. Then

$$\frac{\mathrm{d}\mu}{\mathrm{d}\zeta} = \frac{\mathrm{d}\nu}{\mathrm{d}\zeta} = \frac{\mathrm{d}\mu}{\mathrm{d}\zeta}\,\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\,\frac{\mathrm{d}\mu}{\mathrm{d}\zeta} = \left(\frac{\mathrm{d}\mu}{\mathrm{d}\zeta}\right)^2\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$$

and hence

$$H(\mu,\nu) = \int_\Omega \sqrt{\frac{\mathrm{d}\nu}{\mathrm{d}\mu}}\,\frac{\mathrm{d}\mu}{\mathrm{d}\zeta}\,\mathrm{d}\zeta = \int_\Omega \sqrt{\frac{\mathrm{d}\nu}{\mathrm{d}\mu}}\,\mathrm{d}\mu.$$

Example C.41. Let $\Omega = \mathbb{R}$, $\mu = N_\lambda$, $\nu = N_{a,\lambda}$, $a \in \mathbb{R}$, and $\lambda > 0$. Then

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) = \mathrm{e}^{-\frac{a^2}{2\lambda}+\frac{ax}{\lambda}}, \quad x \in \mathbb{R},$$

and hence

$$H(\mu,\nu) = \mathrm{e}^{-\frac{a^2}{2\lambda}}\int_{\mathbb{R}} \mathrm{e}^{\frac{ax}{2\lambda}}N_\lambda(\mathrm{d}x) = \mathrm{e}^{-\frac{a^2}{8\lambda}}.$$

Proposition C.42. If $H(\mu,\nu) = 0$, the μ and ν are singular.

Proof. Let

$$f = \frac{\mathrm{d}\mu}{\mathrm{d}\zeta}, \quad g = \frac{\mathrm{d}\nu}{\mathrm{d}\zeta}, \quad \zeta = \frac{1}{2}(\mu + \nu).$$

Then $fg = 0$ ζ-a.e., since

$$H(\mu, \nu) = \int_\Omega \sqrt{fg}\, \mathrm{d}\zeta = 0.$$

Define the sets

$$A = \{\omega \in \Omega\,;\ f(\omega) = 0\}, \quad B = \{\omega \in \Omega\,;\ g(\omega) = 0\},$$
$$C = \{\omega \in \Omega\,;\ (fg)(\omega) = 0\}.$$

Then $\zeta(C) = 1$, hence $\mu(C) = \nu(C) = 1$. Moreover, $\mu(A) = \int_A f\, \mathrm{d}\zeta = 0$ and $\nu(B) = \int_B g\, \mathrm{d}\zeta = 0$. Therefore, $\mu(B \setminus A) = 1$ and $\nu(A \setminus B) = 1$, *i.e.*, μ and ν are mutually singular. □

Kakutani's theorem

If $H(\mu, \nu) = 0$, μ and ν are mutually singular. Conversely, if $H(\mu, \nu) > 0$, then μ and ν are not necessarily equivalent, in general, *unless* μ, ν are countable products of equivalent 'factor measures'. This is Kakutani's theorem. We prepare its exposition with products of two measures.

Lemma C.43. Let μ_i, ν_i, $i = 1, 2$, be probability measures on (Ω, Σ). Then

$$H(\mu_1 \otimes \mu_2, \nu_1 \otimes \nu_2) = H(\mu_1, \nu_1) H(\mu_2, \nu_2).$$

Proof. Let ζ_1, ζ_2 be probability measures on (Ω, Σ) such that

$$\mu_1 \ll \zeta_1, \quad \nu_1 \ll \zeta_1, \quad \mu_2 \ll \zeta_2, \quad \nu_2 \ll \zeta_2.$$

Then, by Fubini's theorem,

$$\mu_1 \otimes \mu_2 \ll \zeta_1 \otimes \zeta_2 \quad \wedge \quad \nu_1 \otimes \nu_2 \ll \zeta_1 \otimes \zeta_2.$$

