Numerical Methods for Boundary Value Problems on Random Domains

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Michael Peters

aus

Rheinbach, Deutschland

Rheinbach, 2014

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Helmut Harbrecht
Prof. Dr. Christoph Schwab

Basel, den 14.10.2014

Prof. Dr. Jörg Schibler
Dekan

Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 3.0 Schweiz
(CC BY-NC-ND 3.0 CH)

Sie dürfen: **Teilen** — den Inhalt kopieren, verbreiten und zugänglich machen

Unter den folgenden Bedingungen:

Namensnennung — Sie müssen den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen.

Keine kommerzielle Nutzung — Sie dürfen diesen Inhalt nicht für kommerzielle Zwecke nutzen.

Keine Bearbeitung erlaubt — Sie dürfen diesen Inhalt nicht bearbeiten, abwandeln oder in anderer Weise verändern.

Wobei gilt:

- **Verzichtserklärung —** Jede der vorgenannten Bedingungen kann **aufgehoben** werden, sofern Sie die ausdrückliche Einwilligung des Rechteinhabers dazu erhalten.
- **Public Domain (gemeinfreie oder nicht-schützbare Inhalte) —** Soweit das Werk, der Inhalt oder irgendein Teil davon zur Public Domain der jeweiligen Rechtsordnung gehört, wird dieser Status von der Lizenz in keiner Weise berührt.
- **Sonstige Rechte —** Die Lizenz hat keinerlei Einfluss auf die folgenden Rechte:
    - Die Rechte, die jedermann wegen der Schranken des Urheberrechts oder aufgrund gesetzlicher Erlaubnisse zustehen (in einigen Ländern als grundsätzliche Doktrin des **fair use** bekannt);
    - Die **Persönlichkeitsrechte** des Urhebers;
    - Rechte anderer Personen, entweder am Lizenzgegenstand selber oder bezüglich seiner Verwendung, zum Beispiel für **Werbung** oder Privatsphärenschutz.
- **Hinweis —** Bei jeder Nutzung oder Verbreitung müssen Sie anderen alle Lizenzbedingungen mitteilen, die für diesen Inhalt gelten. Am einfachsten ist es, an entsprechender Stelle einen Link auf diese Seite einzubinden.

Quelle: http://creativecommons.org/licenses/by-nc-nd/3.0/ch/ Datum: 12.11.2013

VORWORT

Die vorliegende Dissertationsschrift ist im Wesentlichen eine Zusammenstellung der beiden veröffentlichten Artikel

- H. Harbrecht and M. Peters. Comparison of fast boundary element methods on parametric surfaces. *Computer Methods in Applied Mechanics and Engineering*, 261–262:39–55, 2013.
- H. Harbrecht, M. Peters, and M. Siebenmorgen. Combination technique based k-th moment analysis of elliptic problems with random diffusion. *Journal of Computational Physics*, 252:128–141, 2013.

und der beiden Vorabdrucke

- H. Harbrecht, M. Peters, and M. Siebenmorgen. Efficient approximation of random fields for numerical applications. *Preprint 2014-01, Mathematisches Institut Universität Basel*, 2014.
- H. Harbrecht, M. Peters, and M. Siebenmorgen. Numerical solution of elliptic diffusion problems on random domains. *Preprint 2014-08, Mathematisches Institut Universität Basel*, 2014.

Die Resultate dieser Arbeiten sollen hier in einen gemeinsamen Kontext gesetzt werden: Die numerische Lösung von Randwertproblemen auf stochastischen Gebieten. Ich hoffe, dass sich daraus eine neue Perspektive ergibt und das Ganze zu mehr wird als der Summe seiner Teile. In diesem Sinne wurden hier einige Details ergänzt, die in den einzelnen Arbeiten ausgelassen wurden.

Ich möchte die Gelegenheit nutzen, um meine Dankbarkeit zu bekunden. An erster Stelle gebührt mein Dank Herrn Prof. Dr. Helmut Harbrecht, der diese Dissertation betreut hat. Er nahm seine Rolle als Doktorvater wörtlich und hatte immer ein offenes Ohr für meine Anliegen, mögen sie wissenschaftlicher oder trivialer Natur gewesen sein. Ich danke Herrn Prof. Dr. Christoph Schwab für die Übernahme des Korreferats. Ferner möchte ich hier meine Eltern Petra und Uwe Peters erwähnen. Ich vermag nicht in schöne Worte zu fassen, was sie alles für mich getan haben und wie tief die Dankbarkeit ist, die ich für sie empfinde. Daher versuche ich es mit einfachen Worten: Danke Mama, danke Papa. Mein besonderer Dank gilt Markus Siebenmorgen. Im Jahr 2004 haben wir in Bonn zusammen unser Studium der Mathematik aufgenommen. Seither ist er mir ein treuer Freund und wir haben uns vielen Herausforderungen des Lebens gemeinsam gestellt. So haben wir beide im Jahr 2010 in Stuttgart unser Doktorat in Mathematik bei Herrn Prof. Dr. Helmut Harbrecht begonnen und sind mit ihm im Jahr 2011 nach Basel gewechselt. Schließlich danke ich dem Schweizerischen Nationalfonds (SNSF), der diese Arbeit durch das Projekt „Rapid Solution of Boundary Value Problems on Stochastic Domains“ gefördert hat.

CONTENTS

Chapter I

INTRODUCTION

Various problems in science and engineering can be formulated as boundary value problems for an unknown function. In general, the numerical simulation is well understood provided that the input parameters are known exactly. In many applications, however, the input parameters are not known exactly. Especially, the treatment of uncertainties in the computational domain has become of growing interest, see e.g. [CK07, HSS08b, TX06, XT06]. In this thesis, we consider the elliptic diffusion equation

$$-\operatorname{div}\left(\alpha \nabla u(\omega)\right) = f \text{ in } D(\omega), \quad u(\omega) = 0 \text{ on } \partial D(\omega), \tag{1.1}$$

as a model problem where the underlying domain $D(\omega) \subset \mathbb{R}^d$ or respectively its boundary $\partial D(\omega)$ are random. For example, one might think of tolerances in the shape of products fabricated by line production, or shapes which stem from inverse problems, like e.g. tomography. Of course, besides a scalar diffusion coefficient $\alpha(\mathbf{x})$, one could also consider a diffusion matrix $\mathbf{A}(\mathbf{x})$. Even so, the emphasis of our considerations will be laid on the case $\alpha(\mathbf{x}) \equiv 1$, i.e. the Poisson equation. As we will see, the case of an arbitrary positive diffusion coefficient and even the case of a symmetric positive diffusion matrix can also be deduced from the presented framework.

Besides the fictitious domain approach considered in [CK07], one might essentially distinguish two approaches: the *perturbation method* and the *domain mapping method*, both of which shall be considered in this thesis.

The perturbation method starts with a prescribed perturbation field

$$\mathbf{V}(\omega)\colon \partial D_{\text{ref}} \to \mathbb{R}^d$$

at the boundary ∂D_{ref} and uses a *shape Taylor expansion* with respect to this perturbation field to represent the solution to the model problem, see e.g. [HL13, HSS08b]. Whereas, the domain mapping method requires that the perturbation field is also known in the interior of the domain D_{ref}, i.e.

$$\mathbf{V}(\omega)\colon \overline{D_{\text{ref}}} \to \mathbb{R}^d.$$

Then, the problem is transformed to the nominal, fixed domain D_{ref}. This yields a partial differential equation with a random diffusion matrix and a stochastic right hand side which are correlated, cf. [CNT13, MNK11, TX06, XT06].

The major drawback of the perturbation method is that it is only feasible for relatively small perturbations. Thus, in order to treat larger perturbations, the domain

mapping method is the method of choice. Nevertheless, it might in practice be much easier to obtain measurements from the outside of a work-piece to estimate the perturbation field $\mathbf{V}(\omega)$ rather than from its interior. If no information of the vector field inside the domain is available, it has to be extended appropriately, for example by the Laplacian, as proposed in [MNK11, TX06].

We would like to point out that the two approaches are in fact not comparable at all. In the perturbation method, we use a problem description in terms of *Eulerian coordinates*, which means that we keep each point fixed and perturb just the domain's boundary. When considering the domain mapping method, we change to *Lagrangian coordinates*, which means that we keep track of the movement of each point. The correspondence between those two approaches can be expressed in terms of the *local shape derivative* $\delta u[\mathbf{V}(\omega)]$ and the *material derivative* $\dot{u}[\mathbf{V}(\omega)]$ of a given function u. They differ by a transport term, cf. [SZ92]:

$$\dot{u}[\mathbf{V}(\omega)] = \delta u[\mathbf{V}(\omega)] + \langle \nabla u, \mathbf{V}(\omega) \rangle.$$

For both methods, namely the domain mapping method and the perturbation method, the starting point for our considerations will be the knowledge of an appropriate description of the the random field $\mathbf{V}(\omega)$. To that end, we assume that the random vector field is described in terms of its mean

$$\mathbb{E}[\mathbf{V}] \colon D_{\mathrm{ref}} \to \mathbb{R}^d, \quad \mathbb{E}[\mathbf{V}](\mathbf{x}) = \big[\mathbb{E}[v_1](\mathbf{x}), \dots, \mathbb{E}[v_d](\mathbf{x})\big]^\intercal$$

and its (matrix valued) covariance function

$$\mathrm{Cov}[\mathbf{V}] \colon D_{\mathrm{ref}} \times D_{\mathrm{ref}} \to \mathbb{R}^{d \times d}, \quad \mathrm{Cov}[\mathbf{V}](\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \mathrm{Cov}_{1,1}(\mathbf{x}, \mathbf{y}) & \cdots & \mathrm{Cov}_{1,d}(\mathbf{x}, \mathbf{y}) \\ \vdots & & \vdots \\ \mathrm{Cov}_{d,1}(\mathbf{x}, \mathbf{y}) & \cdots & \mathrm{Cov}_{d,d}(\mathbf{x}, \mathbf{y}) \end{bmatrix}.$$

For the perturbation method, this representation of the random vector field is already sufficient. To make the vector field $\mathbf{V}(\omega)$ feasible for the domain mapping method, we introduce the *Karhunen-Loève expansion*.

The Karhunen-Loève expansion separates the spatial variable $\mathbf{x}$ and the stochastic variable ω. It is also used to make random diffusion coefficients or random right hand sides applicable for numerical computations in the stochastic Galerkin or the stochastic collocation method, see e.g. [BNT07, BTZ04, FST05, GS91, MK05, SG11] and the references therein. Thus, one naturally aims at efficient algorithms for the computation of the Karhunen-Loève expansion. In this context, approaches to efficiently compute the Karhunen-Loève expansion (for scalar valued random fields) have been made by means of the *Fast Multipole Method* (FMM) based on interpolation (cf. [Gie01]) in [ST06] and with the aid of $\mathcal{H}$-matrices (cf. [Hac09]) in [EEU07]. The idea in these works is to provide a data-sparse representation of the covariance operator which is then used to solve the related eigen-problem numerically by a Krylov subspace method, cf. [Saa92]. Of course, another algorithm for the efficient approximation of non-local operators, like the *Adaptive Cross Approximation* (ACA), cf. [Beb00, BR03], or the *Wavelet Galerkin Scheme* (WGS), cf. [DHS06, HS06], can be considered as well for the data sparse representation of the covariance operator.

In any case, the major drawback of these approaches is that the number of eigenvalues to be computed has to be known in advance which might be a strong requirement in practice. To overcome this obstruction, we present an alternative approach based on the *Pivoted Cholesky Decomposition* (PCD). The PCD is an established tool in the simulation of Gaussian processes and and the computation of low-rank approximations to covariance matrices, see e.g. [RW05, BL77, FWA+09]. It can be interpreted as a single-block ACA with total pivoting, cf. [HPS12]. Hence, only the main diagonal of the discretized operator has to be precomputed, which can be performed in essentially, i.e. up to possible poly-logarithmic terms, linear complexity if the quadrature proposed in [SS97] is applied to discretize the underlying operator. Then, in each step of the algorithm, the quality of the approximation to the random field is controllable by means of the trace. If the desired accuracy is achieved, the algorithm stops with an M-term approximation to the operator. If M is substantially smaller than the dimension of the ansatz space, we end up with a remarkable computational speed-up. The related Karhunen-Loève expansion might then be computed in a post-processing step. Notice that in this case the PCD yields a full but relatively small eigen-problem if the operator under consideration exhibits a certain smoothness.

The outline of this thesis is as follows.

Chapter II provides the theoretical background for the further considerations. Here, we will introduce tensor products of Hilbert spaces, which are the theoretical foundation for the representation of random fields. Moreover, we define the *Karhunen-Loève expansion* and introduce the *Matérn class* of covariance functions. These covariance functions will serve as a benchmark for the numerical computations. For the numerical realization of the domain mapping method, we propose in this thesis the use of parametric finite elements. These are also introduced here.

In Chapter III, we consider the numerical approximation of random fields as in [HPS14a]. Especially, we transfer here the results provided by this work to the case of vector valued random fields and show how these fields can be represented by means of the PCD. Moreover, we present special variants of the ACA and the FMM based on parametric representations of the underlying geometry as introduced in [HP13]. This approach yields very efficient variants of the two methods. Furthermore, in order to speed up the matrix-vector product for the Krylov subspace method, we present a related and improved $\mathcal{H}^2$-matrix, cf. [HB02], version of the FMM.

Now, the following question arises: which approach is more efficient in practice? We will answer this question by numerically comparing ACA, FMM and the PCD. As Krylov subspace method for ACA and FMM, we use the *Implicit Restarted Arnoldi Method* (IRAM), cf. [LS96, LSY98, Sor92]. For the sake of simplicity, we consider here only scalar valued covariance functions. Notice that, for matrix valued covariance functions, we would have to compress each block $\mathrm{Cov}_{i,j}$ of the related covariance operator separately for ACA and FMM, since no global smoothness between two particular blocks is feasible. The PCD does not suffer from this fact since it is independent of any smoothness assumption. Thus, it can approximate the whole covariance operator en bloc.

Chapter IV deals with the domain mapping method as presented in [HPS14b]. In [CNT13], it is shown for a specific class of variation fields that the solution to (1.1) provides analytic regularity with respect to the stochastic parameter. Thus, the random

solution can be approximated by using the *isotropic* variant of the stochastic collocation method from [BNT07]. We will generalize the analysis from [CNT13] to arbitrary domain perturbation fields which are described by their mean and their covariance. Taking the Karhunen-Loève expansion of $\mathbf{V}(\omega)$ as the starting point, we show rates of decay for the derivatives of the solution to (1.1) with respect to the stochastic parameter. From this, we immediately derive the tractability of the Quasi-Monte Carlo method based on the Halton sequence, cf. [Hal60, HPS13b, Nie92]. Furthermore, the decay estimates can be sharpened in case of univariate derivatives which yields the applicability and related rates of convergence for the *anisotropic* variant of the stochastic collocation method from [BNT07].

Employing parametric finite elements, we are able to approximate the mean and the variance of the solution to (1.1) by computing each sample on the particular realization $D(\omega_i) = \mathbf{V}(\omega_i, D_{\mathrm{ref}})$ of the random domain rather than on the reference domain D_{ref}. This yields a non-intrusive approach to solve the problem under consideration. Actually, any available finite element solver can be employed to compute the particular samples. Using this approach rather than mapping the diffusion problem always to the reference domain, we can easily treat also stochastic interface problems, cf. [HL13].

Chapter V treats the perturbation method for the numerical approximation of the solution to (1.1). Having the mean and the covariance of the random vector field at hand, we aim at the computation of the corresponding statistics of the unknown random solution. Making use of sensitivity analysis, we linearize the solution's nonlinear dependence on the random vector field $\mathbf{V}(\omega)$. Based on this, we derive deterministic equations, which compute, to leading order, the mean field and the covariance. In particular, the covariance solves a tensor product boundary value problem on the product domain $D_{\mathrm{ref}} \times D_{\mathrm{ref}}$.

In difference to previous works, we do not explicitly use wavelets [HSS08b, ST03a, ST03b] or multilevel frames [Har10b, HSS08a] for the discretization in a sparse tensor product space. Instead, we define the complement spaces which enter the sparse tensor product construction by Galerkin projections. The Galerkin discretization leads then to a system of linear equations which decouples into sub-problems with respect to full tensor product spaces of small size. These sub-problems can be solved by standard multilevel finite element methods. In our particular realization, we need only the access to the stiffness matrix, the BPX preconditioner (cf. [BPX90]) and the sparse grid interpolant (cf. [BG04]) of the two-point correlation function of the random vector field under consideration. In this sense, our approach can be considered to be weakly intrusive. The resulting representation of the covariance is known as the combination technique [GSZ92]. Nevertheless, in difference to [GSZ92, PZ99, Rei13, XZ04], this representation is a consequence of the Galerkin method in the sparse tensor product space and is not an additional approximation step.

Throughout this thesis, in order to avoid the repeated use of generic but unspecified constants, by $C \lesssim D$ we mean that C can be bounded by a multiple of D, independently of parameters which C and D may depend on. Obviously, $C \gtrsim D$ is defined as $D \lesssim C$, and $C \eqsim D$ as $C \lesssim D$ and $C \gtrsim D$.

Chapter II

PRELIMINARIES

1. Tensor products of Hilbert spaces

Tensor products have been considered in the context of Banach spaces at first in [Sch50]. The construction presented there extends to the tensor product of Hilbert spaces. There exists a close connection between tensor products of Hilbert spaces, Hilbert-Schmidt operators and trace-class operators, cf. [KR86, Sch50, Sch60]. This connection is later on exploited in order to compute separable expansions of random vector fields. Thus, we think it is convenient to outline here in brief the construction of tensor products of Hilbert spaces. We follow the representation in [KR86], where the tensor product of Hilbert spaces is defined by its universal property. For more details and the related proofs, we refer the reader to [KR86]. In the sequel, let $(\mathscr{X}, (\cdot,\cdot)_{\mathscr{X}})$, $(\mathscr{Y}, (\cdot,\cdot)_{\mathscr{Y}})$ and $(\mathscr{Z}, (\cdot,\cdot)_{\mathscr{Z}})$ denote separable Hilbert spaces over the field of real numbers $\mathbb{R}$.

(1.1) **Definition.** The bounded, bilinear functional $f\colon \mathscr{X} \times \mathscr{Y} \to \mathbb{R}$ is called *Hilbert-Schmidt functional* if

$$\sum_i \sum_j |f(\varphi_i, \psi_j)|^2 < \infty \tag{1.2}$$

holds for two orthonormal bases $\{\varphi_i\}_i \subset \mathscr{X}$ and $\{\psi_i\}_i \subset \mathscr{Y}$.

It can be shown that if (1.2) is satisfied for one pair of orthonormal bases $\{\varphi_i\}_i \subset \mathscr{X}$ and $\{\psi_i\}_i \subset \mathscr{Y}$, it holds for all pairs of orthonormal basis, cf. [KR86, Proposition 2.6.1]. Moreover, the Hilbert-Schmidt functionals on $\mathscr{X} \times \mathscr{Y}$ form itself a Hilbert space.

(1.3) **Theorem.** The set $\mathfrak{HSF}$ of all Hilbert-Schmidt functionals on $\mathscr{X} \times \mathscr{Y}$ forms a Hilbert space with respect to the linear structure

$$(\alpha f_1 + \beta f_2)(x, y) = \alpha f_1(x, y) + \beta f_2(x, y)$$

for any two Hilbert-Schmidt functionals $f_1, f_2 \in \mathfrak{HSF}$ and $\alpha, \beta \in \mathbb{R}$. The related inner product is given by

$$(f_1, f_2)_{\mathfrak{HSF}} = \sum_i \sum_j f_1(\varphi_i, \psi_j) f_2(\varphi_i, \psi_j)$$

for an arbitrary pair of orthonormal bases $\{\varphi_i\}_i \subset \mathscr{X}$ and $\{\psi_i\}_i \subset \mathscr{Y}$. The sum is absolutely convergent and independent of the choice of the orthonormal bases. Moreover, the related norm is given by $\|f\|_{\mathfrak{HSF}} := \sqrt{\langle f, f\rangle_{\mathfrak{HSF}}}$. Finally, for each $u \in \mathscr{X}$ and $v \in \mathscr{Y}$, it holds

$$f_{u,v}(x,y) := (x,u)_{\mathscr{X}}(y,v)_{\mathscr{Y}} \in \mathfrak{HSF}$$

and the set $\{f_{\varphi_i,\psi_j}\}_{i,j}$ forms an orthonormal basis of $\mathfrak{HSF}$.

Proof. For the proof, see [KR86, Proposition 2.6.2]. □

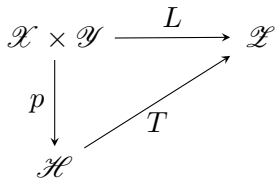

Figure II.1: Universal property of the tensor product of Hilbert spaces.

(1.4) **Definition.** The bounded, bilinear mapping $L\colon \mathscr{X} \times \mathscr{Y} \to \mathscr{Z}$ is called *weak Hilbert-Schmidt mapping* if

$$L_z(x,y) := \big(L(x,y), z\big)_{\mathscr{Z}} \quad \text{for } x \in \mathscr{X}, y \in \mathscr{Y}$$

defines for each $z \in \mathscr{Z}$ a Hilbert-Schmidt functional on $\mathscr{X} \times \mathscr{Y}$. The norm $\|L\|$ of L is the smallest value $c \in \mathbb{R}$ such that $\|L_z\|_{\mathfrak{HSF}} \leqslant c\|z\|_{\mathscr{Z}}$.

With this definition at hand, the tensor product $\mathscr{X} \otimes \mathscr{Y}$ of the Hilbert spaces $\mathscr{X}$ and $\mathscr{Y}$ may be characterized as follows.

(1.5) **Theorem.** There exists a Hilbert space $\big(\mathscr{H}, (\cdot,\cdot)_{\mathscr{H}}\big)$ and a weak Hilbert-Schmidt mapping $p\colon \mathscr{X} \times \mathscr{Y} \to \mathscr{H}$ such that for an arbitrary weak Hilbert-Schmidt mapping $L\colon \mathscr{X} \times \mathscr{Y} \to \mathscr{Z}$ there exists a unique, bounded linear mapping $T\colon \mathscr{H} \to \mathscr{Z}$ with $L = Tp$ and $\|T\| = \|L\|$. The space $\mathscr{H}$ and the mapping p are uniquely determined up to an isometric isomorphism. Moreover, it holds

$$\big(p(x_1,y_1), p(x_2,y_2)\big)_{\mathscr{H}} = (x_1,x_2)_{\mathscr{X}}(y_1,y_2)_{\mathscr{Y}}$$

for any $x_1, x_2 \in \mathscr{X}$, $y_1, y_2 \in \mathscr{Y}$. If $\{\varphi_i\}_i \subset \mathscr{X}$ and $\{\psi_i\}_i \subset \mathscr{Y}$ are two orthonormal bases, the set $\{p(\varphi_i,\psi_j)\}_{i,j}$ forms an orthonormal basis of $\mathscr{H}$ and the operator norm of p satisfies $\|p\| = 1$.

Proof. For a proof of this statement, see [KR86, Proposition 2.6.4]. □

(1.6) **Remark.** In the sequel, we set $\mathscr{X} \otimes \mathscr{Y} := \mathscr{H}$. The elements $x \otimes y := p(x, y)$ for $x \in \mathscr{X}$, $y \in \mathscr{Y}$ are called *dyads*. Their finite sums $\sum_{i=1}^{n} x_i \otimes y_i$ form an everywhere dense subspace of $\mathscr{X} \otimes \mathscr{Y}$, cf. [KR86, Proposition 2.6.6]. In fact, the space

$$\mathscr{H}_0 := \left\{ \sum_{i=1}^{n} x_i \otimes y_i : \{x_i\}_{i=1}^{n} \subset \mathscr{X}, \{y_i\}_{i=1}^{n} \subset \mathscr{Y}, n \in \mathbb{N} \right\}$$

corresponds to the *algebraic tensor product of* $\mathscr{X}$ *and* $\mathscr{Y}$, cf. [KR86, Remark 2.6.7]. Thus, we can consider $\mathscr{X} \otimes \mathscr{Y}$ as the completion of the space $\mathscr{H}_0$ endowed with the inner product $(x_1 \otimes y_1, x_2 \otimes y_2)_{\mathscr{H}_0} = (x_1, x_2)_{\mathscr{X}} (y_1, y_2)_{\mathscr{Y}}$ defined for dyads and extended by linearity to $\mathscr{H}_0$. This definition of the tensor product of Hilbert spaces conforms to the construction in [LC85].