Define

$$f_i(\omega_i) := \frac{\mathrm{d}\mu_i}{\mathrm{d}\zeta_i}(\omega_i), \quad g_i(\omega_i) := \frac{\mathrm{d}\nu_i}{\mathrm{d}\zeta_i}(\omega_i), \quad i = 1, 2.$$

Then

$$\frac{\mathrm{d}(\mu_1 \otimes \mu_2)}{\mathrm{d}(\zeta_1 \otimes \zeta_2)} = f_1(\omega_1) f_2(\omega_2), \quad \frac{\mathrm{d}(\nu_1 \otimes \nu_2)}{\mathrm{d}(\zeta_1 \otimes \zeta_2)} = g_1(\omega_1) g_2(\omega_2).$$

Hence

$$\begin{aligned} H(\mu_1 \otimes \mu_2, \nu_1 \otimes \nu_2) &= \int_{\Omega \times \Omega} \big((f_1, g_1)(\omega_1)(f_2, g_2)(\omega_2)\big)^{1/2} \zeta_1(\mathrm{d}\omega_1) \zeta_2(\mathrm{d}\omega_2) \\ &= H(\mu_1, \nu_1) H(\mu_2, \nu_2). \end{aligned}$$

□

Kakutani's theorem is an infinite-dimensional generalization of the previous result.

Theorem C.44. (Kakutani) Let $(\mu_k)_k$, $(\nu_k)_k$ be sequences of probability measures on $(\mathbb{R}, \Sigma)$, such that $\mu_k \sim \nu_k$ for all $k \in \mathbb{N}$, and define

$$\mu := \bigotimes_{k=1}^{\infty} \mu_k, \quad \nu := \bigotimes_{k=1}^{\infty} \nu_k.$$

If $H(\mu, \nu) > 0$, then $\mu \sim \nu$ and

$$\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) = \lim_{n\to\infty} \prod_{k=1}^{n} \frac{\mathrm{d}\nu_k}{\mathrm{d}\mu_k}(x_k) \quad \text{in} \quad L^1(\mathbb{R}^\infty, \mu). \tag{C.60}$$

If $H(\mu, \nu) = 0$, then μ and ν are singular.

We refer to Kakutani (1948) and Da Prato (2006) for a proof of Theorem C.44.

REFERENCES

I. Babuška (1961), 'On randomised solutions of Laplace's equation', *Časopis Pěst. Mat.* **86**, 269–276.

I. Babuška (1970/71), 'Error-bounds for finite element method', *Numer. Math.* **16**, 322–333.

I. Babuška, F. Nobile and R. Tempone (2005), 'Worst case scenario analysis for elliptic problems with uncertainty', *Numer. Math.* **101**, 185–219.

I. Babuška, F. Nobile and R. Tempone (2007*a*), 'Reliability of computational science', *Numer. Methods Partial Differential Equations* **23**, 753–784.

I. Babuška, F. Nobile and R. Tempone (2007*b*), 'A stochastic collocation method for elliptic partial differential equations with random input data', *SIAM J. Numer. Anal.* **45**, 1005–1034.

A. Barth (2010), 'A finite element method for martingale-driven stochastic partial differential equations', *Comm. Stoch. Anal.* **4**, 355–375.

A. Barth and A. Lang (2009), Almost sure convergence of a Galerkin–Milstein approximation for stochastic partial differential equations. In review.

A. Barth, C. Schwab and N. Zollinger (2010), Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients. Technical Report 2010-18, Seminar for Applied Mathematics, ETH Zürich. To appear in *Numer. Math.*

H. Bauer (1996), *Probability Theory*, Vol. 23 of De Gruyter Studies in Mathematics, Walter de Gruyter. Translation by R. B. Burckel.