Next, we establish the connection between the tensor product of Hilbert spaces and the class of Hilbert-Schmidt operators. To that end, we associate the bilinear form

$$b_T \colon \mathscr{X} \times \mathscr{Y} \to \mathbb{R}, \quad b_T(x, y) := (Tx, y)_{\mathscr{Y}}$$

to a given bounded linear operator $T \colon \mathscr{X} \to \mathscr{Y}$. Obviously, the map $T \mapsto b_T$ is one-to-one from the space of bounded linear operators to the bounded bilinear functionals.

(1.7) **Definition.** The bounded linear operator $T \colon \mathscr{X} \to \mathscr{Y}$ is called *Hilbert-Schmidt operator* if b_T is a Hilbert-Schmidt functional. The linear space of Hilbert-Schmidt operators is denoted by $\mathfrak{HSO}$.

Due to the mapping $T \mapsto b_T$, the Hilbert space structure on $\mathfrak{HSF}$ directly transfers to $\mathfrak{HSO}$ and yields the inner product

$$(S, T)_{\mathfrak{HSO}} := \sum_i \sum_j (S\varphi_i, \psi_j)_{\mathscr{Y}} (T\varphi_i, \psi_j)_{\mathscr{Y}}$$

for any two orthonormal bases $\{\varphi_i\}_i \subset \mathscr{X}$ and $\{\psi_i\}_i \subset \mathscr{Y}$. The related norm is denoted by $\|T\|_{\mathfrak{HSO}} := \sqrt{\langle T, T \rangle_{\mathfrak{HSO}}}$. The definition of the inner product is equivalent to

(1.8) $$(S, T)_{\mathfrak{HSO}} = \sum_i (S\varphi_i, T\varphi_i)_{\mathscr{Y}}$$

by Parseval's identity. Notice that $\|T\|_{\mathfrak{HSO}} < \infty$ already implies the boundedness of T in the operator norm. This is easily seen by completing a given function $\varphi \in \mathscr{X}$ with $\|\varphi\|_{\mathscr{X}} = 1$ to an orthogonal basis of $\mathscr{X}$ and then observing that $\|T\varphi\|_{\mathscr{Y}} \leqslant \|T\|_{\mathfrak{HSO}}$.

Now, we have the following identification of the tensor product of Hilbert spaces.

(1.9) **Theorem.** For each $u \in \mathscr{X}$ and $v \in \mathscr{Y}$, the operator

$$T_{u,v} \colon \mathscr{X} \to \mathscr{Y}, \quad T_{u,v} x := (x, u)_{\mathscr{X}} v \quad \text{for } x \in \mathscr{X},$$

defines a Hilbert-Schmidt operator. Moreover, there exists an isometric isomorphism U from $\mathscr{X} \otimes \mathscr{Y}$ to $\mathfrak{HSO}$ such that $U(u \otimes v) = T_{u,v}$ for any $u \in \mathscr{X}$ and $v \in \mathscr{Y}$.

Proof. For a proof of this result, see [KR86, Proposition 2.6.9]. □

The theorem suggests that we may consider expressions of the form $z = \sum_{i=1}^{n} x_i \otimes y_i \in \mathscr{H}_0$ to define the operators

$$T_z \colon \mathscr{X} \to \mathscr{Y}, \quad T_z u := \sum_{i=1}^{n} (u, x_i)_{\mathscr{X}} y_i \in \mathscr{Y} \quad \text{for } u \in \mathscr{X}$$

which are at most of rank n. Extension by continuity, i.e. setting $T_z = \lim_{z_0 \in \mathscr{H}_0, z_0 \to z} T_{z_0}$ for $z \in \mathscr{X} \otimes \mathscr{Y} \setminus \mathscr{H}_0$, then yields, together with the isometric isomorphism U, the expression T_z for an arbitrary $z \in \mathscr{X} \otimes \mathscr{Y}$. More precisely, for $z = \sum_{i \in \mathcal{I}} x_i \otimes y_i$, where $\mathcal{I} \subset \mathbb{N}$, and for an arbitrary orthonormal basis $\{\varphi_i\}_i \subset \mathscr{X}$ we have that

$$\begin{aligned} (T_z, T_z)_{\mathfrak{HSO}} &= \sum_i \Big(\sum_{j \in \mathcal{I}} (\varphi_i, x_j)_{\mathscr{X}} y_j, \sum_{k \in \mathcal{I}} (\varphi_i, x_k)_{\mathscr{X}} y_k \Big)_{\mathscr{Y}} \\ &= \sum_i \sum_{j \in \mathcal{I}} \sum_{k \in \mathcal{I}} (\varphi_i, x_j)_{\mathscr{X}} (\varphi_i, x_k)_{\mathscr{X}} (y_j, y_k)_{\mathscr{Y}} \\ &= \sum_{j \in \mathcal{I}} \sum_{k \in \mathcal{I}} (x_j, x_k)_{\mathscr{X}} (y_j, y_k)_{\mathscr{Y}} = \|z\|^2_{\mathscr{X} \otimes \mathscr{Y}} \end{aligned} \tag{1.10}$$

by Parseval's identity.

We complete this section by showing that each element $z \in \mathscr{X} \otimes \mathscr{Y}$ provides an orthogonal representation of the form

$$z = \sum_{i \in \mathcal{I}} \sigma_i \varphi_i \otimes \psi_i \tag{1.11}$$

with two orthonormal families $\{\varphi_i\}_{i \in \mathcal{I}} \subset \mathscr{X}$ and $\{\psi_i\}_{i \in \mathcal{I}} \subset \mathscr{Y}$ and non-negative real numbers $\{\sigma_i\}_{i \in \mathcal{I}}$. To that end, we consider the operators

$$T_z^\star T_z \colon \mathscr{X} \to \mathscr{X} \quad \text{and} \quad T_z T_z^\star \colon \mathscr{Y} \to \mathscr{Y},$$

where $T_z^\star \colon \mathscr{Y} \to \mathscr{X}$ denotes the Hilbert space adjoint of T_z, i.e.

$$T_z^\star v = \sum_{i \in \mathcal{I}} (v, y_i)_{\mathscr{Y}} x_i \quad \text{for all } v \in \mathscr{Y}.$$

The set of all products $S^\star T$ of two operators $S, T \in \mathfrak{HSO}$ form a subspace of the Hilbert-Schmidt operators, cf. [Sch60].

(1.12) **Definition.** The products of two operators in $\mathfrak{HSO}$ form the *trace-class*. In particular, an operator $T \colon \mathscr{X} \to \mathscr{X}$ is in the trace-class if

$$\operatorname{Tr} T := \sum_i (T\varphi_i, \varphi_i)_{\mathscr{X}} < \infty$$

for an arbitrary orthonormal basis $\{\varphi_i\}_i \subset \mathscr{X}$.

By the definition of the trace, we have for $A = S^\star T$ with $S, T \colon \mathscr{X} \to \mathscr{Y}$

$$\operatorname{Tr} A = \sum_i (A\varphi_i, \varphi_i)_{\mathscr{X}} = \sum_i (T\varphi_i, S\varphi_i)_{\mathscr{Y}} = (S, T)_{\mathfrak{HSO}}, \tag{1.13}$$

cf. [Sch60, Remark 1].

(1.14) **Remark.** The class of Hilbert-Schmidt operators contains those operators for which the sequence of eigenvalues satisfy $\{\lambda_i\}_{i\in\mathcal{I}} \in \ell^2(\mathcal{I})$, whereas the eigenvalues of trace-class operators satisfy $\{\lambda_i\}_{i\in\mathcal{I}} \in \ell^1(\mathcal{I})$.

The operator T_z is the norm limit of finite rank operators and thus compact, cf. [Alt07, Lemma 8.2]. Hence, the operators $T_z^\star T_z$ and $T_z T_z^\star$ are also compact and additionally symmetric. For example, we have

$$(T_z^\star T_z u_1, u_2)_{\mathscr{X}} = (T_z u_1, T_z u_2)_{\mathscr{Y}} = (u_1, T_z^\star T_z u_2)_{\mathscr{X}}.$$

Consequently, the spectral theorem for compact and normal operators, cf. [Alt07, Theorem 10.12], applies to $T_z^\star T_z$ and $T_z T_z^\star$. The spectral theorem reads as follows.

(1.15) **Theorem.** Let $T\colon \mathscr{X} \to \mathscr{X}$ be a symmetric and compact operator. Then, there exists an orthonormal family $\{\varphi_i\}_{i\in\mathcal{I}}$ with $\mathcal{I} \subset \mathbb{N}$ and $\{\lambda_i\}_{i\in\mathcal{I}} \subset \mathbb{R}\setminus\{0\}$ such that $T\varphi_i = \lambda_i\varphi_i$ for all $i \in \mathcal{I}$. Furthermore, it holds

$$Tu = \sum_{i\in\mathcal{I}} \lambda_i (u, \varphi_i)_{\mathscr{X}} \varphi_i \quad \text{for all } u \in \mathscr{X}.$$

Proof. For a proof of this theorem, we refer to [Alt07, Theorem 10.12]. □

In the following, we assume that the index set $\mathcal{I}$ provides a meaningful numbering, i.e. we assume that either $\mathcal{I} = \{1, 2, \dots, n\}$ for some $n \in \mathbb{N}$ or $\mathcal{I} = \mathbb{N}\setminus\{0\}$.

For the eigenvalues of $T_z^\star T_z$ it holds $\{\lambda_i\}_{i\in\mathcal{I}} \in (0, \infty)$ due to

$$0 \leqslant \|T_z u\|_{\mathscr{Y}}^2 = (T_z u, T_z u)_{\mathscr{Y}} = (T_z^\star T_z u, u)_{\mathscr{X}} \quad \text{for all } u \in \mathscr{X}.$$

The same argumentation implies the positivity of the eigenvalues of $T_z T_z^\star$. Especially, we have the following connection between the eigen-pairs of $T_z^\star T_z$ and $T_z T_z^\star$.

(1.16) **Lemma.** Let the set $\{(\lambda_i, \varphi_i)\}_{i\in\mathcal{I}}$ denote the eigen-pairs of $T_z^\star T_z$. Then, it holds that $\{(\lambda_i, 1/\sqrt{\lambda_i} T_z\varphi_i)\}_{i\in\mathcal{I}}$ are precisely the eigen-pairs of $T_z T_z^\star$.

Proof. Let (λ_i, φ_i) for $i \in \mathcal{I}$ be an eigen-pair of $T_z^\star T_z$. It holds

$$T_z T_z^\star (T_z\varphi_i) = T_z (T_z^\star T_z \varphi_i) = \lambda_i T_z \varphi_i.$$

Moreover, we have for another eigenfunction $T_z\varphi_j$ that

$$(T_z\varphi_i, T_z\varphi_j)_{\mathscr{Y}} = (T_z^\star T_z \varphi_i, \varphi_j)_{\mathscr{X}} = \delta_{i,j}\lambda_i.$$

This shows that $(\lambda_i, 1/\sqrt{\lambda_i} T_z\varphi_i)$ is an eigen-pair $T_z T_z^\star$ for every eigen-pair (λ_i, φ_i) of $T_z^\star T_z$. Interchanging the roles of $T_z^\star T_z$ and $T_z T_z^\star$ in the preceding argumentation yields that the cardinality of the set of eigen-pairs for both operators coincides. □

We have

$$\sum_{i\in\mathcal{I}} \lambda_i = \sum_{i\in\mathcal{I}} (T_z^\star T_z \varphi_i, \varphi_i)_{\mathscr{X}} = \operatorname{Tr} T_z^\star T_z,$$

which is easily seen by completing the eigenfunctions $\{\varphi_i\}_{i\in\mathcal{I}}$ to an orthonormal basis of $\mathscr{X}$. Since the eigenvalues of $T_z^\star T_z$ and $T_z T_z^\star$ coincide, it holds

$$\operatorname{Tr} T_z^\star T_z = \operatorname{Tr} T_z T_z^\star = (T_z^\star, T_z^\star)_{\mathfrak{HS}} = (T_z, T_z)_{\mathfrak{HS}} = \|z\|^2_{\mathscr{X}\otimes\mathscr{Y}} \tag{1.17}$$

due to the isometry (1.10). This relationship will serve later on as a measure of the approximation error for random fields.

Next, the following theorem gives us that each $z \in \mathscr{X}\otimes\mathscr{Y}$ exhibits a decomposition similar to (1.11).

(1.18) **Theorem.** Every $z \in \mathscr{X}\otimes\mathscr{Y}$ can be represented in the form

$$z = \sum_{i\in\mathcal{I}} \sigma_i \varphi_i \otimes \psi_i,$$

where $\sigma_i = \sqrt{\lambda_i}$, $\psi_i = 1/\sqrt{\lambda_i} T_z \varphi_i$ and $\{(\lambda_i, \varphi_i)\}_{i\in\mathcal{I}}$ corresponds to the eigen-pairs of $T_z^\star T_z$.

Proof. Due to the equivalence of the spaces $\mathfrak{HS}$ and $\mathscr{X}\otimes\mathscr{Y}$, cf. Theorem (1.9), it suffices to show that the operators defined by $z = \sum_{i\in\mathcal{I}'} x_i \otimes y_i$ and the orthogonal representation $\tilde{z} = \sum_{i\in\mathcal{I}} \sigma_i\varphi_i \otimes \psi_i$ coincide. We complete the eigenfunctions $\{\varphi_i\}_{i\in\mathcal{I}}$ of T_z to an orthonormal basis of $\mathscr{X}$. Then, it holds

$$T_{\tilde{z}}\varphi_j = \sum_{i\in\mathcal{I}} \sqrt{\lambda_i}(\varphi_j, \varphi_i)\frac{1}{\sqrt{\lambda_i}} T_z\varphi_i = 0 \quad \text{for } j \notin \mathcal{I}$$

and

$$T_{\tilde{z}}\varphi_j = \sum_{i\in\mathcal{I}} \sqrt{\lambda_i}(\varphi_j, \varphi_i)\frac{1}{\sqrt{\lambda_i}} T_z\varphi_i = T_z\varphi_j \quad \text{for } j \in \mathcal{I}.$$

It remains to show that $T_z\varphi_j = 0$ for all $\varphi_j \notin \mathcal{I}$. To that end, assume that $T_z\varphi_j \neq 0$ for some $j \notin \mathcal{I}$. Therefore, we have $0 < \|T_z\varphi_j\|^2_{\mathscr{Y}} = (T_z^\star T_z\varphi_j, \varphi_j)$. Hence,

$$0 \neq T_z^\star T_z\varphi_j = \sum_{i\in\mathcal{I}} \lambda_i(\varphi_j, \varphi_i)\varphi_i,$$

which is a contradiction to the orthogonality of the basis $\{\varphi_i\} \subset \mathscr{X}$. □

(1.19) **Remark.** Without loss of generality, we assume that the *singular values* are sorted in decreasing order, i.e. $\sigma_1 \geqslant \sigma_2 \geqslant \ldots$. The representation (1.11) is unique up to isometries of the eigen-spaces. The representation becomes unique if we prescribe either the orthonormal basis in $\{\varphi_i\}_i$ in $\mathscr{X}$ or the orthonormal basis $\{\psi_i\}_i$ in $\mathscr{Y}$, see also [ST06].

This statement is seen as follows. Without loss of generality, we prescribe the orthonormal basis $\{\psi_i\}_i$ in $\mathscr{Y}$. Now, let $z = \sum_i x_i \otimes \psi_i = \sum_i \tilde{x}_i \otimes \psi_i$. Thus, it holds $0 = \sum_i (x_i - \tilde{x}_i) \otimes \psi_i$ and therefore

$$\begin{aligned} 0 &= \Big(\sum_i (x_i - \tilde{x}_i) \otimes \psi_i, \sum_i (x_i - \tilde{x}_i) \otimes \psi_i \Big)_{\mathscr{X}\otimes\mathscr{Y}} \\ &= \sum_{i,j} \big((x_i - \tilde{x}_i), (x_i - \tilde{x}_i)\big)_{\mathscr{X}} (\psi_i, \psi_j)_{\mathscr{Y}} = \sum_i \|x_i - \tilde{x}_i\|^2_{\mathscr{X}}. \end{aligned}$$

This shows $x_i = \tilde{x}_i$ for all i.

(1.20) **Corollary.** Let $\sigma_1 \geqslant \sigma_2 \geqslant \dots$ denote the singular values of $z \in \mathscr{X} \otimes \mathscr{Y}$. Then, we have that $\|T_z\| = \sigma_1$ for the operator norm of the associated Hilbert-Schmidt operator T_z.

Proof. It holds $\|T_z\| = \sup_{\|u\|_{\mathscr{X}}=1} \|T_z u\|_{\mathscr{Y}}$. Since the preimage of T_z is spanned by the eigenfunctions $\{\varphi_i\}_{i\in\mathcal{I}}$ of $T_z^\star T_z$, it suffices to consider linear combinations of these functions. Let $u = \sum_{i\in\mathcal{I}} \alpha_i \varphi_i$ with $\sum_{i\in\mathcal{I}} \alpha_i^2 = 1$. Then we have

$$\|T_z\|^2 = \sup_{\|\boldsymbol{\alpha}\|_{\ell^2}=1} \Big\| T_z \sum_{i\in\mathcal{I}} \alpha_i \varphi_i \Big\|_{\mathscr{Y}}^2 = \sup_{\|\boldsymbol{\alpha}\|_{\ell^2}=1} \Big\| \sum_{i\in\mathcal{I}} \alpha_i T_z \varphi_i \Big\|_{\mathscr{Y}}^2 = \sup_{\|\boldsymbol{\alpha}\|_{\ell^2}=1} \sum_{i\in\mathcal{I}} \alpha_i^2 \sigma_i^2 = \sigma_1^2. \qquad \square$$

2. Random fields

The natural environment for the consideration of random fields are the so called *Lebesgue-Bochner spaces*. These spaces quantify the integrability of Banach space valued functions and have originally been introduced in [Boc33]. In this section, we want to provide some facts and results on Lebesgue-Bochner spaces. For more details on this topic, we refer to the works [AE08, Alt07, DU77, HP57, LC85]. Especially in [AE08, Alt07], the Lebesgue spaces L^p are defined in a rather abstract fashion for Banach space valued functions and thus coincide with our conception of Lebesgue-Bochner spaces. We will collect here results from these works but directly modify them for probability spaces. In the sequel, we will consider both, random scalar fields and random vector fields. Thus, in this section, we will introduce the underlying spaces for both cases.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a complete and separable probability space with σ-algebra $\mathcal{F}$ and probability measure $\mathbb{P}$. Here, complete means that $\mathcal{F}$ contains all $\mathbb{P}$-null sets. The separability is e.g. obtained if $\mathcal{F}$ is countably generated, cf. [Hal76, Theorem 40.B].

Furthermore, let $(\mathscr{B}, \|\cdot\|_{\mathscr{B}})$ be a Banach space over $\mathbb{R}$. Its Borel σ-algebra, which is defined with respect to the open sets of the metric induced by $\|\cdot\|_{\mathscr{B}}$, is called $\mathcal{B}$. We start by specifying measurability for functions $u\colon \Omega \to \mathscr{B}$.

(2.1) **Definition.** A function $u\colon \Omega \to \mathscr{B}$ is called strongly $\mathbb{P}$-measurable if for any $O \in \mathcal{B}$ it holds that $u^{-1}(O) \in \mathcal{F}$ and if there exists a $\mathbb{P}$-null set N such that $u(\Omega \setminus N)$ is separable.

Notice that the second part of the definition is automatically satisfied if $\mathscr{B}$ is a separable space itself. Equivalently to the definition, there exists a sequence of simple functions $u_n = \sum_{i=1}^n x_i \chi_{A_i}$, where χ_{A_i} is the characteristic function of the set $A_i := u_n^{-1}(x_i)$, such that

$$\lim_{n\to\infty} \|u_n(\omega,\cdot) - u(\omega,\cdot)\|_{\mathscr{B}} = 0 \quad \mathbb{P}\text{-almost everywhere},$$

cf. [LC85, Lemmata 10.1, 10.3, 10.5]. The following lemma indicates that $\|u(\omega,\cdot)\|_{\mathscr{B}}$ is a random variable if $u\colon \Omega \to \mathscr{B}$ is strongly $\mathbb{P}$-measurable.

(2.2) **Lemma.** Let $u\colon \Omega \to \mathscr{B}$ be strongly $\mathbb{P}$-measurable. Then, $\|u(\omega,\cdot)\|_{\mathscr{B}}\colon \Omega \to \mathbb{R}$ is a measurable function in the classical sense.

Proof. For a proof of this result, see [HP57, Theorem 3.5.2]. □

Thus, we may now define the Lebesgue-Bochner spaces as follows.

(2.3) **Definition.** For $p \geqslant 0$, the *Lebesgue-Bochner space* $L^p_{\mathbb{P}}(\Omega;\mathscr{B})$ consists of all equivalence classes of strongly $\mathbb{P}$-measurable maps $u\colon \Omega \to \mathscr{B}$ with finite norm

$$\|u\|_{L^p_{\mathbb{P}}(\Omega;\mathscr{B})} := \begin{cases} \left(\int_\Omega \|u(\omega,\cdot)\|^p_{\mathscr{B}} \,\mathrm{d}\mathbb{P}\right)^{1/p}, & p < \infty \\ \operatorname*{ess\,sup}_{\omega\in\Omega} \|u(\omega,\cdot)\|_{\mathscr{B}}, & p = \infty. \end{cases} \tag{2.4}$$

Here, $\int_\Omega \cdot \,\mathrm{d}\mathbb{P}$ denotes the standard integral for $\mathbb{R}$-valued measurable maps. Furthermore, $u, v\colon \Omega \to \mathscr{B}$ are identified if they coincide $\mathbb{P}$-almost everywhere, i.e. if $\mathbb{P}[\{u \neq v\}] = 0$.

The spaces $L^p_{\mathbb{P}}(\Omega;\mathscr{B})$ are for all $p \in [1,\infty]$ complete with respect to the norm defined in (2.4) and thus Banach spaces, see e.g. [AE08] for a proof of this statement. Notice that that $L^2_{\mathbb{P}}(\Omega)$ is separable if $(\Omega,\mathcal{F},\mathbb{P})$ is separable, cf. [Hal76, Exercise 43.(1)]. Thus, if $p = 2$ and $\mathscr{B} = (\mathscr{H},(\cdot,\cdot)_{\mathscr{H}})$ is a separable Hilbert spaces, then $L^2_{\mathbb{P}}(\Omega;\mathscr{H})$ is also a separable Hilbert space equipped with the inner product

$$(u,v)_{L^2_{\mathbb{P}}(\Omega;\mathscr{H})} := \int_\Omega \big(u(\omega,\cdot), v(\omega,\cdot)\big)_{\mathscr{H}} \,\mathrm{d}\mathbb{P}.$$

In particular, it holds $L^2_{\mathbb{P}}(\Omega;\mathscr{H}) \cong L^2_{\mathbb{P}}(\Omega) \otimes \mathscr{H}$, cf. [RS80, Theorem II.10].

(2.5) **Definition.** A strongly $\mathbb{P}$-measurable map $u\colon \Omega \to \mathscr{B}$ is *Bochner integrable* if there exists a sequence of simple functions $\{u_n\}_n$ such that

$$\lim_{n\to\infty} \int_\Omega \|u_n(\omega,\cdot) - u(\omega,\cdot)\|_{\mathscr{B}} \,\mathrm{d}\mathbb{P} = 0.$$

In this case, we define for a set $A \in \mathcal{F}$ the *Bochner integral*

$$\int_A u(\omega,\cdot) \,\mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_A u_n(\omega,\cdot) \,\mathrm{d}\mathbb{P}.$$

Especially for $p = 1$, the space $L^1_{\mathbb{P}}(\Omega;\mathscr{B})$ coincides with the space of *Bochner integrable* functions, cf. [DU77, Theorem 2.4]. We summarize some important facts about the Bochner integral.

(2.6) **Theorem.**

(a) The Bochner integral $\int_\Omega \cdot \,\mathrm{d}\mathbb{P}\colon \Omega \to \mathscr{B}$ is a linear map.

(b) For $u \in L^1_{\mathbb{P}}(\Omega;\mathscr{B})$ it holds $\|\int_A u(\omega,\cdot)\,\mathrm{d}\mathbb{P}\|_{\mathscr{B}} \leqslant \int_A \|u(\omega,\cdot)\|_{\mathscr{B}} \,\mathrm{d}\mathbb{P}$ for all $A \in \mathcal{F}$.

(c) Let $\{u_n\}_n$ be a sequence of Bochner integrable functions with $\lim_{n\to\infty} u_n = u$ in $\mathbb{P}$-measure and g a Lebesgue integrable function on Ω such that $\|u_n\| \leqslant g$ $\mathbb{P}$-almost everywhere. Then, u is Bochner integrable and $\lim_{n\to\infty} \int_A u_n \,\mathrm{d}\mathbb{P} = \int_A u \,\mathrm{d}\mathbb{P}$ for all $A \in \mathcal{F}$. Moreover, it holds $\lim_{n\to\infty} \int_\Omega \|u_n - u\|_{\mathscr{B}} \,\mathrm{d}\mathbb{P} = 0$.

(d) Let $T\colon U \to \mathscr{C}$ be a closed linear operator for some Banach space $\mathscr{C}$ and $U \subseteq \mathscr{B}$. If u and Tu are Bochner integrable, then $T(\int_A u \,\mathrm{d}\mathbb{P}) = \int_A Tu \,\mathrm{d}\mathbb{P}$ for all $A \in \mathcal{F}$.

Proof. The statement (a) is shown in [HP57, Theorem 3.7.5]. For a proof of (b), see [HP57, Theorem 3.7.6]. The result (c) is proven in [HP57, Theorem 3.7.9] and finally, a proof of (d) is given by [HP57, Theorem 3.7.12]. □

Additionally, we have an analogue to Fubini's theorem in case of Bochner integrals.

(2.7) **Theorem.** Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be two σ-finite measure spaces and $u \in L^1_{\mu_1\times\mu_2}(\Omega_1 \times \Omega_2; \mathscr{B})$, where $\mu_1 \times \mu_2$ denotes the product measure on the product σ-algebra $\mathcal{F}_1 \times \mathcal{F}_2$. Then, $u(\omega_1, \cdot)\colon \Omega_2 \to \mathscr{B}$ is Bochner integrable for μ_1-almost every $\omega_1 \in \Omega_1$ and $u(\cdot, \omega_2)\colon \Omega_1 \to X$ is Bochner integrable for μ_2-almost every $\omega_2 \in \Omega_2$. Furthermore, it holds

$$\int_{\Omega_1\times\Omega_2} u \,\mathrm{d}(\mu_1 \times \mu_2) = \int_{\Omega_1}\int_{\Omega_2} u(\omega_1, \omega_2)\,\mathrm{d}\mu_2\,\mathrm{d}\mu_1 = \int_{\Omega_2}\int_{\Omega_1} u(\omega_1, \omega_2)\,\mathrm{d}\mu_1\,\mathrm{d}\mu_2.$$

Proof. A proof of this theorem, can be found in e.g. [HP57, Theorem 3.7.13]. □

Consider a sufficiently smooth domain $D \subset \mathbb{R}^{d_1}$ and let $\mathscr{B} = L^2(D; \mathbb{R}^{d_2})$, where we equip $L^2(D; \mathbb{R}^{d_2}) = [L^2(D)]^{d_2}$ with the inner product

$$(\mathbf{u}, \mathbf{v})_{L^2(D;\mathbb{R}^{d_2})} := \int_D \langle \mathbf{u}, \mathbf{v}\rangle \,\mathrm{d}\mathbf{x} \quad \text{for all } \mathbf{u}, \mathbf{v} \in L^2(D; \mathbb{R}^{d_2}).$$

Here, $\langle\cdot,\cdot\rangle$ denotes the canonical inner product in $\mathbb{R}^{d_2}$. In the case of random scalar fields, we have $d_2 = 1$. For random vector fields, we will especially consider $d_1 = d_2 = d$. There exists the following relationship between the spaces under consideration. It holds

$$L^2_{\mathbb{P}}(\Omega) \otimes L^2(D; \mathbb{R}^{d_2}) \cong L^2_{\mathbb{P}}(\Omega; L^2(D; \mathbb{R}^{d_2})) \cong L^2_{\mathbb{P}\times\lambda}(\Omega \times D; \mathbb{R}^{d_2}),$$

where each relation holds by an isometric isomorphism, cf. [LC85, Theorem 1.39].

Now, let $\mathbf{u} \in L^2_{\mathbb{P}}(\Omega; L^2(D; \mathbb{R}^{d_2}))$ be represented according to

$$\mathbf{u}(\omega, \mathbf{x}) = [u_1(\omega, \mathbf{x}), \dots, u_{d_2}(\omega, \mathbf{x})]^\intercal.$$

Then, we can define the *mean* of $\mathbf{u}$ in terms of the Bochner integral

$$\mathbb{E}[\mathbf{u}](\mathbf{x}) := \int_\Omega \mathbf{u}(\omega, \mathbf{x}) \,\mathrm{d}\mathbb{P}(\Omega) \in L^2(D; \mathbb{R}^{d_2}).$$

Especially, it holds $\mathbb{E}[u_i](\mathbf{x}) = \int_\Omega u_i(\omega, \mathbf{x}) \,\mathrm{d}\mathbb{P}(\Omega)$. By identifying $\mathbf{u}$ with its representative in $L^2_{\mathbb{P}}(\Omega) \otimes L^2(D; \mathbb{R}^{d_2})$, this definition coincides with $\mathbb{E}[\mathbf{u}] = T_{\mathbf{u}}1$. With respect to the *centered* random field

$$\mathbf{u}_0 = \mathbf{u} - \mathbb{E}[\mathbf{u}],$$

we introduce the (matrix valued) *covariance function* of $\mathbf{u}$ according to

$$\mathrm{Cov}[\mathbf{u}](\mathbf{x},\mathbf{y}) = [\mathrm{Cov}_{i,j}(\mathbf{x},\mathbf{y})]_{i,j=1}^{d_2}$$

with

$$\mathrm{Cov}_{i,j}(\mathbf{x},\mathbf{y}) = \mathbb{E}[u_{0,i}(\omega,\mathbf{x})u_{0,j}(\omega,\mathbf{y})]. \tag{2.8}$$

The boundedness of $\mathrm{Cov}_{i,j}(\mathbf{x},\mathbf{y})$ in $L^2(D\times D)$ follows from the Cauchy-Schwarz inequality and the application of Fubini's theorem. Since $\mathrm{Cov}_{i,j}(\mathbf{x},\mathbf{y}) \in L^2(D\times D)$ holds, we conclude $\mathrm{Cov}[\mathbf{u}](\mathbf{x},\mathbf{y}) \in L^2(D\times D;\mathbb{R}^{d_2\times d_2})$, where we consider the space $\mathbb{R}^{d_2\times d_2}$ to be endowed with the inner product

$$\mathbf{A}:\mathbf{B} := \sum_{i,j=1}^{d_2} a_{i,j}b_{i,j} \quad \text{for } \mathbf{A},\mathbf{B}\in\mathbb{R}^{d_2\times d_2} \quad \text{with } \mathbf{A}=[a_{i,j}]_{i,j=1}^{d_2},\ \mathbf{B}=[b_{i,j}]_{i,j=1}^{d_2}.$$

The related norm $\|\mathbf{A}\|_F := \sqrt{\mathbf{A}:\mathbf{A}}$ is the Frobenius norm. The inner product particularly induces the inner product on $L^2(D\times D;\mathbb{R}^{d_2\times d_2})$ given by

$$(\mathbf{A},\mathbf{B})_{L^2(D\times D;\mathbb{R}^{d_2\times d_2})} := \int_D\int_D \mathbf{A}:\mathbf{B}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y} \quad \text{for } \mathbf{A},\mathbf{B}\in L^2(D\times D;\mathbb{R}^{d_2\times d_2}).$$

By identifying $\mathbf{u}_0$ with its representative in $L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^{d_2})$, we observe that $\mathrm{Cov}[\mathbf{u}] \in L^2(D;\mathbb{R}^{d_2})\otimes L^2(D;\mathbb{R}^{d_2})$ is exactly the element corresponding to the trace-class operator $T_{\mathbf{u}_0}T_{\mathbf{u}_0}^\star$. More precisely, for the orthogonal decomposition of $\mathbf{u}_0$ in accordance with (1.11), i.e. $\mathbf{u}_0 = \sum_{i\in\mathcal{I}}\sigma_i X_i\otimes\boldsymbol{\varphi}_i$, where $\{X_i\}_{i\in\mathcal{I}}\subset L^2_{\mathbb{P}}(\Omega)$ and $\{\boldsymbol{\varphi}_i\}_{k\in\mathcal{I}}\subset L^2(D;\mathbb{R}^{d_2})$ are orthonormal families, it holds that

$$T_{\mathbf{u}_0}T_{\mathbf{u}_0}^\star\mathbf{v} = \sum_{i\in\mathcal{I}}\sigma_i\Big(\sum_{j\in\mathcal{I}}\sigma_j(\mathbf{v},\boldsymbol{\varphi}_j)_{L^2(D;\mathbb{R}^{d_2})}X_j, X_i\Big)_{L^2_{\mathbb{P}}(\Omega)}\boldsymbol{\varphi}_i = \sum_{i\in\mathcal{I}}\sigma_i^2(\mathbf{v},\boldsymbol{\varphi}_i)_{L^2(D;\mathbb{R}^{d_2})}\boldsymbol{\varphi}_i.$$

This implies the identity

$$\mathrm{Cov}[\mathbf{u}] = \sum_{i\in\mathcal{I}}\sigma_i^2\boldsymbol{\varphi}_i\otimes\boldsymbol{\varphi}_i.$$

For $\lim_{\mathbf{y}\to\mathbf{x}}\mathrm{Cov}[\mathbf{u}](\mathbf{x},\mathbf{y})$, we especially obtain the *variance*

$$\mathbb{V}[\mathbf{u}](\mathbf{x}) := \int_\Omega \mathbf{u}_0^2(\omega,\mathbf{x})\,\mathrm{d}\mathbb{P}(\Omega) - \big(\mathbb{E}[\mathbf{u}](\mathbf{x})\big)^2 \in L^2(D;\mathbb{R}^{d_2}).$$

Finally, in order to approximate random fields in $L^2_{\mathbb{P}}(\Omega;L^2(D;\mathbb{R}^{d_2}))$, we have to provide additional regularity with respect to the spatial variable $\mathbf{x}$ in terms of Sobolev smoothness. To that end, we define the Sobolev spaces $H^q(D;\mathbb{R}^{d_2}) := [H^q(D)]^{d_2}$ for $q>0$ with respect to the inner product

$$(\mathbf{u},\mathbf{w})_{H^q(D;\mathbb{R}^{d_2})} := \sum_{|\boldsymbol{\alpha}|\leqslant q}\int_D \langle\partial^{\boldsymbol{\alpha}}\mathbf{u},\partial^{\boldsymbol{\alpha}}\mathbf{w}\rangle\,\mathrm{d}\mathbf{x}$$

for $q\in\mathbb{N}$ and

$$(\mathbf{u},\mathbf{w})_{H^q(D;\mathbb{R}^{d_2})} := (\mathbf{u},\mathbf{w})_{H^{\lfloor q\rfloor}(D;\mathbb{R}^{d_2})} + \sum_{|\boldsymbol{\alpha}|=\lfloor q\rfloor}\int_D\int_D \frac{\|\partial^{\boldsymbol{\alpha}}\mathbf{u}(\mathbf{x})-\partial^{\boldsymbol{\alpha}}\mathbf{w}(\mathbf{y})\|_2^2}{\|\mathbf{x}-\mathbf{y}\|_2^{d_1+2s}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}\mathbf{y}$$

for $q=\lfloor q\rfloor+s$ with $s\in(0,1)$. Its dual space with respect to the L^2-duality pairing is denoted as $\tilde{H}^{-q}(D;\mathbb{R}^{d_2})$.

(2.9) **Remark.** The sequence of inclusions $H^q(D;\mathbb{R}^{d_2}) \subset L^2(D;\mathbb{R}^{d_2}) \subset \big[H^q(D;\mathbb{R}^{d_2})\big]'$ forms a *Gelfand triple*, cf. [Wlo82, Definition 17.1]. The completion of $L^2(D;\mathbb{R}^{d_2})$ with respect to the norm

$$\|\mathbf{u}\|_{\tilde{H}^{-q}(D;\mathbb{R}^{d_2})} := \sup_{\mathbf{0}\neq\mathbf{v}\in H^q(D;\mathbb{R}^{d_2})} \frac{(\mathbf{v},\mathbf{u})_{L^2(D;\mathbb{R}^{d_2})}}{\|\mathbf{v}\|_{H^q(D;\mathbb{R}^{d_2})}}, \quad \mathbf{u}\in L^2(D;\mathbb{R}^{d_2}),$$

is $\tilde{H}^{-q}(D;\mathbb{R}^{d_2})$ and it holds $\big[H^q(D;\mathbb{R}^{d_2})\big]' = \tilde{H}^{-q}(D;\mathbb{R}^{d_2})$ due to [Wlo82, Theorem 17.3]. Especially, $(\cdot,\cdot)_{L^2(D;\mathbb{R}^{d_2})}$ extends to a bilinear form on $H^q(D;\mathbb{R}^{d_2}) \times \tilde{H}^{-q}(D;\mathbb{R}^{d_2})$. In the sequel, this bilinear form will also be denoted by $(\cdot,\cdot)_{L^2(D;\mathbb{R}^{d_2})}$.

Now, in particular, we obtain a generalized Cauchy-Schwarz inequality.

(2.10) **Lemma.** Let $\mathbf{u}\in H^q(D;\mathbb{R}^{d_2})$ and $\mathbf{f}\in\tilde{H}^{-q}(D;\mathbb{R}^{d_2})$. Then, there holds

$$(\mathbf{u},\mathbf{f})_{L^2(D;\mathbb{R}^{d_2})} \leqslant \|\mathbf{u}\|_{H^q(D;\mathbb{R}^{d_2})}\|\mathbf{f}\|_{\tilde{H}^{-q}(D;\mathbb{R}^{d_2})}$$

Proof. The case $\mathbf{u}=\mathbf{0}$ is trivial. Hence let $\mathbf{u}\neq\mathbf{0}$. We conclude

$$\begin{aligned}(\mathbf{u},\mathbf{f})_{L^2(D;\mathbb{R}^{d_2})} &= \|\mathbf{u}\|_{H^q(D;\mathbb{R}^{d_2})}\frac{(\mathbf{u},\mathbf{f})_{L^2(D;\mathbb{R}^{d_2})}}{\|\mathbf{u}\|_{H^q(D;\mathbb{R}^{d_2})}}\\ &\leqslant \|\mathbf{u}\|_{H^q(D;\mathbb{R}^{d_2})}\sup_{\mathbf{0}\neq\mathbf{v}\in H^q(D;\mathbb{R}^{d_2})}\frac{(\mathbf{v},\mathbf{f})_{L^2(D;\mathbb{R}^{d_2})}}{\|\mathbf{v}\|_{H^q(D;\mathbb{R}^{d_2})}}\\ &= \|\mathbf{u}\|_{H^q(D;\mathbb{R}^{d_2})}\|\mathbf{f}\|_{\tilde{H}^{-q}(D;\mathbb{R}^{d_2})}.\end{aligned}$$

□

3. The Karhunen-Loève expansion

In order to make the random (vector-) field $\mathbf{u}(\omega,\mathbf{x})\in L^2_{\mathbb{P}}(\Omega;L^2(D;\mathbb{R}^{d_2}))$ feasible for numerical computations, we consider here its *Karhunen-Loève expansion*, cf. [Loè77]. This representation is easily obtained with the machinery provided in the first section. Since we may identify $L^2_{\mathbb{P}}(\Omega;L^2(D;\mathbb{R}^{d_2})) \cong L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^{d_2})$, we already know that $\mathbf{u}_0(\omega,\mathbf{x})$ exhibits the orthogonal decomposition

$$\mathbf{u}_0 = \sum_{i\in\mathcal{I}} \sigma_i X_i \otimes \boldsymbol{\varphi}_i,$$

where $\{\boldsymbol{\varphi}_i\}_{i\in\mathcal{I}}\subset L^2(D;\mathbb{R}^{d_2})$ and $\{X_i\}_{i\in\mathcal{I}}\subset L^2_{\mathbb{P}}(\Omega)$ are orthonormal families. With respect to the canonical map

$$L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^{d_2}) \to L^2_{\mathbb{P}}(\Omega;L^2(D;\mathbb{R}^{d_2})), \quad X\otimes\varphi \mapsto X(\omega)\varphi(\mathbf{x}),$$

we end up with the following

(3.1) **Definition.** Let $\mathbf{u}(\omega,\mathbf{x})$ be a vector field in $L^2_{\mathbb{P}}(\Omega;L^2(D;\mathbb{R}^{d_2}))$. The expansion

$$\mathbf{u}(\omega,\mathbf{x}) = \mathbb{E}[\mathbf{u}](\mathbf{x}) + \sum_{i\in\mathcal{I}} \sigma_i X_i(\omega)\boldsymbol{\varphi}_i(\mathbf{x})$$

is called *Karhunen-Loève expansion* of $\mathbf{u}(\omega,\mathbf{x})$.

(3.2) **Remark.** The knowledge of the random vector field $\mathbf{u}(\omega, \mathbf{x})$ is sufficient to compute the related Karhunen-Loève expansion. This is achieved in complete analogy to the construction of the representation (1.11), i.e. by solving the eigenvalue problem for the trace-class operator $T^\star_{\mathbf{u}_0} T_{\mathbf{u}_0}$ associated with $\mathbf{u}_0 = \mathbf{u} - \mathbb{E}[\mathbf{u}]$. In practice, however, the random field is often only provided in terms of its (empirical) mean $\mathbb{E}[\mathbf{u}]$ and its (empirical) covariance function $\mathrm{Cov}[\mathbf{u}]$. In this case, the orthogonal basis in $L^2_{\mathbb{P}}(\Omega)$ is only determined up to isometry since $T_{\mathrm{Cov}[\mathbf{u}]} = T_{\mathbf{u}_0} T^\star_{\mathbf{u}_0} = (T_{\mathbf{u}_0} U)(T_{\mathbf{u}_0} U)^\star$ for any isometry $U \colon L^2_{\mathbb{P}}(\Omega) \to L^2_{\mathbb{P}}(\Omega)$, see also Remark (1.19). In this situation, the law of the random variables $\{X_i\}_{i \in \mathcal{I}}$ has to be approximated appropriately, e.g. by a *maximum likelihood estimate*, cf. [ST06].

We impose some common assumptions on the properties of the Karhunen-Loève expansion.

(3.3) **Assumption.**

(1) The random variables $\{X_i\}_{i \in \mathcal{I}}$ are centered, i.e. $\mathbb{E}[X_i] = 0$, and take values in $[-1, 1]$ for all $i \in \mathcal{I}$ and almost every $\omega \in \Omega$.

(2) The random variables $\{X_i\}_{i \in \mathcal{I}}$ are independent and identically distributed.

(3) The sequence

$$\{\gamma_i\}_i := \{\|\sigma_i \boldsymbol{\varphi}_i\|_{W^{1,\infty}(D;\mathbb{R}^{d_2})}\}_i \tag{3.4}$$

is at least in $\ell^1(\mathcal{I})$. We denote its norm by

$$c_\gamma := \sum_{i \in \mathcal{I}} \gamma_i. \tag{3.5}$$

Here, and in the following, we shall equip the space $W^{1,\infty}(D; \mathbb{R}^{d_2})$ with the equivalent norm

$$\|\mathbf{v}\|_{W^{1,\infty}(D;\mathbb{R}^{d_2})} = \max\{\|\mathbf{v}\|_{L^\infty(D;\mathbb{R}^{d_2})}, \|\mathbf{v}'\|_{L^\infty(D;\mathbb{R}^{d_2 \times d_2})}\},$$

where $\mathbf{v}'$ denotes the Jacobian of $\mathbf{v}$ and

$$\|\mathbf{v}'\|_{L^\infty(D;\mathbb{R}^{d_2 \times d_2})} := \operatorname*{ess\,sup}_{\mathbf{x} \in D} \|\mathbf{v}'(\mathbf{x})\|_2.$$

In the last expression, $\|\cdot\|_2$ corresponds to the usual 2-norm of matrices, i.e. the largest singular value. Notice that this norm is equivalent to the Frobenius norm defined earlier. Nevertheless, in the subsequent error estimates, the 2-norm provides smaller constants. Regard moreover that the norm $\|\cdot\|_{L^\infty(D_{\mathrm{ref}};\mathbb{R}^{d_2 \times d_2})}$ is consistent in the following way: for $\mathbf{v} \in L^\infty(D_{\mathrm{ref}}; \mathbb{R}^{d_2})$ and $\mathbf{M} \in L^\infty(D_{\mathrm{ref}}; \mathbb{R}^{d_2 \times d_2})$ it holds

$$\|\mathbf{M}\mathbf{v}\|_{L^\infty(D_{\mathrm{ref}};\mathbb{R}^{d_2})} \leqslant \|\mathbf{M}\|_{L^\infty(D_{\mathrm{ref}};\mathbb{R}^{d_2 \times d_2})} \|\mathbf{v}\|_{L^\infty(D_{\mathrm{ref}};\mathbb{R}^{d_2})}. \tag{3.6}$$

4. The Matérn class of covariance functions

Later on, for our numerical experiments, we assume that the random (vector-) field $\mathbf{u}(\omega, \mathbf{x})$ is described in terms of its mean $\mathbb{E}[\mathbf{u}]$ and its covariance function $\mathrm{Cov}[\mathbf{u}]$. Thus, we have especially to prescribe the scalar valued covariance functions $\mathrm{Cov}_{i,j}$ for $i, j = 1, \dots, d_2$, cf. (2.8). To that end, we consider a special class of covariance functions, namely the *Matérn class*, cf. [Mat86]. They are very often used as covariance functions for the definition of stochastic fields in applications. In accordance with [RW05], they are defined as follows.

(4.1) **Definition.** Let $r := \|\mathbf{x} - \mathbf{y}\|_2$ and $\ell \in (0, \infty)$. Then, the *Matérn covariance function* of order $\nu > 0$ is given by

$$k_\nu(r) := \frac{2^{1-\nu}}{\Gamma(\nu)} \Big(\frac{\sqrt{2\nu} r}{\ell} \Big)^\nu K_\nu \Big(\frac{\sqrt{2\nu} r}{\ell} \Big). \tag{4.2}$$

Here, Γ denotes the gamma function and K_ν denotes the modified Bessel function of the second kind of order ν, cf. [AS64].