H. Bauer (2001), *Measure and Integration Theory*, Vol. 26 of De Gruyter Studies in Mathematics, Walter de Gruyter. Translation by R. B. Burckel.

R. Beatson and L. Greengard (1997), A short course on fast multipole methods, in *Wavelets, Multilevel Methods and Elliptic PDEs* (Leicester 1996), Numerical Mathematics and Scientific Computation, Oxford University Press, pp. 1–37.

M. Bebendorf and W. Hackbusch (2003), 'Existence of $\mathcal{H}$-matrix approximants to the inverse FE-matrix of elliptic operators with L^∞-coefficients', *Numer. Math.* **95**, 1–28.

M. Bieri (2009*a*), A sparse composite collocation finite element method for elliptic sPDEs. Technical Report 2009-8, Seminar for Applied Mathematics, ETH Zürich. To appear in *SIAM J. Numer. Anal.*

M. Bieri (2009*b*), Sparse tensor discretization of elliptic PDEs with random input data. PhD thesis, ETH Zürich. ETH Dissertation no. 18598.

M. Bieri, R. Andreev and C. Schwab (2009), 'Sparse tensor discretization of elliptic SPDEs', *SIAM J. Sci. Comput.* **31**, 4281–4304.

V. I. Bogachev (1998), *Gaussian Measures*, Vol. 62 of Mathematical Surveys and Monographs, AMS, Providence, RI.

V. I. Bogachev (2007), *Measure Theory*, Vols I and II, Springer.

D. Braess (2007), *Finite Elements: Theory, Fast Solvers, and Applications in Elasticity Theory*, third edition, Cambridge University Press. Translation by L. L. Schumaker.

S. C. Brenner and L. R. Scott (2002), *The Mathematical Theory of Finite Element Methods*, Vol. 15 of Texts in Applied Mathematics, second edition, Springer.

H.-J. Bungartz and M. Griebel (2004), Sparse grids. In *Acta Numerica*, Vol. 13, Cambridge University Press, pp. 147–269.

R. H. Cameron and W. T. Martin (1947), 'The orthogonal development of non-linear functionals in series of Fourier–Hermite functionals', *Ann. of Math.* (2) **48**, 385–392.

E. R. Canfield, P. Erdős and C. Pomerance (1983), 'On a problem of Oppenheim concerning "factorisatio numerorum"', *J. Number Theory* **17**, 1–28.

J. Charrier (2010), Strong and weak error estimates for the solutions of elliptic partial differential equations with random coefficients. Technical Report 7300, INRIA.

A. Chernov and C. Schwab (2009), 'Sparse p-version BEM for first kind boundary integral equations with random loading', *Appl. Numer. Math.* **59**, 2698–2712.

O. Christensen (2008), *Frames and Bases: An Introductory Course*, Applied and Numerical Harmonic Analysis, Birkhäuser.

O. Christensen (2010), *Functions, Spaces, and Expansions: Mathematical Tools in Physics and Engineering*, Applied and Numerical Harmonic Analysis, Birkhäuser.

P. G. Ciarlet (1978), *The Finite Element Method for Elliptic Problems*, Vol. 4 of Studies in Mathematics and its Applications, North-Holland.

A. Cohen (2003), *Numerical Analysis of Wavelet Methods*, Vol. 32 of Studies in Mathematics and its Applications, North-Holland.

A. Cohen, R. A. DeVore and C. Schwab (2010), Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs. *J. Found. Comput. Math.* **10**, 615–646.

A. Cohen, R. A. DeVore and C. Schwab (2011), Analytic regularity and polynomial approximation of parametric stochastic elliptic PDEs. *Anal. Appl.* **9**, 1–37.

G. Da Prato (2006), *An Introduction to Infinite-Dimensional Analysis*, revised and extended edition, Universitext, Springer.

G. Da Prato and J. Zabczyk (1992), *Stochastic Equations in Infinite Dimensions*, Vol. 44 of Encyclopedia of Mathematics and its Applications, Cambridge University Press.