The expression (4.2) simplifies if $\nu = p + 1/2$ with $p \in \mathbb{N}$. In this case, [RW05] provides

$$k_{p+1/2}(r) = \exp\Big(- \frac{\sqrt{2\nu} r}{\ell} \Big) \frac{p!}{(2p)!} \sum_{i=0}^{p} \frac{(p+i)!}{i!(p-i)!} \Big(\frac{\sqrt{8\nu} r}{\ell} \Big)^{p-i}.$$

Especially, we deduce

$$\begin{aligned}
&\nu = \frac{1}{2}, && k_{1/2}(r) = \exp\Big(- \frac{r}{\ell} \Big), \\
&\nu = \frac{3}{2}, && k_{3/2}(r) = \Big(1 + \frac{\sqrt{3} r}{\ell} \Big) \exp\Big(- \frac{\sqrt{3} r}{\ell} \Big), \\
&\nu = \frac{5}{2}, && k_{5/2}(r) = \Big(1 + \frac{\sqrt{5} r}{\ell} + \frac{5 r^2}{3\ell^2} \Big) \exp\Big(- \frac{\sqrt{5} r}{\ell} \Big), \\
&\nu = \frac{7}{2}, && k_{7/2}(r) = \Big(1 + \frac{\sqrt{7} r}{\ell} + \frac{14 r^2}{5\ell^2} + \frac{49\sqrt{7} r^3}{15\ell^3} \Big) \exp\Big(- \frac{\sqrt{7} r}{\ell} \Big), \\
&\nu = \frac{9}{2}, && k_{9/2}(r) = \Big(1 + \frac{3r}{\ell} + \frac{27 r^2}{7\ell^2} + \frac{18 r^3}{7\ell^3} + \frac{27 r^4}{35\ell^3} \Big) \exp\Big(- \frac{3r}{\ell} \Big), \\
&\nu = \infty, && k_\infty(r) = \exp\Big(- \frac{r^2}{2\ell^2} \Big).
\end{aligned} \tag{4.3}$$

A visualization of these kernels for different values of ν is given in Figure II.2. Obviously, the Sobolev smoothness of the kernel k_ν is controlled by the *smoothness parameter* ν.

For increasing values of ν, the respective kernel function k_ν exhibits successively more regularity. Especially, the eigenvalues of the Matérn correlation kernels decay like

$$\lambda_m \leqslant C m^{-(1+\frac{2\nu}{d})} \tag{4.4}$$

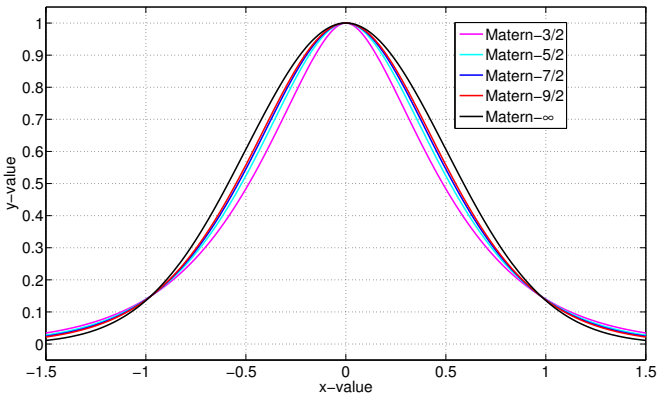

Figure II.2: Visualization of $k_\nu(|x|)$ for different values of the smoothness parameter ν.

for some $C > 0$, cf. [GKN+13]. Thus, since the decay of the covariance operator's eigenvalues is known in advance, they are very well suited for numerical examples.

Obviously, the Matérn kernels provide rotational symmetry, i.e. they are invariant under isometries of D, since they are only dependent on the particular distance of the points $\mathbf{x}$ and $\mathbf{y}$. Thus, we obtain analytic expressions for the eigenvalues of the underlying Hilbert-Schmidt operators if we choose $D = \mathbb{S}^{d-1}$ to be the unit sphere in $\mathbb{R}^d$. More precisely, we may apply the Funk-Hecke formula, cf. [Mül98], which reads as follows.

(4.5) **Theorem.** Let $\mathbf{x} \in \mathbb{S}^{d-1}$ and $f \in C([-1,1])$, then it holds

$$\int_{\mathbb{S}^{d-1}} f(\mathbf{x}^\intercal \mathbf{y}) Y_m(\mathbf{y}) \, \mathrm{d}\sigma_{\mathbf{y}} = \lambda_m Y_m(\mathbf{x})$$

with

$$\lambda_m = |\mathbb{S}^{d-2}| \int_{-1}^{1} P_m(d;t) f(t) (1-t^2)^{\frac{d-3}{2}} \, \mathrm{d}t.$$

Here, Y_m corresponds to a spherical harmonic of order m and $P_m(d;t)$ denotes the polynomial

$$P_m(d;t) := m! \, \Gamma\left(\frac{d-1}{2}\right) \sum_{i=0}^{\lfloor m/2 \rfloor} \left(\frac{-1}{4}\right)^i \frac{(1-t^2)^i t^{m-2i}}{i!(m-2i)!\Gamma(i+\frac{d-1}{2})}.$$

Proof. A proof of this theorem can be found in [Mül98]. □

Especially, for the case $d = 3$, the polynomials $P_m(3;t)$ correspond to the Legendre polynomials, cf. [Mül98]. Moreover, the Funk-Hecke formula applies to all kernel functions on $\mathbb{S}^{d-1}$ which depend only on the Euclidean distance $r(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_2$. This is easily seen due to

$$r(\mathbf{x},\mathbf{y}) = r(\mathbf{x}^\intercal \mathbf{y}) = \sqrt{2 - 2\mathbf{x}^\intercal \mathbf{y}} \quad \text{for all } \mathbf{x},\mathbf{y} \in \mathbb{S}^{d-1}.$$

Figure II.3 shows the distribution of the Matérn-kernels' eigenvalues for the parameters $\nu = 3/2, 5/2, 7/2, 9/2$ on $\mathbb{S}^2$ up to an order of magnitude of 10^{-10} for the correlation length $\ell = 1$. The constant C is estimated by a least-square fit for the ratio of the

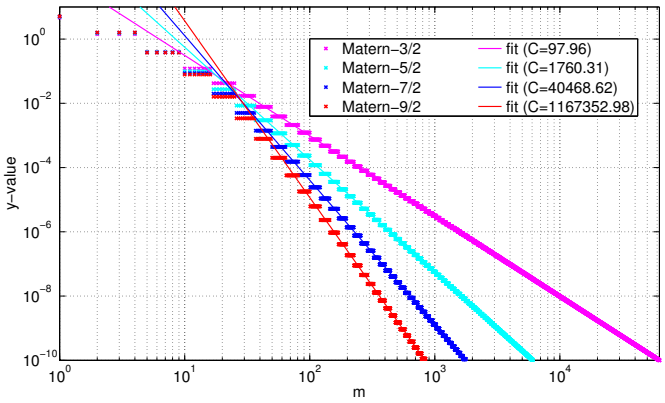

Figure II.3: Decay of the eigenvalues with related fits.

rate given by formula (4.4) for $C = 1$ and the exact eigenvalues given by Theorem (4.5). The obtained values of C for each kernel under consideration are denoted in the legend of Figure II.3. The plot indicates, that the fitted rates perfectly match the asymptotic behavior of the eigenvalues.

Finally, we remark that for continuous covariance functions, we have especially

$$\operatorname{Tr} T_{\operatorname{Cov}[u]} = \int_D \operatorname{Cov}[u](\mathbf{x}, \mathbf{x}) \, \mathrm{d}\mathbf{x}, \tag{4.6}$$

which is a consequence of Mercer's theorem, cf. [Mer09]. Hence, in case of the Matérn covariance functions, we can easily compute the trace related to the covariance operator by computing the measure of D, since

$$\int_D k_\nu(\|\mathbf{x} - \mathbf{x}\|_2) \, \mathrm{d}\mathbf{x} = \int_D k_\nu(0) \, \mathrm{d}\mathbf{x} = \int_D 1 \, \mathrm{d}\mathbf{x}.$$

5. Parametric representation of geometries

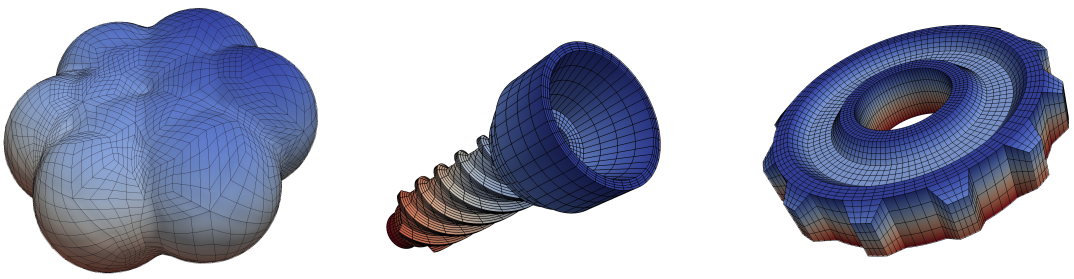

Figure II.4: Different parametric geometries.

In this section and the following one, we introduce a parametric representation for geometries and the related finite element spaces. Since we are considering both, the traditional finite element method for the discretization of partial differential equations on

the domain $D \subset \mathbb{R}^d$ (of interest to us, are the cases $d = 2$ and $d = 3$) and techniques associated to the boundary element method, we will present here the underlying framework for both of them. To that end, we shall consider here both, triangular and quadrangular meshes. In particular, the special variant of the Fast Multipole Method introduced later on heavily relies on quadrangular meshes. We remark that of course any quadrangular mesh can be transformed into a triangular mesh by subdividing each quadrangular element along a diagonal into two triangular elements.

We suppose that D is a Lipschitz domain which is given as a collection of smooth *patches*. More precisely, let $\triangle$ denote the reference simplex in $\mathbb{R}^d$ and $\square = [0,1]^d$ the reference hypercube. We assume that the domain D is partitioned into $K \in \mathbb{N}$ patches

$$\overline{D} = \bigcup_{i=1}^{K} \tau_{i,0},$$

where the intersection $\tau_{i,0} \cap \tau_{i',0}$ consists at most of a lower dimensional face for $i \neq i'$. Herein, it holds $\tau_{i,0} = \boldsymbol{\kappa}_i(\mathcal{M})$ for $\mathcal{M} \in \{\triangle, \square\}$, where $\boldsymbol{\kappa}_i \colon \mathbb{R}^d \to \mathbb{R}^d$ denotes a smooth, i.e. analytic, diffeomorphism for $i = 1, 2, \ldots, K$. Thus, we have especially that

$$\frac{\sup\{\|\boldsymbol{\kappa}_i'(\mathbf{s})\mathbf{x}\|_2 : \mathbf{s} \in \mathcal{M}, \|\mathbf{x}\|_2 = 1\}}{\inf\{\|\boldsymbol{\kappa}_i'(\mathbf{s})\mathbf{x}\|_2 : \mathbf{s} \in \mathcal{M}, \|\mathbf{x}\|_2 = 1\}} \leqslant \rho_i \quad \text{for all } i = 1, \ldots, K \text{ and } \mathcal{M} \in \{\triangle, \square\}, \tag{5.1}$$

where $\boldsymbol{\kappa}_i'$ denotes the Jacobian of $\boldsymbol{\kappa}_i$. Since there are only finitely many patches, we may set $\rho := \max_{i=1}^{K} \rho_i$. Moreover, to obtain a regular mesh, we impose the following *matching condition*: there exists a bijective, affine mapping $\boldsymbol{\Xi} \colon \mathcal{M} \to \mathcal{M}$ for $\mathcal{M} \in \{\triangle, \square\}$ such that for each $\mathbf{x} = \boldsymbol{\kappa}_i(\mathbf{s})$ on a common interface of $\tau_{i,0}$ and $\tau_{i',0}$ it holds $\boldsymbol{\kappa}_{i,0}(\mathbf{s}) = (\boldsymbol{\kappa}_{i',0} \circ \boldsymbol{\Xi})(\mathbf{s})$. This means, the parameterizations $\boldsymbol{\kappa}_i$ and $\boldsymbol{\kappa}_{i'}$ coincide on a common interface except for orientation.

Many of such parametric representations of geometries are available as technical surfaces generated by tools from Computer Aided Design (CAD). The most common geometry representation in CAD is defined by the IGES (Initial Graphics Exchange Specification) standard. Here, the initial CAD object is a solid, bounded by a closed surface that is given as a collection of parametric surfaces which can be trimmed or untrimmed. An untrimmed surface is already a four-sided patch, parameterized over a rectangle. Whereas, a trimmed surface is just a piece of a supporting untrimmed surface, described by boundary curves. There are several representations of the parameterizations including B-splines, NURBS (nonuniform rational B-Splines), surfaces of revolution, and tabulated cylinders [HL89]. Such geometries are also recently studied in *isogeometric analysis*, where finite elements based on B-splines are considered, cf. [HCB05] and the references therein.

(5.2) **Remark.** In the context of parametric boundary element methods, we have the situation that the surface $\Gamma := \partial D$ is represented in exactly this way. In [HR10], an algorithm has been developed to decompose a technical surface, described in the IGES format, into a collection of parameterized four-sided patches, fulfilling all the above requirements. In [HR09, HR11], the algorithm has been extended to molecular surfaces. Figure II.4 visualizes three parameterizations which satisfy the present requirements. Since in this situation Γ is a $(d-1)$-dimensional manifold, we also consider the reference hypercube $\square = [0,1]^{d-1}$ together with the smooth diffeomorphisms $\boldsymbol{\kappa}_i \colon \mathbb{R}^{d-1} \to \mathbb{R}^d$, cf. [HP13] and the references therein.

The proposed parametric representation yields an exact representation of the geometry under consideration, which is in contrast to the common approximation of geometries by simplices. Especially, there is no further approximation step required if the geometry is given in this form. As a result, the rate of convergence is not limited by the accuracy of the geometry approximation.

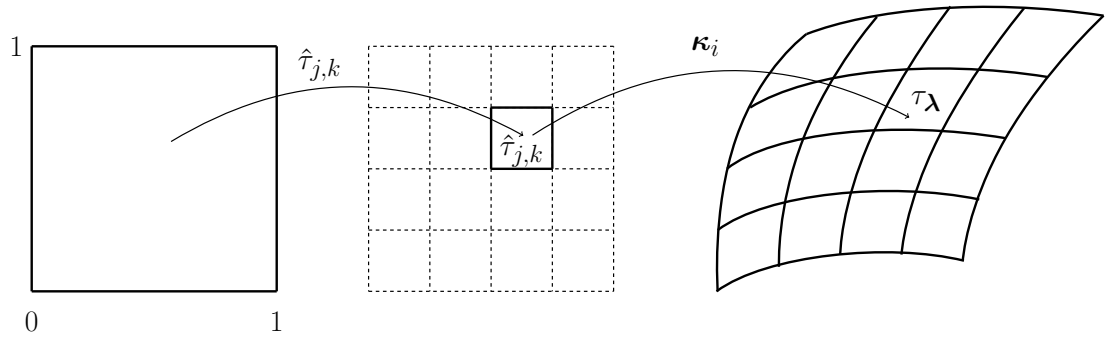

Figure II.5: Localized parameterization

Given a geometry in this fashion, a nested sequence of meshes can be easily constructed. A mesh on level j on D is induced by regular subdivisions of depth j of the reference element $\mathcal{M} \in \{\triangle, \square\}$ into 2^{jd} sub-domains $\mathcal{M}_j = \{\hat{\tau}_{j,k}\}_k$ and lifting the elements in $\mathcal{M}_j$ to D by the diffeomorphisms $\boldsymbol{\kappa}_i$ for $i = 1, \ldots, K$. This generates the

$$N_j = 2^{jd} K \tag{5.3}$$

elements on level j. We will refer to the particular elements as $\tau_{i,j,k} := \boldsymbol{\kappa}_i(\hat{\tau}_{j,k})$, where i is the index of the diffeomorphism $\boldsymbol{\kappa}_i$, j is the level of the element and k is the index of the element in hierarchical order. To simplify the notation we will also denote the triple (i, j, k) by $\boldsymbol{\lambda} := (i, j, k)$ with $|\boldsymbol{\lambda}| := j$. A visualization of the lifting procedure is shown in Figure II.5 for $\mathcal{M} = \square$. Notice that we, here and later on, slightly abuse the notation and also refer to $\hat{\tau}_{j,k}$ as the local element mapping from the reference element to the k-th element in $\mathcal{M}_j$. Finally, we denote the obtained mesh on level j by

$$\mathcal{T}_j := \{\tau_{i,j,k} : i = 1 \ldots, K, k = 0, \ldots, 2^{jd} - 1\}$$

with $\mathcal{T}_0 \subset \mathcal{T}_1 \subset \ldots \subset \mathcal{T}_J$.

In this construction, the local element mappings $\mathcal{M} \to \tau_{i,j,k}$ satisfy for the reference element $\mathcal{M} \in \{\triangle, \square\}$ the same bound (5.1). This result is easily derived by the application of the chain rule. Therefore, especially the uniformity condition for (iso-) parametric finite elements is fulfilled, cf. [Bra07, Len86].

6. Multilevel finite elements

We have now the prerequisites to define nested finite element spaces. We define the finite element ansatz functions on D with respect to the parameterizations $\{\boldsymbol{\kappa}_i\}_{i=1}^K$. To that end,

we lift Lagrangian finite elements from $\mathcal{M} \in \{\triangle, \square\}$ to D by the composition with the inverse mappings $\boldsymbol{\kappa}_i^{-1}$. We will distinguish between piecewise continuous finite elements defined in simplices and piecewise discontinuous finite elements defined on hypercubes.

We begin with the definition of the continuous finite elements. For this purpose, we define on the j-th subdivision $\triangle_j$ of the reference element the standard Lagrangian piecewise polynomial continuous finite elements $\Phi_j = \{\hat{\varphi}_{j,k} : k \in \mathcal{I}_j\}$, where $\mathcal{I}_j$ denotes an appropriate index set. The corresponding finite element spaces on the reference domain are then given by

$$\hat{V}^s_{\triangle,j} = \operatorname{span}\{\hat{\varphi}_{j,k} : k \in \mathcal{I}_j\} = \{u \in C(\triangle) : u|_\tau \in \Pi_s \text{ for all } \tau \in \triangle_j\} \in H^1(\triangle)$$

with $\dim \hat{V}^s_{\triangle,j} \eqsim 2^{jd}$ and Π_s denoting the space of polynomials of total degree at most s. Continuous basis functions whose support overlaps with several patches are obtained by gluing across patch boundaries, using the C^0 inter-patch compatibility. This yields a (nested) sequence of finite element spaces

$$V^s_{\triangle,j} := \{\varphi \circ \boldsymbol{\kappa}_i^{-1} \in C(D) : \varphi \in \hat{V}^s_{\triangle,j},\ i = 1, \dots, K\} \subset H^1(D)$$

with $\dim V^s_{\triangle,j} \eqsim N_j$, cf. (5.3). A basis of this space is given by the functions

$$\varphi_{j,k} = \begin{cases} (\hat{\varphi}_{j,k} \circ \boldsymbol{\kappa}_i^{-1})(\mathbf{x}), & \text{if } \mathbf{x} \in \tau_{i,0} \\ 0, & \text{otherwise,} \end{cases}$$

where $\hat{\varphi}_{j,k} \in \Phi_j$.

Analogously, for hypercubes on the j-th subdivision $\square_j$ of the reference element, we define the finite element spaces

$$\hat{V}^s_{\square,j} := \{u \colon \square \to \mathbb{R} : u|_\tau \in \mathcal{Q}_s \text{ for all } \tau \in \square_j\} \subset L^2(\square)$$

with $\dim \hat{V}^s_{\square,j} \eqsim 2^{jd}$ and $\mathcal{Q}_s$ denoting the space of tensor product polynomials of degree at most s. Then, the related ansatz space $V^s_{\square,j}$ on level j is given by

$$V^s_{\square,j} := \{\varphi \circ \boldsymbol{\kappa}_i^{-1} : \varphi \in \hat{V}^s_{\square,j},\ i = 1, \dots, K\} \subset L^2(D).$$

Both constructions of the finite element spaces obviously result in a nested sequence

$$V_0 \subset V_1 \subset \cdots \subset V_J \subset H^t(D), \tag{6.1}$$

where the Sobolev smoothness t depends on the global smoothness of the functions $v_j \in V^s_j$. Especially, for transported piecewise constant functions ($s = 0$), we have $t < 1/2$ and, for globally continuous, transported piecewise linear functions ($s = 1$), we have $t < 3/2$.

It is well known that the spaces $V^s_{\mathcal{M},j}$ for $\mathcal{M} \in \{\triangle, \square\}$ satisfy the following Jackson and Bernstein type estimates, see e.g. [DeV98], for all $t \le q \le s + 1$, $0 \le r \le t$

$$\inf_{v_j \in V^s_{\mathcal{M},j}} \|u - v_j\|_{H^t(D)} \lesssim h_j^{q-t} \|u\|_{H^q(D)}, \quad u \in H^q(D), \tag{6.2}$$

and

$$\|v_j\|_{H^t(D)} \lesssim h_j^{r-t} \|v_j\|_{H^r(D)}, \quad v_j \in V^s_{\mathcal{M},j}, \tag{6.3}$$

uniformly in j, where we set $h_j := 2^{-j}$.

(6.4) **Remark.** For a proof of these statements, cf. [Sch98, Theorems 6.1.1, 6.1.2]. The proof can be sketched as follows: Since the $\boldsymbol{\kappa}_i$ are smooth functions, there holds $\|v\|_{H^q(\tau_{i,0})} \eqsim \|v \circ \boldsymbol{\kappa}_i\|_{H^q(\mathcal{M})}$ with a constant dependent on q. For the reference domain, we have the classical approximation result for finite elements, cf. [BS08, Theorem 4.4.20] and a related inverse estimate, cf. [BS08, Theorem 4.5.11]. Now, (6.2) and (6.3) are obtained by the norm equivalence of the piece-wise defined norm with respect to the spaces $H^q(\tau_{i,0})$ and the norm on $H^q(D)$, cf. [SS11, Lemma 4.1.49].

Note that, by construction, h_j scales like the mesh size $\max_k\{\operatorname{diam}\tau_{\boldsymbol{\lambda}}\}$, i.e. it holds $h_j \eqsim \max_k\{\operatorname{diam}\tau_{\boldsymbol{\lambda}}\}$ uniformly in $j \in \mathbb{N}$ due to (5.1). In particular, for $t = 0$, we obtain an estimate for the L^2-orthogonal projection onto $V^s_{\mathcal{M},j}$, i.e.

$$\|v - Q_j v\|_{L^2(D)} := \inf_{v_j \in V^s_{\mathcal{M},j}} \|v - v_j\|_{L^2(D)} \lesssim h_j^q \|v\|_{H^q(D)}. \tag{6.5}$$

(6.6) **Remark.** Of course, we can define continuous finite elements on hypercubes and discontinuous finite elements on polyhedrons in complete analogy. If the shape of the reference domain and the hierarchy of the finite element spaces is not of interest, we will simply refer to the finite element spaces for a mesh of mesh-width h and a polynomial degree s by V_h^s.

In the sequel, we will also deal with finite element approximations to vector fields in $\mathbb{R}^d$. Therefore, we shall also introduce the related L^2-orthogonal projection. Let $Q_h \colon L^2(D) \to V_h^s$ denote the L^2-orthogonal projection onto V_h^s. The orthogonal projection Q_h can be extended to an orthogonal projection

$$\mathbf{Q}_h \colon L^2(D;\mathbb{R}^d) \to [V_h^s]^d, \quad [v_1 \dots, v_d]^\intercal \mapsto [Q_h v_1, \dots, Q_h v_d]^\intercal, \tag{6.7}$$

i.e. the component-wise projection of a function in $L^2(D;\mathbb{R}^d)$.

(6.8) **Lemma.** The operator $\mathbf{Q}_h$ defined by (6.7) is the L^2-orthogonal projection from $L^2(D;\mathbb{R}^d)$ to $[V_h^s]^d$. Moreover, it holds

$$\|(\mathbf{I} - \mathbf{Q}_h)\mathbf{v}\|_{L^2(D;\mathbb{R}^d)} \lesssim h^q \|\mathbf{v}\|_{H^q(D;\mathbb{R}^d)} \quad \text{for } \mathbf{v} \in H^q(D;\mathbb{R}^d). \tag{6.9}$$

Here, $\mathbf{I} \in \mathbb{R}^{d\times d}$ denotes the identity matrix.

Proof. Obviously, it holds $\operatorname{Im}\mathbf{Q}_h = [V_h^s]^d$. Let $\mathbf{v} \in H^q(D;\mathbb{R}^d)$ and $\mathbf{w} \in [V_h^s]^d$. We deduce

$$\big((\mathbf{I} - \mathbf{Q}_h)\mathbf{v}, \mathbf{q}\big)_{L^2(D;\mathbb{R}^d)} = \sum_{i=1}^d (v_i - Q_h v_i, w_i)_{L^2(D)} = 0$$

due to the orthogonality of Q_h. This shows the orthogonality of $\mathbf{Q}_h$. Moreover, we have by (6.5) that

$$\begin{aligned}\|(\mathbf{I} - \mathbf{Q}_h)\mathbf{v}\|^2_{L^2(D;\mathbb{R}^d)} &= \sum_{i=1}^d \|v_i - Q_h v_i\|^2_{L^2(D)} \\ &\lesssim \sum_{i=1}^d h^{2q} \|v_i\|^2_{H^q(D)} = h^{2q} \|\mathbf{v}\|^2_{H^q(D;\mathbb{R}^d)}.\end{aligned}$$

□

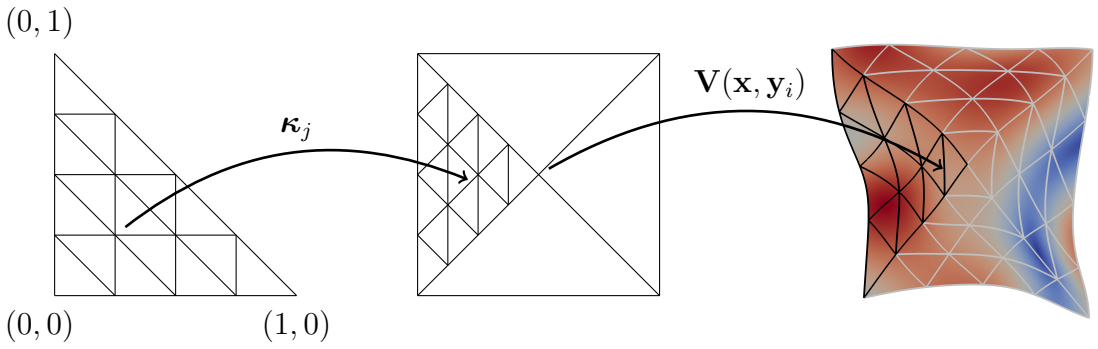

Figure II.6: Construction of parametric finite elements for mapped domains.

For the analysis of the regularity of the solution to elliptic diffusion problems on random domains in Chapter IV, we will exploit that there exists a one-to-one correspondence between the deterministic problem on the random domain and the random problem pulled-back to the reference domain. Nevertheless, for the numerical computations, in contrast to [CNT13, TX06], we do not aim at mapping the equation to a fixed reference domain D_{ref} but rather intend to solve the equation on each particular realization $D(\mathbf{y}_i) = \mathbf{V}(D_{\mathrm{ref}}, \mathbf{y}_i)$ for suitable parameters $\mathbf{y}_i$, where $i \in \mathcal{I}$ for an appropriate index set $\mathcal{I}$. Here, we assume that the vector field $\mathbf{V}(\mathbf{y})$ corresponds for each choice of the parameter $\mathbf{y}$ to a C^2-diffeomorphism.

A first step towards this approach is made by [MNK11], where a random boundary variation is assumed and a mesh on the realization of the reference domain D_{ref} is generated via the solution of the Laplacian. Under the assumption that the realizations of D_{ref} are provided by a sufficiently smooth vector field, we may also employ *mapped* parametric finite elements to directly approximate functions on the mapped domains $D(\mathbf{y}_i)$. Figure II.6 visualizes this procedure.

The argumentation in this situation is similar to that presented before. We have to assume that the singular values of the vector fields Jacobian $\mathbf{J}(\mathbf{x}, \mathbf{y})$ are bounded in the following way: There exists constants $0 < \underline{\sigma} \leqslant \overline{\sigma} < \infty$ such that

$$\underline{\sigma} \leqslant \min\{\sigma(\mathbf{J}(\mathbf{x}, \mathbf{y}))\} \leqslant \max\{\sigma(\mathbf{J}(\mathbf{x}, \mathbf{y}))\} \leqslant \overline{\sigma}$$

holds for the singular values of $\mathbf{J}(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x} \in D_{\mathrm{ref}}$ and almost every $\mathbf{y}$ in the set of admissible parameters. This condition guarantees that (5.1) is satisfied with $\rho = \overline{\sigma}/\underline{\sigma}$. Also the Jackson and Bernstein type estimates (6.2) and (6.3) are still valid, where the only limitation is imposed by the smoothness of $\mathbf{V}(\mathbf{x}, \mathbf{y})$. For example, in our case that $\mathbf{V}(\mathbf{x}, \mathbf{y})$ is a C^2-diffeomorphism, we have the restriction $q \leq 2$ such that

$$\inf_{v_\ell \in V_j(\mathbf{y})} \|u - v_j\|_{H^t(D(\mathbf{y}))} \lesssim h_j^{q-t} \|u\|_{H^q(D(\mathbf{y}))}$$

for all $0 \leqslant t \leqslant 3/2$, $t \leq q \leq 2$, where

$$V_j(\mathbf{y}) := \{\varphi \circ \mathbf{V}(\mathbf{y})^{-1} : \varphi \in V^s_{\triangle,j}\} \subset H^1(D(\mathbf{y})). \tag{6.10}$$

(6.11) **Remark.** For elliptic diffusion problems, the H^2-regularity of the mapped problem, i.e. on $D(\mathbf{y})$, follows from the H^2-regularity of the problem on the reference domain D_{ref} if the vector field $\mathbf{V}(\mathbf{x},\mathbf{y})$ is at least a C^2-diffeomorphism. Especially, if $\mathbf{V}(\mathbf{x},\mathbf{y}) = \mathbf{x} + \mathbf{V}_0(\mathbf{x},\mathbf{y})$ is a perturbation of the identity as in (IV.3.1) and $\mathbf{V}_0(\mathbf{x},\mathbf{y})$ is a vector field of class C^2, then $\mathbf{V}(\mathbf{x},\mathbf{y})^{-1}$ is also a C^2-diffeomorphism provided that $\|\mathbf{V}_0(\cdot,\mathbf{y})\|_{C^2(\overline{D_{\text{ref}}};\mathbb{R}^d)} < 1/2$, cf. [Sim80].

Chapter III

APPROXIMATION OF RANDOM FIELDS

As it has been discussed in Chapter II, the main task in the computation of a Karhunen-Loève expansion is the solution of a symmetric and positive semidefinite eigen-problem. In this chapter, we present and compare different numerical approaches for the approximation of random fields. These are cluster methods, which are well established in the business of discretizing non-local operators, namely the Adaptive Cross Approximation (ACA) and the Fast Multipole Method (FMM) on the one hand, and the Pivoted Cholesky Decomposition (PCD) on the other hand. As we will see, the PCD can be considered as a black-box algorithm to compute low-rank approximations to symmetric and positive semi-definite operators. Although, the results presented here are valid for $d_1 \neq d_2$, we restrict ourselves in the subsequent analysis to the case $d_1 = d_2 = d$, i.e., for fixed $\omega \in \Omega$, $D \subset \mathbb{R}^d$, we consider mappings $\mathbf{u}(\omega)\colon D \to \mathbb{R}^d$.

We start our considerations by providing approximation results for random fields. To that end, we have to assume that the random field provides additional spatial regularity, i.e.

$$\mathbf{u}(\omega, \mathbf{x}) \in L^2_{\mathbb{P}}(\Omega) \otimes H^q(D; \mathbb{R}^d) \quad \text{with } q > 0.$$

Thus, the associated Hilbert-Schmidt operator satisfies $T_{\mathbf{u}_0}\colon L^2_{\mathbb{P}}(\Omega) \to H^q(D; \mathbb{R}^d)$. In particular, it holds

$$\mathcal{C} := T_{\mathrm{Cov}[\mathbf{u}]} = T_{\mathbf{u}_0} T^\star_{\mathbf{u}_0}\colon H^q(D; \mathbb{R}^d) \to H^q(D; \mathbb{R}^d).$$

We shall be first concerned with the approximability of the covariance operator and derive error estimates involving the trace of the covariance operator $\mathcal{C}$.

1. Error bounds in terms of the trace

In the beginning of this section, we show that the covariance operator $\mathcal{C}$ extends to an operator $\mathcal{C}\colon \tilde{H}^{-q}(D; \mathbb{R}^d) \to H^q(D; \mathbb{R}^d)$.

(1.1) **Lemma.** By defining the dual operator $T^\star_{\mathbf{u}_0}\colon \tilde{H}^{-q}(D; \mathbb{R}^d) \to L^2_{\mathbb{P}}(\Omega)$ with respect to the $L^2(D; \mathbb{R}^d)$-inner product, i.e. $(T^\star_{\mathbf{u}_0}\mathbf{f}, \psi)_{L^2_{\mathbb{P}}(\Omega)} = (\mathbf{f}, T_{\mathbf{u}_0}\psi)_{L^2(D;\mathbb{R}^d)}$, the operator $T_{\mathrm{Cov}[\mathbf{u}]}$ extends to a continuous and compact operator from $\tilde{H}^{-q}(D; \mathbb{R}^d)$ to $H^q(D; \mathbb{R}^d)$ in accordance with

$$T_{\mathrm{Cov}[\mathbf{u}]}\mathbf{f} = \sum_{i \in \mathcal{I}} \sigma_i^2 (\mathbf{f}, \boldsymbol{\varphi}_i)_{L^2(D;\mathbb{R}^d)} \boldsymbol{\varphi}_i.$$

Proof. From $(T^\star_{\mathbf{u}_0}\mathbf{f},\psi)_{L^2_{\mathbb{P}}(\Omega)} = (\mathbf{f}, T_{\mathbf{u}_0}\psi)_{L^2(D;\mathbb{R}^d)}$, we deduce

$$T^\star_{\mathbf{u}_0}\mathbf{f} = \sum_{i\in\mathcal{I}} \sigma_i (\mathbf{f},\boldsymbol{\varphi}_i)_{L^2(D;\mathbb{R}^d)} X_i.$$

Hence, it holds

$$T_{\mathbf{u}_0}T^\star_{\mathbf{u}_0}\mathbf{f} = \sum_{i\in\mathcal{I}} \sigma_i \Big(\sum_{j\in\mathcal{I}} \sigma_j (\mathbf{f},\boldsymbol{\varphi}_j)_{L^2(D;\mathbb{R}^d)} X_j, X_i \Big)_{L^2_{\mathbb{P}}(\Omega)} \boldsymbol{\varphi}_i = \sum_{i\in\mathcal{I}} \lambda_i (\mathbf{f},\boldsymbol{\varphi}_i)_{L^2(D;\mathbb{R}^d)} \boldsymbol{\varphi}_i.$$

By applying the generalized Cauchy-Schwarz inequality, cf. Lemma (II.2.10), we arrive at

$$\begin{aligned}
\|T^\star_{\mathbf{u}_0}\mathbf{f}\|^2_{L^2_{\mathbb{P}}(\Omega)} &= \Big(\sum_{i\in\mathcal{I}} \sigma_i (\mathbf{f},\boldsymbol{\varphi}_i)_{L^2(D;\mathbb{R}^d)} X_i, \sum_{j\in\mathcal{I}} \sigma_j (\mathbf{f},\boldsymbol{\varphi}_j)_{L^2(D;\mathbb{R}^d)} X_j \Big)_{L^2_{\mathbb{P}}(\Omega)} \\
&= \sum_{i\in\mathcal{I}} \sigma_i^2 \big((\mathbf{f},\boldsymbol{\varphi}_i)_{L^2(D;\mathbb{R}^d)} \big)^2 \|X_i\|^2_{L^2_{\mathbb{P}}(\Omega)} \\
&\leqslant \sum_{i\in\mathcal{I}} \sigma_i^2 \|\mathbf{f}\|^2_{\check{H}^{-q}(D;\mathbb{R}^d)} \|\boldsymbol{\varphi}_i\|^2_{H^q(D;\mathbb{R}^d)} \|X_i\|^2_{L^2_{\mathbb{P}}(\Omega)} \\
&= \|\mathbf{u}_0\|^2_{L^2_{\mathbb{P}}(\Omega)\otimes H^q(D;\mathbb{R}^d)} \|\mathbf{f}\|^2_{\check{H}^{-q}(D;\mathbb{R}^d)}.
\end{aligned}$$

Thus, $T^\star_{\mathbf{u}_0}$ is a bounded operator. Since $T_{\mathbf{u}_0}$ is compact, so is $T_{\mathbf{u}_0}T^\star_{\mathbf{u}_0} = T_{\mathrm{Cov}[\mathbf{u}]}$. □

Combining equations (II.1.10) and (II.1.13), we obtain

$$\operatorname{Tr}\mathcal{C} = \|\mathbf{u}_0\|^2_{L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^d)}, \tag{1.2}$$

where the trace is defined with respect to the $L^2(D;\mathbb{R}^d)$-inner product. This identity gives rise to an approximation result in the finite element space V_h^s which bounds the error in terms of the trace, cf. [ST06, Theorem 2.10].

(1.3) **Theorem.** Let $N = \dim V_h^s$, let $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant 0$ be the eigenvalues of the covariance operator $\mathcal{C}$ and let $\lambda_{1,h} \geqslant \lambda_{2,h} \geqslant \ldots \geqslant \lambda_{dN,h} \geqslant 0$ be the eigenvalues of $\mathcal{C}_h := \mathbf{Q}_h\mathcal{C}\mathbf{Q}_h$, where $\mathbf{Q}_h$ is given by (II.6.7). Then, it holds

$$\|\mathbf{u}_0 - (\mathrm{Id}\otimes\mathbf{Q}_h)\mathbf{u}_0\|^2_{L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^d)} = \operatorname{Tr}\mathcal{C} - \operatorname{Tr}\mathcal{C}_h$$

and therefore

$$\|\mathbf{u}_0 - (\mathrm{Id}\otimes\mathbf{Q}_h)\mathbf{u}_0\|^2_{L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^d)} = \sum_{i=1}^{dN} (\lambda_i - \lambda_{i,h}) + \sum_{i=dN+1}^{\infty} \lambda_i,$$

where we set $\lambda_i = 0$ for $i > \#\mathcal{I}$.

Proof. Let $\{\boldsymbol{\varphi}_i\}_i$ be an orthonormal basis of $L^2(D;\mathbb{R}^d)$ such that either $\boldsymbol{\varphi}_i \in \operatorname{Im}\mathbf{Q}_h$ or $\boldsymbol{\varphi}_i \in \operatorname{Im}(\mathbf{I}-\mathbf{Q}_h)$ holds. Then, one easily derives $(T^\star_{\mathbf{u}_0}(\mathbf{I}-\mathbf{Q}_h)\boldsymbol{\varphi}_i, T^\star_{\mathbf{u}_0}\mathbf{Q}_h\boldsymbol{\varphi}_i)_{L^2_{\mathbb{P}}(\Omega)} = 0$. Thus, we infer by (1.2) that

$$\begin{aligned}
&\|\mathbf{u}_0 - (\mathrm{Id}\otimes\mathbf{Q}_h)\mathbf{u}_0\|^2_{L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^d)} \\
&\quad = \sum_{i=1}^{\infty} (T^\star_{\mathbf{u}_0}\boldsymbol{\varphi}_i, T^\star_{\mathbf{u}_0}\boldsymbol{\varphi}_i)_{L^2_{\mathbb{P}}(\Omega)} - (T^\star_{\mathbf{u}_0}\mathbf{Q}_h\boldsymbol{\varphi}_i, T^\star_{\mathbf{u}_0}\mathbf{Q}_h\boldsymbol{\varphi}_i)_{L^2_{\mathbb{P}}(\Omega)} = \operatorname{Tr}\mathcal{C} - \operatorname{Tr}\mathcal{C}_h.
\end{aligned}$$

□

Even though this result has already been derived in [ST06, Theorem 2.10], we have presented here an alternative proof which employs another technique required for our considerations later on. Notice that the proof of the theorem heavily relies on the properties of orthogonal projections, especially on the self-adjointness. Thus, we cannot weaken here this supposition to arbitrary projections.

For the rest of this thesis, we refer to $\{(\lambda_i, \boldsymbol{\varphi}_i)\}_{i\in\mathcal{I}}$ as the eigen-pairs of $\mathcal{C}$ (in decreasing order) and to $\{(\lambda_{i,h}, \boldsymbol{\varphi}_{i,h})\}_{i=1}^{dN}$ as the eigen-pairs of $\mathcal{C}_h$. By the application of Theorem (1.3) and the approximation estimate (II.6.5) it is straightforward to show the following

(1.4) **Corollary.** The trace error satisfies

$$0 \leqslant \operatorname{Tr}\mathcal{C} - \operatorname{Tr}\mathcal{C}_h \lesssim h^{2\min\{s+1,q\}}$$

if $\mathbf{u} \in L^2_{\mathbb{P}}(\Omega) \otimes H^q(D;\mathbb{R}^d)$.

Theorem (1.3) remains valid for $\mathcal{C}_{h,M} := \mathbf{P}_h\mathcal{C}\mathbf{P}_h$, where $\mathbf{P}_h\colon L^2(D;\mathbb{R}^d) \to U$ denotes an arbitrary orthogonal projection onto an M-dimensional subspace $U \subset [V_h^s]^d$. Thus, we obtain

(1.5) **Theorem.** Let $\mathcal{C}_h = \mathbf{Q}_h\mathcal{C}\mathbf{Q}_h$ and $\mathcal{C}_{h,M} = \mathbf{P}_h\mathcal{C}\mathbf{P}_h$. Then, there holds

$$\|\mathbf{u} - (\mathrm{Id}\otimes\mathbf{P}_h)\mathbf{u}\|^2_{L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^d)} \lesssim h^{2\min\{s+1,q\}} + (\operatorname{Tr}\mathcal{C}_h - \operatorname{Tr}\mathcal{C}_{h,M}), \tag{1.6}$$

where the hidden constant involves the norm of $\mathbf{u}$ in $L^2_{\mathbb{P}}(\Omega) \otimes H^q(D;\mathbb{R}^d)$.

Proof. By Theorem (1.3) and Corollary (1.4), it holds

$$\begin{aligned}
&\|\mathbf{u} - (\mathrm{Id}\otimes\mathbf{P}_h)\mathbf{u}\|^2_{L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^d)} \\
&\quad\leqslant \|(\mathbf{I}-\mathbf{Q}_h)\mathbb{E}[\mathbf{u}]\|^2_{L^2(D;\mathbb{R}^d)} + (\operatorname{Tr}\mathcal{C} - \operatorname{Tr}\mathcal{C}_h) + (\operatorname{Tr}\mathcal{C}_h - \operatorname{Tr}\mathcal{C}_{h,M}) \\
&\quad\lesssim h^{2\min\{s+1,q\}} + (\operatorname{Tr}\mathcal{C}_h - \operatorname{Tr}\mathcal{C}_{h,M}).
\end{aligned}$$

□

The theorem indicates that, after fixing the ansatz space V_h^s, the approximation error of the stochastic field is controllable in terms of the discretized operators $\mathcal{C}_h$ and $\mathcal{C}_{h,M}$. The optimal choice of $\mathbf{P}_h$ in terms of minimizing the trace error is the orthogonal projection onto the dominant invariant subspace of $\mathcal{C}_h$, i.e. $U_{M,h} := \operatorname{span}\{\boldsymbol{\varphi}_{1,h}, \dots, \boldsymbol{\varphi}_{M,h}\}$ corresponding to the M dominant eigenvalues of $\mathcal{C}_h$. If $U_{M,h}$ and thus $\mathbf{P}_h$ are not known exactly, they have to be approximated appropriately. This induces an additional error and we have to assume that $\lambda_{M+1,h}$ is distinct from $\lambda_{M,h}$, cf. [Kny97, Ovt06]. Nevertheless, any subspace is feasible as long as the difference $\operatorname{Tr}\mathcal{C}_h - \operatorname{Tr}\mathcal{C}_{h,M}$ becomes small for moderate sizes of M.

2. Decay rates for the eigenvalues of the covariance

Usually, the index set $\mathcal{I}$ which occurs in the Karhunen-Loève expansion

$$\mathbf{u}(\omega,\mathbf{x}) = \mathbb{E}[\mathbf{u}](\mathbf{x}) + \sum_{i\in\mathcal{I}} \sigma_i X_i(\omega)\boldsymbol{\varphi}_i(\mathbf{x}),$$

cf. Definition (II.3.1), is not finite. Therefore, for numerical applications, the Karhunen-Loève expansion has to be truncated appropriately after $M \in \mathbb{N}$ terms. The question how small M can actually be chosen in order to achieve a prescribed precision in the approximation of the covariance operator is closely related to the decay of the eigenvalues of $\mathcal{C}$ and $\mathcal{C}_h$, respectively, which depends on the smoothness index q. Results on the decay of the eigenvalues have already been established for periodic functions in [Tem86]. Nevertheless, since we do not want to restrict ourselves to this situation, we refer here to the more general results in [GH14, ST06] and extend them to the case of vector fields.

For $\mathbf{u} \in L^2_{\mathbb{P}}(\Omega) \otimes H^q(D;\mathbb{R}^d)$, it obviously holds

$$\operatorname{Cov}[\mathbf{u}] \in H^q(D;\mathbb{R}^d) \otimes H^q(D;\mathbb{R}^d) \quad \text{with} \quad \operatorname{Cov}[\mathbf{u}] = \sum_{i\in\mathcal{I}} \lambda_i \boldsymbol{\varphi}_i \otimes \boldsymbol{\varphi}_i,$$

where $\lambda_i = \sigma_i^2$ denote the eigenvalues of $T_{\operatorname{Cov}[\mathbf{u}]}$. At the same time, we will consider the operator

$$T_{\widetilde{\operatorname{Cov}}[\mathbf{u}]} = T^\star_{\mathbf{u}_0} T_{\mathbf{u}_0} \colon L^2_{\mathbb{P}}(\Omega) \to L^2_{\mathbb{P}}(\Omega).$$

Here, we derive

$$\widetilde{\operatorname{Cov}}[\mathbf{u}] \in L^2_{\mathbb{P}}(\Omega) \otimes L^2_{\mathbb{P}}(\Omega) \quad \text{with} \quad \widetilde{\operatorname{Cov}}[\mathbf{u}] = \sum_{i\in\mathcal{I}} \lambda_i X_i \otimes X_i.$$

The following theorem for the decay of the covariance operator's eigenvalues is a modification of the related theorem in [GH14] for the case of $\mathbb{R}^d$ valued functions, see also [ST06].

(2.1) **Theorem.** Let $\mathbf{u} \in L^2_{\mathbb{P}}(\Omega) \otimes H^q(D;\mathbb{R}^d)$. Then, the eigenvalues of the covariance operator $T_{\operatorname{Cov}[\mathbf{u}]} \colon \tilde{H}^{-q}(D;\mathbb{R}^d) \to H^q(D;\mathbb{R}^d)$ decay like

$$\lambda_i \lesssim \left(\frac{i}{d}\right)^{-2q/d} \lambda_1 \quad \text{as } i \to \infty. \tag{2.2}$$

Proof. We shall focus on the approximation space $\big[V_h^{\lfloor q \rfloor}\big]^d \subset L^2(D;\mathbb{R}^d)$ of piecewise polynomial (discontinuous) finite elements. Let $N = \dim V_h^{\lfloor q \rfloor}$. Then, we have for the L^2-orthogonal projection $\mathbf{Q}_h$, cf. (II.6.7), $dN = \dim(\operatorname{Im} \mathbf{Q}_h)$.