W. Dahmen (1997), Wavelet and multiscale methods for operator equations. In *Acta Numerica*, Vol. 6, Cambridge University Press, pp. 55–228.

W. Dahmen, H. Harbrecht and R. Schneider (2006), 'Compression techniques for boundary integral equations: Asymptotically optimal complexity estimates', *SIAM J. Numer. Anal.* **43**, 2251–2271.

R. Dalang, D. Khoshnevisan, C. Mueller, D. Nualart and Y. Xiao (2009), *A Minicourse on Stochastic Partial Differential Equations* (Salt Lake City 2006; D. Khoshnevisan and F. Rassoul-Agha, eds), Vol. 1962 of Lecture Notes in Mathematics, Springer.

M. Dettinger and J. L. Wilson (1981), 'First order analysis of uncertainty in numerical models of groundwater flow 1: Mathematical development', *Water Res. Res.* **17**, 149–161.

R. A. DeVore (1998), Nonlinear approximation. In *Acta Numerica*, Vol. 7, Cambridge University Press, pp. 51–150.

G. C. Donovan, J. S. Geronimo and D. P. Hardin (1996), 'Intertwining multiresolution analyses and the construction of piecewise-polynomial wavelets', *SIAM J. Math. Anal.* **27**, 1791–1815.

S. C. Eisenstat, H. C. Elman and M. H. Schultz (1983), 'Variational iterative methods for nonsymmetric systems of linear equations', *SIAM J. Numer. Anal.* **20**, 345–357.

O. G. Ernst, A. Mugler, H.-J. Starkloff and E. Ullmann (2010), On the convergence of generalized polynomial chaos expansions. Technical Report 60, DFG Schwerpunktprogramm 1324.

G. S. Fishman (1996), *Monte Carlo: Concepts, Algorithms, and Applications*, Springer Series in Operations Research, Springer.

J. Galvis and M. Sarkis (2009), 'Approximating infinity-dimensional stochastic Darcy's equations without uniform ellipticity', *SIAM J. Numer. Anal.* **47**, 3624–3651.

W. Gautschi (2004), *Orthogonal Polynomials: Computation and Approximation*, Numerical Mathematics and Scientific Computation, Oxford Science Publications, Oxford University Press.

M. Geissert, M. Kovács and S. Larsson (2009), 'Rate of weak convergence of the finite element method for the stochastic heat equation with additive noise', *BIT* **49**, 343–356.

R. G. Ghanem and P. D. Spanos (2007), *Stochastic Finite Elements: A Spectral Approach*, second edition, Dover,

C. J. Gittelson (2010*a*), 'Stochastic Galerkin discretization of the log-normal isotropic diffusion problem', *Math. Models Methods Appl. Sci.* **20**, 237–263.

C. J. Gittelson (2010*b*), Representation of Gaussian fields in series with independent coefficients. Technical Report 2010-15, Seminar for Applied Mathematics, ETH Zürich. Submitted.

C. J. Gittelson (2011*a*) Adaptive Galerkin methods for parametric and stochastic operator equations. ETH Dissertation No. 19533, ETH Zürich.

C. J. Gittelson (2011*b*) Stochastic Galerkin approximation of operator equations with infinite dimensional noise. Technical Report 2011-10, Seminar for Applied Mathematics, ETH Zürich.

C. J. Gittelson (2011*c*) An adaptive stochastic Galerkin method. Technical Report 2011-11, Seminar for Applied Mathematics, ETH Zürich.

C. J. Gittelson (2011*d*) Adaptive stochastic Galerkin methods: Beyond the elliptic case. Technical Report 2011-12, Seminar for Applied Mathematics, ETH Zürich.

I. G. Graham, F. Y. Kuo, D. Nuyens, R. Scheichl and I. H. Sloan (2010), Quasi-Monte Carlo methods for computing flow in random porous media. Technical Report 4/10, Bath Institute for Complex Systems.