Now, we can describe the eigenvalues of the operator $T_{\widetilde{\operatorname{Cov}}[\mathbf{u}]} \colon L^2_{\mathbb{P}}(\Omega) \to L^2_{\mathbb{P}}(\Omega)$ via the min-max principle of Courant-Fisher. For arbitrary subspaces $V_k \subset L^2_{\mathbb{P}}(\Omega)$ with $\dim(V_k) \leqslant k$ it holds that

$$\begin{aligned}\lambda_{k+1} &= \min_{V_k} \max_{\psi\in V_k^\perp, \|\psi\|_{L^2_{\mathbb{P}}(\Omega)}=1} (T_{\widetilde{\operatorname{Cov}}[\mathbf{u}]}\psi, \psi)_{L^2_{\mathbb{P}}(\Omega)} \\ &= \min_{V_k} \max_{\psi\in V_k^\perp, \|\psi\|_{L^2_{\mathbb{P}}(\Omega)}=1} (T_{\mathbf{u}_0}\psi, T_{\mathbf{u}_0}\psi)_{L^2(D;\mathbb{R}^d)}.\end{aligned}$$

For the choice $V_{dN} = \operatorname{Im}(T^\star_{\mathbf{u}_0}\mathbf{Q}_h T_{\mathbf{u}_0})$, the orthogonality of the projection $\mathbf{Q}_h$ yields

$$\begin{aligned}
\lambda_{dN+1} &\leqslant \max_{\psi\perp\operatorname{Im}(T^\star_{\mathbf{u}_0}\mathbf{Q}_h T_{\mathbf{u}_0}),\|\psi\|_{L^2_{\mathbb{P}}(\Omega)}=1} (T_{\mathbf{u}_0}\psi, T_{\mathbf{u}_0}\psi)_{L^2(D;\mathbb{R}^d)} \\
&= \max_{\psi\perp\operatorname{Im}(T^\star_{\mathbf{u}_0}\mathbf{Q}_h T_{\mathbf{u}_0}),\|\psi\|_{L^2_{\mathbb{P}}(\Omega)}=1} (T_{\mathbf{u}_0}\psi, (\mathbf{I}-\mathbf{Q}_h)T_{\mathbf{u}_0}\psi)_{L^2(D;\mathbb{R}^d)} \\
&= \max_{\psi\perp\operatorname{Im}(T^\star_{\mathbf{u}_0}\mathbf{Q}_h T_{\mathbf{u}_0}),\|\psi\|_{L^2_{\mathbb{P}}(\Omega)}=1} \|(\mathbf{I}-\mathbf{Q}_h)T_{\mathbf{u}_0}\psi\|^2_{L^2(D)} \\
&\leqslant \sup_{\|\psi\|_{L^2_{\mathbb{P}}(\Omega)}=1} \|(\mathbf{I}-\mathbf{Q}_h)T_{\mathbf{u}_0}\psi\|^2_{L^2(D)} \\
&\lesssim N^{-2q/d} \sup_{\|\psi\|_{L^2_{\mathbb{P}}(\Omega)}=1} \|T_{\mathbf{u}_0}\psi\|^2_{H^q(D;\mathbb{R})} = N^{-2q/d}\lambda_1,
\end{aligned}$$

where the hidden constant results from (II.6.9). The assertion is finally obtained by substituting N by $(N-1)/d$. □

(2.3) **Remark.** The preceding result can straightforwardly be generalized for $D \subset \mathbb{R}^{d_1}$ and $T_{\operatorname{Cov}[\mathbf{u}]}\colon \tilde{H}^{-q}(D;\mathbb{R}^{d_2}) \to H^q(D;\mathbb{R}^{d_2})$ according to

$$\lambda_i \lesssim \left(\frac{i}{d_2}\right)^{-2q/d_1} \lambda_1 \quad \text{as } i \to \infty.$$

For $d_2 = 1$, this is exactly the result found in [GH14]. Nevertheless, for the case of partial differential equations on random domains, the situation $d_1 = d_2 = d$ is the relevant one.

In accordance with [GH14], an estimation of the Karhunen-Loève expansion's truncation error is provided by the following theorem.

(2.4) **Theorem.** Let $\mathbf{u} \in L^2_{\mathbb{P}}(\Omega) \otimes H^q(D;\mathbb{R}^d)$ with $q > d/2$. Then, it holds

$$\left\|\mathbf{u}_0 - \sum_{i=1}^{M} \sigma_i X_i \otimes \boldsymbol{\varphi}_i\right\|_{L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^d)} = \sqrt{\sum_{i=M+1}^{\infty} \sigma_i^2} \lesssim \sqrt{\frac{d}{2q/d-1}} \left(\frac{M}{d}\right)^{1/2-q/d}.$$

Proof. According to Theorem (2.1), the eigenvalues $\lambda_i = \sigma_i^2$ of the covariance operator related to $\mathbf{u}$ decay like (2.2). Thus, it holds

$$\sum_{i=M+1}^{\infty} \lambda_i^2 \lesssim \sum_{i=M+1}^{\infty} \left(\frac{i}{d}\right)^{-2q/d} \leqslant \int_M^{\infty} \left(\frac{x}{d}\right)^{-2q/d} \mathrm{d}x = \frac{d}{2q/d-1}\left(\frac{M}{d}\right)^{1-2q/d}.$$

□

This theorem tells us that we have to choose

$$M \eqsim d\varepsilon^{\frac{2d}{d-2q}}$$

in order to guarantee an error bound

$$\left\|\mathbf{u}_0 - \sum_{i=1}^{M} \sigma_i X_i \otimes \boldsymbol{\varphi}_i\right\|_{L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^d)} \lesssim \varepsilon.$$

To determine how well the eigenvalues of the covariance operator can be approximated numerically, we have also to take the smoothness of the covariance operator's eigenfunctions into account, cf. [GH14, ST06].

(2.5) **Theorem.** Let $\mathbf{u} \in L^2_{\mathbb{P}}(\Omega) \otimes H^q(D;\mathbb{R}^d)$. Then, the eigenfunctions $\{\boldsymbol{\varphi}_i\}_{i\in\mathcal{I}}$ of the covariance operator $\mathcal{C}$ satisfy

$$\|\boldsymbol{\varphi}_i\|_{H^t(D;\mathbb{R}^d)} \lesssim \left(\frac{\sigma_1}{\sigma_i}\right)^{\frac{t}{q}}, \quad 0 \le t \le q.$$

Proof. From Lemma (II.1.16), we obtain $\boldsymbol{\varphi}_i = T_{\mathbf{u}_0} X_i/\sigma_i$. Thus, from the continuity of $T_{\mathbf{u}_0}$ we deduce

$$\|\boldsymbol{\varphi}_i\|_{H^q(D;\mathbb{R}^d)} = \frac{1}{\sigma_i}\|T_{\mathbf{u}_0} X_i\|_{H^q(D)} \leqslant \frac{\sigma_1}{\sigma_i}\|X_i\|_{L^2_{\mathbb{P}}(\Omega)} = \frac{\sigma_1}{\sigma_i}.$$

Moreover, we have

$$\|\boldsymbol{\varphi}_i\|^2_{L^2(D;\mathbb{R}^d)} = (T_{\mathbf{u}_0} X_i/\sigma_i, T_{\mathbf{u}_0} X_i/\sigma_i)_{L^2(D;\mathbb{R}^d)} = \frac{1}{\lambda_i}(T^\star_{\mathbf{u}_0} T_{\mathbf{u}_0} X_i, X_i)_{L^2_{\mathbb{P}}(\Omega)} = 1.$$

Thus, the assertion is true for the extremal cases $t = 0$ and $t = q$. The result for the intermediate Sobolev spaces is obtained by the interpolation estimate (A.2.6). Note that the hidden constant depends on $t \in (0, q)$. □

(2.6) **Remark.** As a consequence of Theorem (2.5), we can approximate the eigenfunctions in V_h^s according to

$$\inf_{\mathbf{v}_h \in [V_h^s]^d} \|\boldsymbol{\varphi}_i - \mathbf{v}_h\|_{L^2(D;\mathbb{R}^d)} \lesssim h^t \|\boldsymbol{\varphi}_i\|_{H^t(D;\mathbb{R}^d)} \lesssim \left(\frac{\sigma_1}{\sigma_i}\right)^{\frac{t}{q}} h^t, \tag{2.7}$$

given that $0 \leqslant t \leqslant \min\{s+1, q\}$.

The approximation error of the related *Ritz-values* $\lambda_{1,h}, \dots, \lambda_{dN,h}$, $N = \dim(V_h^s)$, is given in terms of the *gap* between the invariant sub-space

$$U_M = \operatorname{span}\{\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_M\} \subset L^2(D;\mathbb{R}^d)$$

($\dim(U_M) = M$) corresponding to the eigenvalues $\lambda_1, \dots, \lambda_M$ and the approximation space V_h^s, i.e.

$$\theta(U_M) := \sup_{\mathbf{v}\in U_M, \|\mathbf{v}\|_{L^2(D;\mathbb{R}^d)}=1} \|(\mathbf{I} - \mathbf{Q}_h)\mathbf{v}\|_{L^2(D;\mathbb{R}^d)},$$

where $\mathbf{Q}_h \colon L^2(D;\mathbb{R}^d) \to [V_h^s]^d$ denotes the $L^2(D;\mathbb{R}^d)$ orthogonal projection onto $[V_h^s]^d$. In order to control the gap, we employ the eigenfunctions' regularity. Then, we can estimate the gap as follows.

(2.8) **Lemma.** Let $U_M = \operatorname{span}\{\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_M\} \subset L^2(D;\mathbb{R}^d)$ be an invariant subspace of $\mathcal{C}$. Then, it holds for $0 \leqslant t \leqslant \min\{s+1, q\}$ and $\lambda_1, \dots, \lambda_M \neq 0$ that

$$\theta(U_M) \lesssim \sqrt{M}\left(\frac{\sigma_1}{\sigma_M}\right)^{\frac{t}{q}} h^t. \tag{2.9}$$

Proof. Let $\mathbf{v} = \sum_{i=1}^{M} \alpha_i \boldsymbol{\varphi}_i$ with $\|\mathbf{v}\|_{L^2(D)} = 1$. Thus, it holds $\sum_{i=1}^{M} \alpha_i^2 = 1$. With $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)$, we have

$$\begin{aligned}
\theta(U_M) &= \sup_{\mathbf{v} \in U_M, \|\mathbf{v}\|_{L^2(D;\mathbb{R}^d)}=1} \|(\mathbf{I} - \mathbf{Q}_h)\mathbf{v}\|_{L^2(D;\mathbb{R}^d)} \\
&= \sup_{\|\boldsymbol{\alpha}\|_{\ell^2}=1} \Big\| \sum_{i=1}^{M} \alpha_i (\mathbf{I} - \mathbf{Q}_h)\boldsymbol{\varphi}_i \Big\|_{L^2(D;\mathbb{R}^d)} \\
&\leqslant \sup_{\|\boldsymbol{\alpha}\|_{\ell^2}=1} \sum_{i=1}^{M} |\alpha_i| \|(\mathbf{I} - \mathbf{Q}_h)\boldsymbol{\varphi}_i\|_{L^2(D;\mathbb{R}^d)} \\
&\lesssim \sup_{\|\boldsymbol{\alpha}\|_{\ell^2}=1} \sum_{i=1}^{M} |\alpha_i| \Big(\frac{\sigma_1}{\sigma_i}\Big)^{\frac{t}{q}} h^t \leqslant \sqrt{M} \Big(\frac{\sigma_1}{\sigma_M}\Big)^{\frac{t}{q}} h^t,
\end{aligned}$$

where we used (2.7) in the second to last step. □

(2.10) **Remark.** In order to achieve convergence for the M-th eigenvalue, we have to guarantee $\theta(U_M) < 1$. This imposes a restriction on the mesh width h of the discretization. Moreover, it is required that $dN = \dim[V_h^s]^d \geqslant M$.

From [DM96, Theorem 9.2.2.2], we obtain the convergence result which relates the eigenvalues' rate of approximation to the gap.

(2.11) **Theorem.** Let $U_i = \operatorname{span}\{\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_i\}$ be an invariant subspace of $\mathcal{C}$ such that $\dim(\mathbf{Q}_h U_i) = i$ for $i = 1, \dots, M$. Then, the approximation $\lambda_{i,h}$ to the i-th eigenvalue determined by the Rayleigh-Ritz method, i.e. $\mathcal{C}_h \boldsymbol{\varphi}_{i,h} = \lambda_{i,h} \boldsymbol{\varphi}_{i,h}$, satisfies the estimate

$$0 \leqslant \lambda_i - \lambda_{i,h} \leqslant \lambda_i \big(\theta(U_i)\big)^2 \quad \text{for all } 1 \leqslant i \leqslant M. \tag{2.12}$$

Proof. For a proof of this result, see [DM96, Theorem 9.2.2.2]. □

As a consequence, we can approximate the eigenvalues in $[V_h^s]^d$ according to

$$0 \leqslant \lambda_i - \lambda_{i,h} \lesssim \lambda_i i \Big(\frac{\lambda_1}{\lambda_i}\Big)^{\frac{t}{q}} h^{2t} \quad \text{for all } 1 \leqslant i \leqslant M \text{ and } 0 \leq t \leq \min\{s+1, q\}.$$

Especially for $s \ll q$, the eigenvalues of $\mathcal{C}_h$ exhibit a similar rate of decay as the eigenvalues of $\mathcal{C}$ up to a relative error of order $ih^{2(s+1)}$.

3. Cluster methods

In this section, we introduce so called *cluster methods*. By name, these are the Adaptive Cross Approximation, cf. [Beb00, BR03], and the (interpolation based) Fast Multipole method, cf. [Gie01, HB02, SLS03]. Our particular realization of these methods is based on a parametric geometry representation by four-sided patches, i.e. the collection $\{\tau_{i,0}\}$ as introduced in Section II.5 with respect to the reference domain $\square$. A specialty of these methods, borrowed from the business of fast boundary element methods, is that they can

be regarded as black-box algorithms for the discretization of Hilbert-Schmidt operators since there is no explicit knowledge of the integral kernel assumed except for its smoothness apart from the diagonal. For the FMM, we exploit the parametric representation of the geometry for the interpolation. This results in a drastic reduction of the computational effort compared to the interpolation in space. The achieved compression in the case of a polynomial expansion of the kernel function is even better than that of $\mathcal{H}^2$-matrices, cf. [HB02]. Notice that we focus here on the discretization of scalar covariance functions, i.e. we only provide a means of compressing the blocks $\mathrm{Cov}_{i,j}\colon D \times D \to \mathbb{R}$. Moreover, the construction is presented here exclusively for $d = 2$. Thus, we either allow that $\kappa_i\colon [0,1]^2 \to \mathbb{R}^2$ (parametric domain) or $\kappa_i\colon [0,1]^2 \to \mathbb{R}^3$ (parametric surface), cf. Remark (II.5.2).

In view of the eigenvalue problem for the covariance operator considered in the preceding section, we refer here to the following situation: For a given kernel function $k(\mathbf{x},\mathbf{y})$, we efficiently want to solve the operator eigenvalue problem

$$(\mathcal{A}u)(\mathbf{x}) = \int_D k(\mathbf{x},\mathbf{y})u(\mathbf{y})\,\mathrm{d}\sigma_{\mathbf{y}} = \lambda u(\mathbf{x}). \tag{3.1}$$

Herein, the integral operator $\mathcal{A}$ is an operator of order $2q$, which means that it maps $H^q(D)$ continuously to $H^{-q}(D)$. Here, since we do not want to distinguish between Sobolev spaces on open domains or closed manifolds, see e.g. [SS11, Ste03], we make the convention that $H^q(D) = [H^{-q}(D)]'$ for $q < 0$. The kernel functions under consideration are supposed to be smooth as functions in the variables $\mathbf{x}$ and $\mathbf{y}$, apart from the diagonal $\{(\mathbf{x},\mathbf{y}) \in D \times D : \mathbf{x} = \mathbf{y}\}$, and may have a singularity on the diagonal. Such kernel functions arise, for instance, by applying a boundary integral formulation to a second order elliptic problem [SS11, Ste03] or as correlation kernels. In general, they decay like a negative power of the distance of the arguments which depends on the order $2q$ of the operator and the spatial dimension.

Since we employ parametric geometry representations, the integration is with respect to a non-constant *surface measure*: On the patch $\tau_{i,0}$, we denote the surface measure by

$$|\boldsymbol{\kappa}_i|(\mathbf{s}) := \left\|\partial_{s_1}\boldsymbol{\kappa}_i(\mathbf{s}) \times \partial_{s_2}\boldsymbol{\kappa}_i(\mathbf{s})\right\|_2. \tag{3.2}$$

Here, the vector product for the situation $\boldsymbol{\kappa}_i\colon [0,1]^2 \to \mathbb{R}^2$, is defined via the embedding $\iota\colon \mathbb{R}^2 \to \mathbb{R}^3$, $(x_1,x_2) \mapsto (x_1,x_2,0)$.

The variational formulation of the eigenvalue problem (3.1) reads as follows:

$$\text{Find } u \in H^q(D) \text{ such that } (\mathcal{A}u,v)_{L^2(D)} = \lambda(u,v)_{L^2(D)} \text{ for all } v \in H^q(D). \tag{3.3}$$

If we insert the parameterizations $\boldsymbol{\kappa}_i$, the bilinear form for $\mathcal{A}$ becomes

$$\begin{aligned}(\mathcal{A}u,v)_{L^2(D)} &= \int_D\int_D k(\mathbf{x},\mathbf{y})u(\mathbf{y})v(\mathbf{x})\,\mathrm{d}\sigma_{\mathbf{y}}\,\mathrm{d}\sigma_{\mathbf{x}}\\ &= \sum_{i,i'=1}^{K}\int_\square\int_\square k_{i,i'}(\mathbf{s},\mathbf{t})u(\boldsymbol{\kappa}_{i'}(\mathbf{t}))v(\boldsymbol{\kappa}_i(\mathbf{s}))\,\mathrm{d}\mathbf{t}\,\mathrm{d}\mathbf{s}\end{aligned}$$

and the L^2-inner product becomes

$$\begin{aligned}(u,v)_{L^2(D)} &= \int_D u(\mathbf{x})v(\mathbf{x})\,\mathrm{d}\sigma_{\mathbf{x}} \\ &= \sum_{i=1}^{K}\int_{\square} u(\boldsymbol{\kappa}_i(\mathbf{s}))v(\boldsymbol{\kappa}_i(\mathbf{s}))|\boldsymbol{\kappa}_i|(\mathbf{s})\,\mathrm{d}\mathbf{s}.\end{aligned}$$

Here, the kernels $k_{i,i'}$ denote the *transported kernel functions*

$$\left.\begin{aligned}&k_{i,i'}\colon \square\times\square \longrightarrow \mathbb{R},\\ &k_{i,i'}(\mathbf{s},\mathbf{t}) := k(\boldsymbol{\kappa}_i(\mathbf{s}),\boldsymbol{\kappa}_{i'}(\mathbf{t}))|\boldsymbol{\kappa}_i|(\mathbf{s})|\boldsymbol{\kappa}_{i'}|(\mathbf{t})\end{aligned}\right\}\quad i,i'=1,2,\ldots,K. \tag{3.4}$$

Since the kernel $k(\mathbf{x},\mathbf{y})$ is in general *asymptotically smooth*, cf. (A.3.1), the analyticity of the parameterizations $\{\boldsymbol{\kappa}_i\}_{i=1}^K$ gives rise to a decay estimate for the transported kernel function which is quite similar to (A.3.1).

(3.5) **Definition.** A kernel function $k(\mathbf{x},\mathbf{y})$ is called analytically standard of order $2q$ if constants $c_k > 0$ and $r_k > 0$ exist such that the partial derivatives of the transported kernel functions $k_{i,i'}(\mathbf{s},\mathbf{t})$ are uniformly bounded by

$$|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\partial_{\mathbf{t}}^{\boldsymbol{\beta}}k_{i,i'}(\mathbf{s},\mathbf{t})| \leq c_k\frac{(|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|)!}{r_k^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}\|\boldsymbol{\kappa}_i(\mathbf{s})-\boldsymbol{\kappa}_{i'}(\mathbf{t})\|_2^{-(2+2q+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|)} \tag{3.6}$$

provided that $2+2q+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|>0$.

(3.7) **Remark.** The parameterizations provide patch-wise smoothness. Hence, under these assumptions, most kernels of boundary integral operators $\mathcal{A}$ of order $2q$ are analytically standard of order $2q$. Indeed, in Section A.3 of the appendix, we present a proof of this statement.

In the context of the Galerkin scheme, it will be convenient to have also access to the *localized kernel functions*. In Section II.5, we have already defined the local element mappings, i.e.

$$\hat{\tau}_{j,k}\colon \square \to \boldsymbol{\kappa}_i^{-1}(\tau_{i,j,k}) \quad \text{for } j=0,1,\ldots,J \text{ and } k=0,1,\ldots,4^jK-1.$$

via dilatation and translation. Thus, the localized kernel functions are given by

$$k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}(\mathbf{s},\mathbf{t}) := k(\boldsymbol{\kappa}_{\boldsymbol{\lambda}}(\mathbf{s}),\boldsymbol{\kappa}_{\boldsymbol{\lambda}'}(\mathbf{t}))|\boldsymbol{\kappa}_{\boldsymbol{\lambda}}|(\mathbf{s})|\boldsymbol{\kappa}_{\boldsymbol{\lambda}'}|(\mathbf{t}) \tag{3.8}$$

with the *localized parameterizations* $\boldsymbol{\kappa}_{\boldsymbol{\lambda}} := \boldsymbol{\kappa}_i\circ\hat{\tau}_{j,k}$ and the corresponding surface measures $|\boldsymbol{\kappa}_{\boldsymbol{\lambda}}| := 2^{-2j}|\boldsymbol{\kappa}_i|\circ\hat{\tau}_{j,k}$ with $|\boldsymbol{\kappa}_i|$ as defined in (3.2), see also Figure II.5.

In the following, we will only consider the localized kernel functions. The subsequent proposition is an immediate consequence of the fact that $\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\hat{\tau}_{j,k}(\mathbf{s}) = 2^{-j}$ if $|\boldsymbol{\alpha}|=1$ and $\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\boldsymbol{\tau}_{j,k}(\mathbf{s})=0$ if $|\boldsymbol{\alpha}|>1$.

(3.9) **Proposition.** Let the kernel function $k(\mathbf{x},\mathbf{y})$ be analytically standard of order $2q$. Then, there exist constants $c_k>0$ and $r_k>0$ such that

$$|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\partial_{\mathbf{t}}^{\boldsymbol{\beta}}k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}(\mathbf{s},\mathbf{t})| \leqslant c_k\frac{(|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|)!}{r_k^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}\frac{2^{-|\boldsymbol{\lambda}|(|\boldsymbol{\alpha}|+2)}2^{-|\boldsymbol{\lambda}'|(|\boldsymbol{\beta}|+2)}}{\|\boldsymbol{\kappa}_{\boldsymbol{\lambda}}(\mathbf{s})-\boldsymbol{\kappa}_{\boldsymbol{\lambda}'}(\mathbf{t})\|_2^{2+2q+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}} \tag{3.10}$$

holds uniformly for all $\boldsymbol{\lambda},\boldsymbol{\lambda}'$ provided that $2+2q+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|>0$.