M. Griebel, P. Oswald and T. Schiekofer (1999), 'Sparse grids for boundary integral equations', *Numer. Math.* **83**, 279–312.

A. Grothendieck (1955), 'Produits tensoriels topologiques et espaces nucléaires', *Mem. Amer. Math. Soc.* **16**, 140.

H. Harbrecht (2001), Wavelet Galerkin schemes for the boundary element method in three dimensions. PhD thesis, Technische Universität Chemnitz.

H. Harbrecht, R. Schneider and C. Schwab (2008), 'Sparse second moment analysis for elliptic problems in stochastic domains', *Numer. Math.* **109**, 385–414.

M. Hervé (1989), *Analyticity in Infinite-Dimensional Spaces*, Vol. 10 of De Gruyter Studies in Mathematics, Walter de Gruyter.

S. Hildebrandt and E. Wienholtz (1964), 'Constructive proofs of representation theorems in separable Hilbert space', *Comm. Pure Appl. Math.* **17**, 369–373.

V. H. Hoang and C. Schwab (2004/05), 'High-dimensional finite elements for elliptic problems with multiple scales', *Multiscale Model. Simul.* **3**, 168–194.

V.-H. Hoang and C. Schwab (2010*a*), Analytic regularity and gpc approximation for parametric and random 2nd order hyperbolic PDEs. Technical Report 2010-19, Seminar for Applied Mathematics, ETH Zürich. To appear in *Anal. Appl.*

V.-H. Hoang and C. Schwab (2010*b*), Sparse tensor Galerkin discretization for parametric and random parabolic PDEs I: Analytic regularity and gpc-approximation. Technical Report 2010-11, Seminar for Applied Mathematics, ETH Zürich. Submitted.

H. Holden, B. Oksendal, J. Uboe and T. Zhang (1996), *Stochastic Partial Differential Equations: A Modeling, White Noise Functional Approach*, Probability and its Applications, Birkhäuser.

G. C. Hsiao and W. L. Wendland (1977), 'A finite element method for some integral equations of the first kind', *J. Math. Anal. Appl.* **58**, 449–481.

S. Janson (1997), *Gaussian Hilbert spaces*, Vol. 129 of Cambridge Tracts in Mathematics, Cambridge University Press.

R. V. Kadison and J. R. Ringrose (1997), *Fundamentals of the Theory of Operator Algebras I: Elementary Theory*, Vol. 15 of Graduate Studies in Mathematics, AMS, Providence, RI.

S. Kakutani (1948), 'On equivalence of infinite product measures', *Ann. of Math.* (2) **49**, 214–224.

N. Kalton (2003), Quasi-Banach spaces. In *Handbook of the Geometry of Banach Spaces*, Vol. 2, North-Holland, pp. 1099–1130.

B. N. Khoromskij and C. Schwab (2011), 'Tensor-structured Galerkin approximation of parametric and stochastic elliptic PDEs', *SIAM J. Sci. Comput.* **33**, 364–385.

M. Kovács, S. Larsson and F. Lindgren (2010*a*), 'Strong convergence of the finite element method with truncated noise for semilinear parabolic stochastic equations with additive noise', *Numer. Algorithms* **53**, 309–320.

M. Kovács, S. Larsson and F. Saedpanah (2010*b*), 'Finite element approximation of the linear stochastic wave equation with additive noise', *SIAM J. Numer. Anal.* **48**, 408–427.

D. Kressner and C. Tobler (2010), Low-rank tensor Krylov subspace methods for parametrized linear systems. Technical Report 2010-16, Seminar for Applied Mathematics, ETH Zürich. Submitted.

M. Ledoux and M. Talagrand (1991), *Probability in Banach Spaces: Isoperimetry and Processes*, Vol. 23 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3), Springer.

W. A. Light and E. W. Cheney (1985), *Approximation Theory in Tensor Product Spaces*, Vol. 1169 of Lecture Notes in Mathematics, Springer.

S. Lototsky and B. Rozovskii (2006), Stochastic differential equations: A Wiener chaos approach. In *From Stochastic Calculus to Mathematical Finance*, Springer, pp. 433–506.

W. McLean (2000), *Strongly Elliptic Systems and Boundary Integral Equations*, Cambridge University Press.

H. G. Matthies and A. Keese (2005), 'Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations', *Comput. Methods Appl. Mech. Engrg* **194**, 1295–1331.

S. Mishra and C. Schwab (2010), Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random intitial data. Technical Report 2010-24, Seminar for Applied Mathematics, ETH Zürich. Submitted.

J.-C. Nédélec and J. Planchard (1973), 'Une méthode variationnelle d'éléments finis pour la résolution numérique d'un problème extérieur dans R^3', *Rev. Française Autom. Inform. Recherche Opérationnelle Sér. Rouge* **7**, 105–129.

H. Nguyen and R. Stevenson (2009), 'Finite element wavelets with improved quantitative properties', *J. Comput. Appl. Math.* **230**, 706–727.

F. Nobile and R. Tempone (2009), 'Analysis and implementation issues for the numerical approximation of parabolic equations with random coefficients', *Internat. J. Numer. Methods Engrg* **80**, 979–1006.

F. Nobile, R. Tempone and C. G. Webster (2008*a*), 'An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data', *SIAM J. Numer. Anal.* **46**, 2411–2442.

F. Nobile, R. Tempone and C. G. Webster (2008*b*), 'A sparse grid stochastic collocation method for partial differential equations with random input data', *SIAM J. Numer. Anal.* **46**, 2309–2345.

J. T. Oden, I. Babuška, F. Nobile, Y. Feng and R. Tempone (2005), 'Theory and methodology for estimation and control of errors due to modeling, approximation, and uncertainty', *Comput. Methods Appl. Mech. Engrg* **194**, 195–204.

A. Oppenheim (1927), 'On an arithmetic function', *J. London Math. Soc.* **s1-2**, 123–130.

S. Peszat and J. Zabczyk (2007), *Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach*, Vol. 113 of Encyclopedia of Mathematics and its Applications, Cambridge University Press.

T. von Petersdorff and C. Schwab (1996), 'Wavelet approximations for first kind boundary integral equations on polygons', *Numer. Math.* **74**, 479–516.

T. von Petersdorff and C. Schwab (2004), 'Numerical solution of parabolic equations in high dimensions', *M2AN Math. Model. Numer. Anal.* **38**, 93–127.

T. von Petersdorff and C. Schwab (2006), 'Sparse finite element methods for operator equations with stochastic data', *Appl. Math.* **51**, 145–180.

C. Prévôt and M. Röckner (2007), *A Concise Course on Stochastic Partial Differential Equations*, Vol. 1905 of Lecture Notes in Mathematics, Springer.

P. E. Protter (2005), *Stochastic Integration and Differential Equations*, Vol. 21 of Stochastic Modelling and Applied Probability, second edition, version 2.1, Springer.

M. Reed and B. Simon (1980), *Functional Analysis*, Vol. 1 of *Methods of Modern Mathematical Physics*, second edition, Academic Press (Harcourt Brace Jovanovich).

R. A. Ryan (2002), *Introduction to Tensor Products of Banach Spaces*, Springer Monographs in Mathematics, Springer.

S. Sauter and C. Schwab (2010), *Boundary Element Methods*, Springer.

R. Schatten (1943), 'On the direct product of Banach spaces', *Trans. Amer. Math. Soc.* **53**, 195–217.

G. Schmidlin, C. Lage and C. Schwab (2003), 'Rapid solution of first kind boundary integral equations in R^3', *Engrg Anal. Boundary Elem.* (special issue on solving large scale problems using BEM) **27**, 469–490.