Now, we shall be concerned with the Galerkin scheme for the discretization of the variational formulation (3.3). By replacing the energy space $H^q(D)$ in the variational formulation (3.3) by the finite dimensional ansatz space $V_J^s \subset H^q(D)$, we arrive at the Galerkin scheme for the operator eigenvalue problem (3.1):

$$\text{Find } u_J \in V_J^s, \text{ such that} \quad \int_D \int_D k(\mathbf{x},\mathbf{y}) u_J(\mathbf{y}) v_J(\mathbf{x}) \,\mathrm{d}\sigma_{\mathbf{y}} \,\mathrm{d}\sigma_{\mathbf{x}} = \lambda \int_D u_J(\mathbf{x}) v_J(\mathbf{x}) \,\mathrm{d}\sigma_{\mathbf{x}} \quad \text{for all } v_J \in V_J^s. \tag{3.11}$$

By setting $\hat{u}_{\boldsymbol{\lambda}} := u_J \circ \boldsymbol{\kappa}_{\boldsymbol{\lambda}}$ and $\hat{v}_{\boldsymbol{\lambda}} := v_J \circ \boldsymbol{\kappa}_{\boldsymbol{\lambda}}$, we may rewrite (3.11) and arrive at the equation

$$\sum_{|\boldsymbol{\lambda}'|=J} \int_\square \int_\square k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}(\mathbf{s},\mathbf{t}) \hat{u}_{\boldsymbol{\lambda}'}(\mathbf{t}) \hat{v}_{\boldsymbol{\lambda}}(\mathbf{s}) \,\mathrm{d}\mathbf{t} \,\mathrm{d}\mathbf{s} = \lambda \int_\square \hat{u}_{\boldsymbol{\lambda}}(\mathbf{s}) \hat{v}_{\boldsymbol{\lambda}}(\mathbf{s}) \kappa_{\boldsymbol{\lambda}}(\mathbf{s}) \,\mathrm{d}\mathbf{s} \tag{3.12}$$

for all $\boldsymbol{\lambda}$ with $|\boldsymbol{\lambda}| = J$. In the case of element-wise supported, piecewise polynomial basis functions for V_J, this leads immediately to the generalized matrix eigenvalue problem

$$\mathbf{A}\mathbf{u} = \lambda \mathbf{B}\mathbf{u} \tag{3.13}$$

with the (block-) diagonal mass matrix $\mathbf{B}$. Otherwise, for basis functions of higher global smoothness, straightforward but obvious modifications have to be made to arrive at the linear system (3.13), cf. [SS11].

In the chosen basis representation, i.e. in the single-scale basis for V_J^s, the system matrix $\mathbf{A}$ in (3.13) is in general densely populated. This yields a rather high computational effort for the assembly and the matrix-vector multiplication. Fortunately, the system matrix is block-wise of low rank, i.e. it is compressible by means of an $\mathcal{H}$-matrix, cf. [Hac99]. The computational complexity can thus be drastically reduced by a block-wise compression scheme.

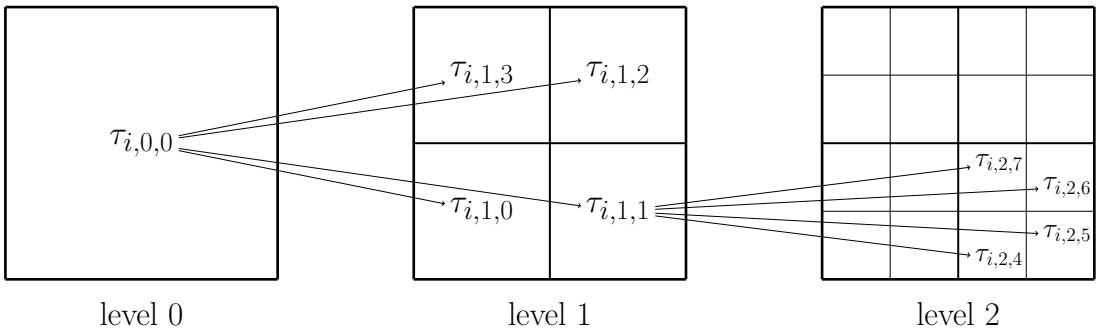

Figure III.1: Visualization of the cluster tree.

To that end, we introduce a tree structure on the nested meshes

$$\mathcal{T}_0 \subset \mathcal{T}_1 \subset \ldots \subset \mathcal{T}_J$$

introduced in Section II.5. Especially for our case $d = 2$, we end up with a *quad-tree* structured sequence of meshes consisting of N_j elements on level j. Therefore, we will also refer to $\tau_{i,j,k}$ as a *cluster*. In this case we think of $\tau_{i,j,k}$ as the union

$$\tau_{i,j,k} = \{\tau_{i,J,k'} : \tau_{i,J,k'} \subset \tau_{i,j,k}\},$$

i.e. the set of all tree leafs appended to $\tau_{i,j,k}$ or its sons. Furthermore, we denote the collection of all clusters, the *cluster tree*, by $\mathcal{T}$. A scheme for the subdivisions of the patch $\tau_{i,0}$ up to level 2 is shown in Figure III.1.

Now, we employ the following *admissibility condition* to determine compressible matrix blocks.

(3.14) **Definition.** The clusters $\tau_{\boldsymbol{\lambda}}$ and $\tau_{\boldsymbol{\lambda}'}$ with $|\boldsymbol{\lambda}| = |\boldsymbol{\lambda}'|$ are called *admissible* if

$$\max\{\operatorname{diam}(\tau_{\boldsymbol{\lambda}}), \operatorname{diam}(\tau_{\boldsymbol{\lambda}'})\} \leqslant \eta \operatorname{dist}(\tau_{\boldsymbol{\lambda}}, \tau_{\boldsymbol{\lambda}'}) \tag{3.15}$$

holds for some fixed $\eta \in (0, 1)$. The collection of admissible blocks $\tau_{\boldsymbol{\lambda}} \times \tau_{\boldsymbol{\lambda}'}$ forms the *far-field* of the operator. The remaining non-admissible blocks correspond to the *near-field* of the operator.

The quad-tree structure of the cluster tree $\mathcal{T}$ yields thus a block partitioning of the system matrix with quadratic blocks and each block on a particular level contains exactly the same number of element-element interactions, see Figure III.2 for a visualization of this special block partitioning of an $\mathcal{H}$-matrix.

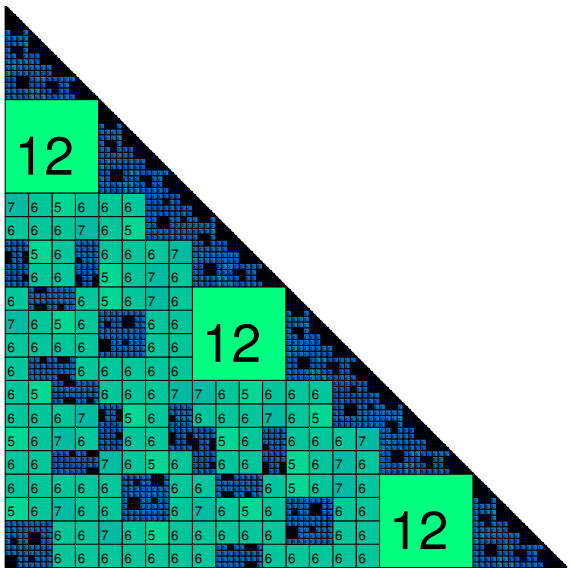

Figure III.2: Partition of the (symmetric) $\mathcal{H}$-matrix for the Matérn-9/2 kernel on $\mathbb{S}^2$ for level 4 with inscribed ranks.

Fast Multipole Method

We start by developing the black-box version of the FMM based on the interpolation of the kernel $k(\mathbf{x},\mathbf{y})$ as firstly proposed in [Gie01]. Note that, later on, this idea was also followed in [HB02] to construct $\mathcal{H}^2$-matrices.

For a given polynomial degree $p \in \mathbb{N}$, let $\{x_0, x_1, \dots, x_p\} \subset [0,1]$ be $p+1$ pairwise distinct points. Furthermore, let $L_m(s)$ for $m = 0, \dots, p$ be the Lagrangian basis polynomials with respect to the interpolation points x_m for $m = 0, \dots, p$. By a tensor product construction, we get the interpolation points $\mathbf{x_m} := (x_{m_1}, x_{m_2})$ and the corresponding tensor product interpolation polynomials $\mathbf{L_m}(\mathbf{s}) := L_{m_1}(s_1) \cdot L_{m_2}(s_2)$ for $m_1, m_2 = 0, \dots, p$. In all admissible blocks $\tau_{\boldsymbol{\lambda}} \times \tau_{\boldsymbol{\lambda}'}$, we approximate

$$k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}(\mathbf{s},\mathbf{t}) \approx \sum_{\|\mathbf{m}\|_\infty, \|\mathbf{m}'\|_\infty \leqslant p} k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}(\mathbf{x_m}, \mathbf{x_{m'}}) \mathbf{L_m}(\mathbf{s}) \mathbf{L_{m'}}(\mathbf{t}). \tag{3.16}$$

Consider now two basis functions $\hat{\varphi}_\ell, \hat{\varphi}_{\ell'} \in \hat{V}^s_{J-|\boldsymbol{\lambda}|}$ of the ansatz space on the level $J - |\boldsymbol{\lambda}|$. Since we employ quadrangular meshes, we may exploit the tensor product structure of the ansatz functions. Therefore, let $\hat{\varphi}_\ell = \hat{\varphi}^{(1)}_\ell \otimes \hat{\varphi}^{(2)}_\ell$ and $\hat{\varphi}_{\ell'} = \hat{\varphi}^{(1)}_{\ell'} \otimes \hat{\varphi}^{(2)}_{\ell'}$, respectively. From this and (3.16), we derive

$$\begin{aligned}
[\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}]_{\ell,\ell'} &\approx \int_\square \int_\square \sum_{\|\mathbf{m}\|_\infty, \|\mathbf{m}'\|_\infty \leqslant p} k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}(\mathbf{x_m}, \mathbf{x_{m'}}) \mathbf{L_m}(\mathbf{s}) \mathbf{L_{m'}}(\mathbf{t}) \hat{\varphi}_\ell(\mathbf{s}) \hat{\varphi}_{\ell'}(\mathbf{t}) \,\mathrm{d}\mathbf{t}\,\mathrm{d}\mathbf{s} \\
&= \sum_{\|\mathbf{m}\|_\infty, \|\mathbf{m}'\|_\infty \leqslant p} k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}(\mathbf{x_m}, \mathbf{x_{m'}}) \int_\square \mathbf{L_m}(\mathbf{s}) \hat{\varphi}_\ell(\mathbf{s}) \,\mathrm{d}\mathbf{s} \int_\square \mathbf{L_{m'}}(\mathbf{t}) \hat{\varphi}_{\ell'}(\mathbf{t}) \,\mathrm{d}\mathbf{t} \\
&=: [\mathbf{M}^\square_{|\boldsymbol{\lambda}|} \mathbf{K}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} (\mathbf{M}^\square_{|\boldsymbol{\lambda}'|})^\intercal]_{\ell,\ell'}.
\end{aligned}$$

By construction, each cluster on a particular level contains the same number of basis functions, namely $\dim(\hat{V}_{J-|\boldsymbol{\lambda}|})$. Additionally, the *moment matrices* $\mathbf{M}^\square_{|\boldsymbol{\lambda}|}$ are independent of the patch parameterization. This yields the

(3.17) **Proposition.** For $j = 1, 2, \dots, J$ and all $|\boldsymbol{\lambda}| = |\boldsymbol{\lambda}'| = j$, it holds

$$\mathbf{M}^\square_{|\boldsymbol{\lambda}|} = \mathbf{M}^\square_{|\boldsymbol{\lambda}'|}. \tag{3.18}$$

As a consequence we have to compute and store only a single moment matrix

$$\mathbf{M}^\square_{|\boldsymbol{\lambda}|} \in \mathbb{R}^{\dim(\hat{V}_{J-|\boldsymbol{\lambda}|}) \times (p+1)^2}$$

for each particular level. These moment matrices can be decomposed further by exploiting the tensor product structure of the basis functions:

$$\begin{aligned}
\int_\square \mathbf{L_m}(\mathbf{s}) \hat{\varphi}_\ell(\mathbf{s}) \,\mathrm{d}\mathbf{s} &= \int_0^1 \int_0^1 L_{m_1}(s_1) \hat{\varphi}^{(1)}_\ell(s_1) L_{m_2}(s_2) \hat{\varphi}^{(2)}_\ell(s_2) \,\mathrm{d}s_1\,\mathrm{d}s_2 \\
&= \int_0^1 L_{m_1}(s_1) \hat{\varphi}^{(1)}_\ell(s_1) \,\mathrm{d}s_1 \int_0^1 L_{m_2}(s_2) \hat{\varphi}^{(2)}_\ell(s_2 \,\mathrm{d}s_2 \\
&=: [\mathbf{M}_{|\boldsymbol{\lambda}|} \otimes \mathbf{M}_{|\boldsymbol{\lambda}|}]_{\ell,(p+1)m_1+m_2}.
\end{aligned}$$

Since

$$\mathbf{M}_{|\boldsymbol{\lambda}|} \in \mathbb{R}^{\sqrt{\dim(\hat{V}_{J-|\boldsymbol{\lambda}|})}\times(p+1)},$$

we end up with a major compression of the far-field.

(3.19) **Remark.** It is convenient to impose a lower threshold for the far-field. Therefore, we consider matrix blocks with $\mathcal{O}(p^4)$ entries as near-field. This yields $\mathcal{O}(N_J p^{-2})$ near-field blocks with a storage cost of $\mathcal{O}(N_J p^2)$.

(3.20) **Theorem.** The complexity for the computation and the storage of the far-field is given by $\mathcal{O}(N_J p^2)$.

Proof. At first, we show inductively that there are $\mathcal{O}(N_j)$ admissible and also $\mathcal{O}(N_j)$ non-admissible clusters on level j. For level 0 this is clearly true. Now, let the assumption hold for level $j-1$. On level $j-1$, for a fixed cluster, there exist $\mathcal{O}(1)$ neighbouring clusters which do not satisfy the admissibility condition (3.15). For such clusters, we have to consider the 4 son clusters on level j. Hence, we face $4\mathcal{O}(N_{j-1}) = \mathcal{O}(N_j)$ non-admissible and also $\mathcal{O}(N_j)$ admissible cluster-cluster interactions on level j.

Furthermore, in accordance with Remark (3.19), the maximum level to be computed is now $\lceil J - 2\log_4 p\rceil$. Due to $N_j = 4^j K$, we thus may estimate

$$\sum_{j=0}^{\lceil J-2\log_4 p\rceil} N_j = \mathcal{O}(K4^{\lceil J-2\log_4 p\rceil}) = \mathcal{O}(K4^J p^{-2}) = \mathcal{O}(N_J p^{-2}).$$

This yields, together with Remark (3.19), overall $\mathcal{O}(N_J p^{-2})$ far-field blocks and accordingly $\mathcal{O}(N_J p^{-2})$ near-field blocks.

For each far-field block, we have to evaluate and store the localized kernel function in $\mathcal{O}(p^4)$ points. The complexity for assembly and storage of the moment matrices is $\mathcal{O}(\sqrt{N_J}p)$ in total. Hence, the far-field complexity is

$$\mathcal{O}(N_J p^{-2}) \cdot \mathcal{O}(p^4) + \mathcal{O}(\sqrt{N_J}p) = \mathcal{O}(N_J p^2). \qquad \square$$

(3.21) **Remark.** Due to the parametric geometry representation, we obtain especially for boundary element methods in three dimensions, i.e. $\Gamma \subset \mathbb{R}^3$, an improved cost complexity. The classical FMM proposes here to interpolate in space. Thus, the polynomial degree enters with $\mathcal{O}(p^3)$, cf. [Gie01, HB02]. Since we only interpolate the transported kernel on the reference domain, we can reduce this cost to $\mathcal{O}(p^2)$.

Storing the moment matrices $\mathbf{M}_{|\boldsymbol{\lambda}|}$ on each particular level can be avoided by the concept of *nested cluster bases*, cf. [HB02], Obviously, since the polynomial degree for each cluster is p, we can represent the Lagrange polynomials of the father cluster by those of the son clusters. Let

$$\{x_m^{(0)}\}_{m=0}^p = \left\{\frac{x_m}{2}\right\}_{m=0}^p \quad \text{and} \quad \{x_m^{(1)}\}_{m=0}^p = \left\{\frac{x_m+1}{2}\right\}_{m=0}^p,$$

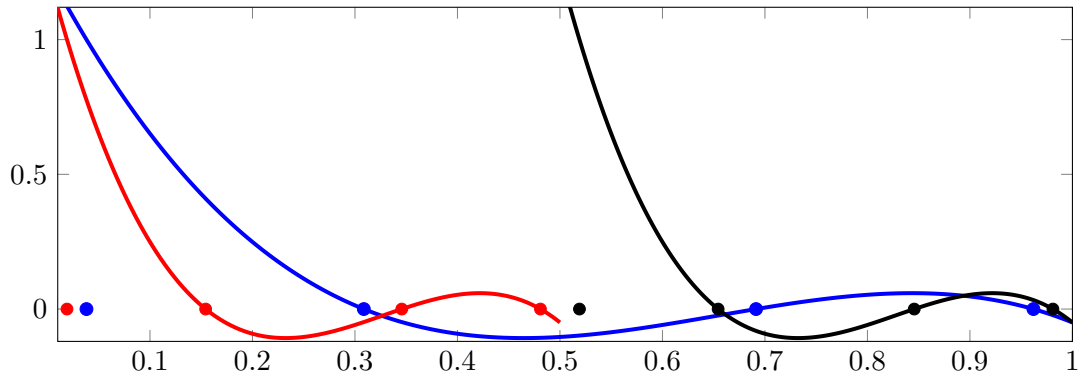

Figure III.3: First Lagrange polynomials of son clusters and father cluster.

respectively, be the interpolation points in the son clusters, see Figure III.3. It holds $\{x_m^{(0)}\}_{m=0}^p \subset [0, 0.5]$ and $\{x_m^{(1)}\}_{m=0}^p \subset [0.5, 1]$. If we denote the related Lagrange polynomials with $L_m^{(0)}(x)$ and $L_m^{(1)}(x)$, respectively, we can now exactly represent the Lagrange polynomials of the father cluster according to

$$L_m(x) = \sum_{i=0}^{p} L_m(x_i^{(0)}) L_i^{(0)}(x) \quad \text{for } x \in [0, 0.5]$$

and

$$L_m(x) = \sum_{i=0}^{p} L_m(x_i^{(1)}) L_i^{(1)}(x) \quad \text{for } x \in [0.5, 1].$$

This gives rise to the transfer matrices

$$\mathbf{T}^{(0)} := [L_i(x_j^{(0)})]_{i,j=0}^p \quad \text{and} \quad \mathbf{T}^{(1)} := [L_i(x_j^{(1)})]_{i,j=0}^p$$

which yields the representation

$$\mathbf{M}_{|\boldsymbol{\lambda}|} = \begin{bmatrix} \mathbf{T}^{(0)} \mathbf{M}_{|\lambda|+1} \\ \mathbf{T}^{(1)} \mathbf{M}_{|\lambda|+1} \end{bmatrix}.$$

By tensor product construction, we then obtain the four transfer matrices

$$\mathbf{T}_{2i+j}^{\square} := \mathbf{T}^{(i)} \otimes \mathbf{T}^{(j)}, \quad i, j = 0, 1,$$

for the reference domain $\square$. Here, we have the refinement relation

$$\mathbf{M}_{|\boldsymbol{\lambda}|}^{\square} = \begin{bmatrix} \mathbf{T}_0^{\square} \mathbf{M}_{|\lambda|+1}^{\square} \\ \mathbf{T}_2^{\square} \mathbf{M}_{|\lambda|+1}^{\square} \\ \mathbf{T}_3^{\square} \mathbf{M}_{|\lambda|+1}^{\square} \\ \mathbf{T}_1^{\square} \mathbf{M}_{|\lambda|+1}^{\square} \end{bmatrix}.$$

Notice that the peculiar order of the transfer matrices results from our hierarchical, counter clock-wise ordering of the elements, cf. Figure III.1. In order to make use of the efficient

implementation of the $\mathcal{H}^2$-matrix-vector product, cf. [HB02], we have only to store $\mathbf{M}_J^\square$ and $\mathbf{T}_0^\square, \mathbf{T}_1^\square, \mathbf{T}_2^\square, \mathbf{T}_3^\square$.

With Definition (3.5) at hand, the proof of convergence for our FMM is straightforward. We present it here for the case that Chebyshev nodes on $I := [0,1]$, i.e. the points

$$x_m := \frac{1}{2}\left[\cos\left(\frac{2m+1}{2(p+1)}\pi\right)+1\right], \quad m = 0,1,\dots,p,$$

are used for the interpolation [Gie01, HB02].

(3.22) **Theorem.** Let $k(\mathbf{x},\mathbf{y})$ be an analytically standard kernel of order $2q$. Then, in an admissible block $\tau_{\boldsymbol{\lambda}} \times \tau_{\boldsymbol{\lambda}'}$, it holds

$$\begin{aligned}\Big\| k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}(\mathbf{s},\mathbf{t}) - \sum_{\|\mathbf{m}\|_\infty,\|\mathbf{m}'\|_\infty \leqslant p} k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}(\mathbf{x}_{\mathbf{m}},\mathbf{x}_{\mathbf{m}'})\mathbf{L}_{\mathbf{m}}(\mathbf{s})\mathbf{L}_{\mathbf{m}'}(\mathbf{t})\Big\|_{L^\infty(\square\times\square)} \\ \lesssim \left(\frac{\eta}{r_k}\right)^{p+1} 2^{-4|\boldsymbol{\lambda}|}\|\boldsymbol{\kappa}_{\boldsymbol{\lambda}}(\mathbf{s}) - \boldsymbol{\kappa}_{\boldsymbol{\lambda}'}(\mathbf{t})\|_{L^\infty(\square\times\square)}^{-2(1+q)}\end{aligned}$$

with $r_k > 0$ being the constant from Definition (3.5).