R. Schneider (1998), *Multiskalen- und Wavelet-Matrixkompression*, Advances in Numerical Mathematics, Teubner. Analysisbasierte Methoden zur effizienten Lösung großer vollbesetzter Gleichungssysteme. [Analysis-based methods for the efficient solution of large nonsparse systems of equations].

W. Schoutens (2000), *Stochastic Processes and Orthogonal Polynomials*, Vol. 146 of Lecture Notes in Statistics, Springer.

C. Schwab (2002), High dimensional finite elements for elliptic problems with multiple scales and stochastic data, in *Proc. International Congress of Mathematicians*, Vol. III (Beijing 2002), Higher Education Press, Beijing, pp. 727–734.

C. Schwab and R. Stevenson (2008), 'Adaptive wavelet algorithms for elliptic PDEs on product domains', *Math. Comp.* **77**, 71–92.

C. Schwab and R. Stevenson (2009), 'Space–time adaptive wavelet methods for parabolic evolution problems', *Math. Comp.* **78**, 1293–1318.

C. Schwab and A. M. Stuart (2011), Sparse deterministic approximation of Bayesian inverse problems. Technical Report 2011-16, Seminar for Applied Mathematics, ETH Zürich.

C. Schwab and R. A. Todor (2003*a*), 'Sparse finite elements for elliptic problems with stochastic loading', *Numer. Math.* **95**, 707–734.

C. Schwab and R. A. Todor (2003*b*), 'Sparse finite elements for stochastic elliptic problems: Higher order moments', *Computing* **71**, 43–63.

S. Smolyak (1963), 'Quadrature and interpolation formulas for tensor products of certain classes of functions', *Sov. Math. Dokl.* **4**, 240–243.
V. Strassen (1964), 'An invariance principle for the law of the iterated logarithm', *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **3**, 211–226 (1964).
G. Szegő (1975), *Orthogonal Polynomials*, fourth edition, Colloquium Publications, Vol. XXIII, AMS, Providence, RI.
G. Szekeres and P. Turán (1933), 'Über das zweite Hauptproblem der "factorisatio numerorum"', *Acta Litt. Sci. Szeged* **6**, 143–154.
V. N. Temlyakov (1993), *Approximation of Periodic Functions*, Computational Mathematics and Analysis Series, Nova Science Publishers, Commack, NY.
R.-A. Todor (2009), 'A new approach to energy-based sparse finite-element spaces', *IMA J. Numer. Anal.* **29**, 72–85.
R. A. Todor and C. Schwab (2007), 'Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients', *IMA J. Numer. Anal.* **27**, 232–261.
N. N. Vakhania, V. I. Tarieladze and S. A. Chobanyan (1987), *Probability Distributions on Banach Spaces*, Vol. 14 of *Mathematics and its Applications* (Soviet Series), Reidel. Translation by W. A. Woyczynski.
G. W. Wasilkowski and H. Woźniakowski (1995), 'Explicit cost bounds of algorithms for multivariate tensor product problems', *J. Complexity* **11**, 1–56.
N. Wiener (1938), 'The homogeneous chaos', *Amer. J. Math.* **60**, 897–936.
D. Xiu (2009), 'Fast numerical methods for stochastic computations: A review', *Commun. Comput. Phys.* **5**, 242–272.
D. Xiu and J. S. Hesthaven (2005), 'High-order collocation methods for differential equations with random inputs', *SIAM J. Sci. Comput.* **27**, 1118–1139.
D. Xiu and G. E. Karniadakis (2002*a*), 'Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos', *Comput. Methods Appl. Mech. Engrg* **191**, 4927–4948.
D. Xiu and G. E. Karniadakis (2002*b*), 'The Wiener–Askey polynomial chaos for stochastic differential equations', *SIAM J. Sci. Comput.* **24**, 619–644.