Proof. We start with the one-dimensional interpolation error for the Chebyshev interpolation. It is well known that for a sufficiently smooth function $f\colon I \to \mathbb{R}$ the error estimate

$$\|f - \Pi_I^p f\|_{L^\infty(I)} \leqslant \frac{2\cdot 4^{-(p+1)}}{(p+1)!}\|\partial^{p+1} f\|_{L^\infty(I)}$$

holds, where the interpolation operator Π_I^p is defined by

$$\Pi_I^p f := \sum_{m=0}^{p} f(x_m) L_m(x).$$

According to [HB02, Estimate A.2], Π_I^p satisfies the stability estimate

$$\|\Pi_I^p f\|_{L^\infty(I)} \leqslant c\log(p+1)\|f\|_{L^\infty(I)} \tag{3.23}$$

for some constant $c > 0$. By tensorization, we obtain the d-dimensional interpolation operator $\boldsymbol{\Pi}_{I^d}^p$ on I^d. From [HB02, Lemma A.1], we know for the interpolation of a function $f\colon B_d \to \mathbb{R}$ in product Chebyshev nodes in $B_d := \prod_{\ell=1}^d [a_\ell, b_\ell]$ that

$$\|f - \boldsymbol{\Pi}_{B_d}^p\|_{L^\infty(B_d)} \leqslant \frac{2^{-p}}{(p+1)!}\sum_{\ell=1}^{d}\big(c\log(p+1)\big)^{\ell-1}\left(\frac{b_\ell - a_\ell}{2}\right)^{p+1}\|\partial_{s_\ell}^{p+1} f\|_{L^\infty(B_d)}.$$

Here, the constant c stems from the stability estimate (3.23). In our case, we interpolate on $\square\times\square$ which is isomorphic to I^4. Hence, the preceding estimate becomes

$$\|f - \boldsymbol{\Pi}_{I^4}^p\|_{L^\infty(I^4)} \lesssim \sum_{\ell=1}^{4}\frac{\big(\log(p+1)\big)^{\ell-1}}{2(p+1)!4^p}\|\partial_{s_\ell}^{p+1} f\|_{L^\infty(I^4)}.$$

Therefore, in view of (3.10), we conclude

$$\begin{aligned}
&\|k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \boldsymbol{\Pi}^p_{\square\times\square} k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}\|_{L^\infty(\square\times\square)} \\
&\quad\lesssim \sum_{\ell=1}^{4} \frac{\big(\log(p+1)\big)^{\ell-1}}{2(p+1)!\,4^p} \|\partial_{s_\ell}^{p+1} k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}\|_{L^\infty(\square\times\square)} \\
&\quad\lesssim \sum_{\ell=1}^{4} \frac{\big(\log(p+1)\big)^{\ell-1}}{2(p+1)!\,4^p} \frac{(p+1)!}{r_k^{p+1}} \|\boldsymbol{\kappa}_{\boldsymbol{\lambda}}(\mathbf{s}) - \boldsymbol{\kappa}_{\boldsymbol{\lambda}'}(\mathbf{t})\|^{-2(1+q)-(p+1)}_{L^\infty(\square\times\square)} 2^{-|\boldsymbol{\lambda}|((p+1)+4)} \\
&\quad\lesssim \sum_{\ell=1}^{4} \frac{\big(\log(p+1)\big)^{\ell-1}}{2r_k^{p+1}4^p} \operatorname{dist}(\tau_{\boldsymbol{\lambda}},\tau_{\boldsymbol{\lambda}'})^{-(p+1)} 2^{-|\boldsymbol{\lambda}|(p+5)} \|\boldsymbol{\kappa}_{\boldsymbol{\lambda}}(\mathbf{s}) - \boldsymbol{\kappa}_{\boldsymbol{\lambda}'}(\mathbf{t})\|^{-2(1+q)}_{L^\infty(\square\times\square)} \\
&\quad\lesssim \frac{2^{-|\boldsymbol{\lambda}|(p+5)}}{r_k^{p+1}\|\boldsymbol{\kappa}_{\boldsymbol{\lambda}}(\mathbf{s}) - \boldsymbol{\kappa}_{\boldsymbol{\lambda}'}(\mathbf{t})\|^{2(1+q)}_{L^\infty(\square\times\square)}} \operatorname{dist}(\tau_{\boldsymbol{\lambda}},\tau_{\boldsymbol{\lambda}'})^{-(p+1)}.
\end{aligned}$$

The admissibility condition (3.15) provides

$$\operatorname{dist}(\tau_{\boldsymbol{\lambda}},\tau_{\boldsymbol{\lambda}'}) \geqslant \frac{\max\{\operatorname{diam}(\tau_{\boldsymbol{\lambda}}),\operatorname{diam}(\tau_{\boldsymbol{\lambda}'})\}}{\eta}.$$

Moreover, the Lipschitz continuity of the parameterizations and their inverses imply

$$\operatorname{diam}\tau_{\boldsymbol{\lambda}} \eqsim 2^{-|\boldsymbol{\lambda}|} \quad \text{for all } |\boldsymbol{\lambda}| = 1,2,\dots,J.$$

Hence, we may bound

$$\operatorname{dist}(\tau_{\boldsymbol{\lambda}},\tau_{\boldsymbol{\lambda}'}) \gtrsim \frac{2^{-|\boldsymbol{\lambda}|}}{\eta}.$$

Inserting this estimate into the above expression finally yields

$$\begin{aligned}
\|k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \boldsymbol{\Pi}^p_{\square\times\square} k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}\|_{L^\infty(\square\times\square)} &\lesssim \frac{2^{-|\boldsymbol{\lambda}|(p+5)}}{r_k^{p+1}\|\boldsymbol{\kappa}_{\boldsymbol{\lambda}}(\mathbf{s}) - \boldsymbol{\kappa}_{\boldsymbol{\lambda}'}(\mathbf{t})\|^{2(1+q)}_{L^\infty(\square\times\square)}} \left(\frac{2^{-|\boldsymbol{\lambda}|}}{\eta}\right)^{-(p+1)} \\
&\lesssim 2^{-4|\boldsymbol{\lambda}|}\left(\frac{\eta}{r_k}\right)^{p+1} \|\boldsymbol{\kappa}_{\boldsymbol{\lambda}}(\mathbf{s}) - \boldsymbol{\kappa}_{\boldsymbol{\lambda}'}(\mathbf{t})\|^{-2(1+q)}_{L^\infty(\square\times\square)}. \qquad \square
\end{aligned}$$

As in [Gie01], we can directly derive from the previous theorem an error estimate for the bilinear form which is associated with the variational formulation (3.3).

(3.24) **Theorem.** Let $\sigma > 0$ be arbitrary but fixed. Then, for the integral operator $\mathcal{A}_J$ which results from an interpolation of degree $p > 0$ of the kernel function in every admissible block and the exact representation of the kernel in all other blocks, there holds

$$|(\mathcal{A}u,v)_{L^2(D)} - (\mathcal{A}_J u,v)_{L^2(D)}| \lesssim 2^{-J\sigma}\|u\|_{L^1(D)}\|v\|_{L^1(D)}$$

provided that $p \eqsim J(2+2q+\sigma)$.

Proof. From Theorem (3.22), one infers for admissible clusters $\tau_{\boldsymbol{\lambda}} \times \tau_{\boldsymbol{\lambda}'}$ that

$$\operatorname{dist}(\tau_{\boldsymbol{\lambda}}, \tau_{\boldsymbol{\lambda}'}) \gtrsim \frac{2^{-|\boldsymbol{\lambda}|}}{\eta} \geqslant 2^{-J}$$

since $\eta < 1$ and $|\boldsymbol{\lambda}| \leqslant J$. Therefore, it holds

$$\|k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \boldsymbol{\Pi}^p_{\square\times\square} k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}\|_{L^\infty(\square\times\square)} \lesssim 2^{-4|\boldsymbol{\lambda}|} \left(\frac{\eta}{r_k}\right)^{p+1} 2^{2J(1+q)}$$

for all $\boldsymbol{\lambda}, \boldsymbol{\lambda}'$ with $|\boldsymbol{\lambda}| = |\boldsymbol{\lambda}'|$, because the kernel representation is exact in non-admissible clusters.

Next, denote by $\mathcal{B} \subset \mathcal{T} \times \mathcal{T}$ the set of all matrix blocks, i.e. the union of all admissible and of all non-admissible blocks. Then, we may write

$$\begin{aligned}
&|(\mathcal{A}u, v)_{L^2(D)} - (\mathcal{A}_J u, v)_{L^2(D)}| \\
&\quad = \left| \sum_{(\boldsymbol{\lambda},\boldsymbol{\lambda}')\in\mathcal{B}} \int_\square \int_\square (k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \boldsymbol{\Pi}^p_{\square\times\square} k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'})(\mathbf{s},\mathbf{t}) \hat{u}_{\boldsymbol{\lambda}'}(\mathbf{t}) \hat{v}_{\boldsymbol{\lambda}}(\mathbf{s}) \,\mathrm{d}\mathbf{t}\,\mathrm{d}\mathbf{s} \right| \\
&\quad \leqslant \left| \sum_{(\boldsymbol{\lambda},\boldsymbol{\lambda}')\in\mathcal{B}} \int_\square \int_\square \|k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \boldsymbol{\Pi}^p_{\square\times\square} k_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}\|_{L^\infty(\square\times\square)} \hat{u}_{\boldsymbol{\lambda}'}(\mathbf{t}) \hat{v}_{\boldsymbol{\lambda}}(\mathbf{s}) \,\mathrm{d}\mathbf{t}\,\mathrm{d}\mathbf{s} \right| \\
&\quad \lesssim \left(\frac{\eta}{r_k}\right)^{p+1} 2^{2J(1+q)} \left| \sum_{(\boldsymbol{\lambda},\boldsymbol{\lambda}')\in\mathcal{B}} 2^{-4|\boldsymbol{\lambda}|} \int_\square \int_\square \hat{u}_{\boldsymbol{\lambda}'}(\mathbf{t}) \hat{v}_{\boldsymbol{\lambda}}(\mathbf{s}) \,\mathrm{d}\mathbf{t}\,\mathrm{d}\mathbf{s} \right| \\
&\quad \lesssim \left(\frac{\eta}{r_k}\right)^{p+1} 2^{2J(1+q)} \|u\|_{L^1(D)} \|v\|_{L^1(D)}.
\end{aligned}$$

In view of

$$\left(\frac{\eta}{r_k}\right)^{p+1} 2^{2J(1+q)} = 2^{-J\sigma} \quad \Longleftrightarrow \quad p+1 = \left| \frac{J(2+2q+\sigma)}{\log_2(\eta) - \log_2(r_k)} \right|,$$

we obtain the assertion. □

(3.25) **Remark.** To maintain the approximation order of the Galerkin method, we have to choose $p \eqsim \log N_J$. This yields an over-all complexity of $\mathcal{O}(N_J (\log N_J)^2)$ for the computation and the storage of the far-field. Nevertheless, if the integrals of the near-field cannot be evaluated with constant effort, then the computational effort of the near-field computation will in general dominate. For example, in the case of tensor product Gaussian quadrature rules and the Duffy trick, cf. [SS97, SS11], to regularize the singular integrals, one has to increase the degree of the quadrature for all singular integrals proportionally to $|\log h_J|$ where $h_J = 2^{-J}$ is the mesh size. Thus, the computational effort is $\mathcal{O}((\log N_J)^4)$ for each entry, which results in a complexity of $\mathcal{O}(N_J (\log N_J)^4)$ for all singular integrals. However, it can be shown that this is also the overall complexity for the whole near-field if the quadrature degree is properly decreased with the distance of the elements.

Adaptive Cross Approximation

We shall also introduce the ACA as an alternative for the compression of admissible matrix blocks. As a starting point, we employ again the admissibility condition (3.15) to partition the system matrix. Then, in each admissible matrix block, we approximate $\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} \in \mathbb{R}^{n\times n}$ with $n = \dim(\hat{V}^s_{J-|\boldsymbol{\lambda}|})$ by a truncated, partially pivoted Gaussian elimination, cf. [Beb00]. To this end, we define the vectors $\boldsymbol{\ell}_m, \mathbf{u}_m \in \mathbb{R}^n$ by the following iterative scheme, where $\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} = [a_{i,j}]_{i,j=1}^n$ is the matrix-block under consideration:

$$\text{for } m = 1, 2, \dots \text{ set } \mathbf{u}_m = \hat{\mathbf{u}}_m / [\hat{\mathbf{u}}_m]_{j_m} \text{ with}$$
$$\hat{\mathbf{u}}_m = [a_{i_m,j}]_{j=1}^n - \sum_{j=1}^{m-1} [\boldsymbol{\ell}_j]_{i_m} \mathbf{u}_j \text{ and } \boldsymbol{\ell}_m = [a_{i,j_m}]_{i=1}^n - \sum_{i=1}^{m-1} [\mathbf{u}_i]_{j_m} \boldsymbol{\ell}_i.$$

A criterion to guarantee the convergence of the algorithm is to choose the pivot element located in (i_m, j_m)-position as the maximum element in modulus of the remainder $\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_{m-1}\mathbf{U}_{m-1}$, where we define the matrices $\mathbf{L}_{m-1} := [\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{m-1}]$ and $\mathbf{U}_{m-1} := [\mathbf{u}_1 \dots, \mathbf{u}_{m-1}]^\intercal$. This would require the assembly of the whole matrix block $\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}$, which is not feasible in practice. Therefore, we employ another pivoting strategy which performs quite well in most cases. We choose j_m such that $[\hat{\mathbf{u}}_m]_{j_m}$ is the largest element in modulus of the row $\hat{\mathbf{u}}_m$.

We finally stop the iteration if the criterion

$$\|\boldsymbol{\ell}_{m+1}\|_2 \|\mathbf{u}_{m+1}\|_2 \leqslant \varepsilon \|\mathbf{L}_m \mathbf{U}_m\|_F \tag{3.26}$$

for some desired accuracy $\varepsilon > 0$ is met. Under the assumption that

$$\|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_{m+1}\mathbf{U}_{m+1}\|_F \leqslant \vartheta \|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_m \mathbf{U}_m\|_F$$

holds uniformly for a fixed $\vartheta < 1$, we arrive at

$$\begin{aligned}\|\boldsymbol{\ell}_{m+1}\|_2 \|\mathbf{u}_{m+1}\|_2 &= \|\mathbf{L}_{m+1}\mathbf{U}_{m+1} - \mathbf{L}_m \mathbf{U}_m\|_F \\ &\leqslant \|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_{m+1}\mathbf{U}_{m+1}\|_F + \|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_m \mathbf{U}_m\|_F \\ &\leqslant (1+\vartheta)\|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_m \mathbf{U}_m\|_F.\end{aligned}$$

On the other hand, we find

$$\begin{aligned}\|\mathbf{L}_{m+1}\mathbf{U}_{m+1} - \mathbf{L}_m \mathbf{U}_m\|_F &\geqslant \|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_m \mathbf{U}_m\|_F - \|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_{m+1}\mathbf{U}_{m+1}\|_F \\ &\geqslant (1-\vartheta)\|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_m \mathbf{U}_m\|_F.\end{aligned}$$

Therefore, we conclude that the approximation error is proportional to the product of the norms $\|\boldsymbol{\ell}_{m+1}\|_2 \|\mathbf{u}_{m+1}\|_2$ of the update vectors

$$(1-\vartheta)\|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_m \mathbf{U}_m\|_F \leqslant \|\boldsymbol{\ell}_{m+1}\|_2 \|\mathbf{u}_{m+1}\|_2 \leqslant (1+\vartheta)\|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_m \mathbf{U}_m\|_F.$$

Thus, together with (3.26), we can guarantee a relative error bound

$$\|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \mathbf{L}_m \mathbf{U}_m\|_F \lesssim \varepsilon \|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}\|_F. \tag{3.27}$$

(3.28) **Theorem.** Let $\mathbf{A}$ be the uncompressed system matrix and $\tilde{\mathbf{A}}$ be the system matrix which is compressed by the ACA. Then, with respect to the Frobenius norm, there holds the error estimate

$$\|\mathbf{A} - \tilde{\mathbf{A}}\|_F \lesssim \varepsilon \|\mathbf{A}\|_F$$

provided that the block-wise error satisfies (3.27).

Proof. In view of (3.27), we have

$$\begin{aligned}\|\mathbf{A} - \tilde{\mathbf{A}}\|_F^2 &= \sum_{j=0}^{J} \sum_{|\boldsymbol{\lambda}|,|\boldsymbol{\lambda}'|=j} \|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'} - \tilde{\mathbf{A}}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}\|_F^2 \\ &\lesssim \varepsilon^2 \sum_{j=0}^{J} \sum_{|\boldsymbol{\lambda}|,|\boldsymbol{\lambda}'|=j} \|\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}\|_F^2 \\ &= \varepsilon^2 \|\mathbf{A}\|_F^2.\end{aligned}$$

Taking square roots on both sides yields the assertion. □

Obviously, the complexity for the computation of the rank-m-approximation $\mathbf{L}_m\mathbf{U}_m$ to the block $\mathbf{A}_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}$ is $\mathcal{O}(m^2 n)$ and the storage cost is $\mathcal{O}(mn)$. The latter one can be further reduced by the application of a singular value decomposition and neglecting non-relevant singular values.

(3.29) **Remark.** The theoretical foundation of ACA for boundary integral equations is the successive interpolation of asymptotically smooth functions, cf. [Beb00]. Traditionally, ACA employs the three-dimensional interpolation theory for estimating the interpolation error relative to the boundary Γ. Since then the interpolation points may lie on a hyperplane for which the interpolation is not unique anymore, cf. [SX95], the traditional ACA may fail to converge. We refer the reader to [BG05] and [BG06], respectively, for a specific example where this happens. Nevertheless, in our framework, such situations are excluded since only the two-dimensional interpolation theory on the unit square is employed.

In the following, we restate the convergence result from [Beb00] and adapt everything to the case that the interpolation is performed on the unit square $\square$ and $\square \times \square$, respectively.

Let the function $f\colon D \times D \to \mathbb{R}$ satisfy Definition (3.5) and let $\tau_{\boldsymbol{\lambda}} \times \tau_{\boldsymbol{\lambda}'}$ be an admissible block. Consider the sequences $\{s_k\}_k, \{r_k\}_k$ given as follows. Set

$$r_0(\mathbf{s},\mathbf{t}) := f_{\boldsymbol{\lambda},\boldsymbol{\lambda}'}(\mathbf{s},\mathbf{t}) \text{ and } s_0(\mathbf{s},\mathbf{t}) := 0,$$

and compute for $k = 0, 1, \ldots$

$$\begin{aligned}r_{k+1}(\mathbf{s},\mathbf{t}) &= r_k(\mathbf{s},\mathbf{t}) - r_k(\mathbf{s}_{i_{k+1}},\mathbf{t}_{j_{k+1}})^{-1} r_k(\mathbf{s},\mathbf{t}_{j_{k+1}}) r_k(\mathbf{s}_{i_{k+1}},\mathbf{t}), \\ s_{k+1}(\mathbf{s},\mathbf{t}) &= s_k(\mathbf{s},\mathbf{t}) + r_k(\mathbf{s}_{i_{k+1}},\mathbf{t}_{j_{k+1}})^{-1} r_k(\mathbf{s},\mathbf{t}_{j_{k+1}}) r_k(\mathbf{s}_{i_{k+1}},\mathbf{t}).\end{aligned}$$

Here, we have to assume explicitly that the points $\mathbf{s}_{i_{k+1}}, \mathbf{t}_{j_{k+1}} \in \square$ are chosen such that

$$r_k(\mathbf{s}_{i_{k+1}},\mathbf{t}_{j_{k+1}})^{-1} \neq 0.$$

Then, with partial pivoting, i.e. $\mathbf{s}_{i_{k+1}}$ is chosen such that

$$|r_k(\mathbf{s}_{i_{k+1}}, \mathbf{t}_{j_{k+1}})| \geqslant |r_k(\mathbf{s}, \mathbf{t}_{j_{k+1}})| \text{ for all } \mathbf{s} \in \square,$$

the following error estimate can be proven, cf. [Beb00],

$$|r_k(\mathbf{s}, \mathbf{t})| \lesssim 2^k \operatorname{dist}(\tau_{\boldsymbol{\lambda}}, \tau_{\boldsymbol{\lambda}'})^{-2(1+q)} \eta^{\sqrt{k}}.$$

Consequently, for sufficiently small η, the remainders $|r_k(\mathbf{s}, \mathbf{t})|$ decay exponentially. According to [Beb00], the factor 2^k is not observed in most of the practical applications. Therefore, we will also omit it here for the complexity considerations which improves the results.

(3.30) **Theorem.** Assume that, for admissible clusters $\tau_{\boldsymbol{\lambda}}$ and $\tau_{\boldsymbol{\lambda}'}$, the remainder $r_k(\mathbf{s}, \mathbf{t})$ satisfies the estimate

(3.31) $$|r_k(\mathbf{s}, \mathbf{t})| \lesssim \operatorname{dist}(\tau_{\boldsymbol{\lambda}}, \tau_{\boldsymbol{\lambda}'})^{-2(1+q)} \eta^{\sqrt{k}}.$$

Then, for $\varepsilon > 0$, it holds $|r_k(\mathbf{s}, \mathbf{t})| \lesssim \varepsilon$ provided that $k \eqsim \big(|\log \varepsilon| + J(2+2q)\big)^2$.

Proof. Analogously to the proof of Theorem (3.24), it holds

$$\operatorname{dist}(\tau_{\boldsymbol{\lambda}}, \tau_{\boldsymbol{\lambda}'}) \gtrsim 2^{-J}.$$

Therefore, the assertion immediately follows from

$$\varepsilon \eqsim 2^{2J(1+q)} \eta^{\sqrt{k}} \quad \Longrightarrow \quad k \eqsim \left(\frac{\log_2 \varepsilon - 2J(1+q)}{\log_2 \eta} \right)^2. \qquad \square$$

(3.32) **Remark.** For the particular choice $\varepsilon = 2^{-J\sigma}$ in the above theorem, we observe that the rank k of the ACA behaves like the rank p^2 for the FMM. In fact, this result is in concordance with the respective results from [Gie01] and [BG05].

Although it is not necessary to introduce a threshold parameter for the far-field in the ACA, as discussed in Remark (3.19) for the FMM, we will consider it here. Hence, we arrive at the following theorem which can be proven rather analogously to Theorem (3.20).

(3.33) **Theorem.** Assume that (3.31) holds uniformly for all k. Furthermore, let p denote the threshold parameter from Remark (3.19). Then, the complexity for the computation of the far-field in the ACA is given by $\mathcal{O}(\lceil J - 2\log_4 p \rceil k^2 N_J)$ and the storage by $\mathcal{O}(\lceil J - 2\log_4 p \rceil k N_J)$.

Proof. In accordance with the proof of Theorem (3.20), the complexity for the far-field computation is given by

$$\begin{aligned} \sum_{j=0}^{\lceil J-2\log_4 p \rceil} \mathcal{O}(N_j) \cdot \mathcal{O}(k^2 N_{J-j}) &= \sum_{j=0}^{\lceil J-2\log_4 p \rceil} \mathcal{O}(K4^j) \cdot \mathcal{O}(k^2 K 4^{J-j}) \\ &= \mathcal{O}(\lceil J - 2\log_4 p \rceil k^2 K^2 4^J) \\ &= \mathcal{O}(\lceil J - 2\log_4 p \rceil k^2 N_J). \end{aligned}$$

A similar computation yields the complexity for the storage. $\square$

4. The pivoted Cholesky decomposition

The discussion in the first two sections of this chapter yields to the spatially discretized Karhunen-Loève expansion according to

$$\mathbf{u}(\omega,\mathbf{x}) \approx \mathbf{Q}_h\mathbb{E}[\mathbf{u}](\mathbf{x}) + \sum_{i=1}^{M} \sigma_{i,h} X_i(\omega)\boldsymbol{\varphi}_{i,h}(\mathbf{x}) \tag{4.1}$$

with

$$X_i := \frac{1}{\sigma_{i,h}} \int_D \left((\mathrm{Id}\otimes\mathbf{Q}_h)\mathbf{u}_0\right)^{\intercal}\boldsymbol{\varphi}_{i,h}\,\mathrm{d}\mathbf{x}$$

and $\{(\lambda_{i,h},\boldsymbol{\varphi}_{i,h})\}_{k=1}^{M}$ are the M dominant eigen-pairs of the discretized covariance operator $\mathcal{C}_h$. Based on the observation in Theorem (1.5), we consider in this section a more general approach for the representation of a random (vector-) field. For this purpose, we rather refer here to the separable decomposition

$$\mathbf{u}_{h,M}(\omega,\mathbf{x}) = \mathbf{Q}_h\mathbb{E}[\mathbf{u}](\mathbf{x}) + \sum_{i=1}^{M} Y_i(\omega)\boldsymbol{\psi}_{i,h}(\mathbf{x}) \tag{4.2}$$

than to the orthogonal decomposition (4.1). In the expansion (4.2), we assume that $\{\boldsymbol{\psi}_{i,h}\}_{i=1}^{M} \subset [V_h^s]^d$ for appropriately modified random variables $\{Y_m\}_{m=1}^{M} \subset L^2_{\mathbb{P}}(\Omega)$.

Data: matrix $\mathbf{A} = [a_{i,j}] \in \mathbb{R}^{n\times n}$ and error tolerance $\varepsilon > 0$
Result: low-rank approximation $\mathbf{A}_M = \sum_{i=1}^{M} \boldsymbol{\ell}_i\boldsymbol{\ell}_i^{\intercal}$ such that $\mathrm{trace}(\mathbf{A}-\mathbf{A}_M) \le \varepsilon$
begin
  set $M := 1$;
  set $\mathbf{d} := \mathrm{diag}(\mathbf{A})$ and $error := \|\mathbf{d}\|_{\ell^1}$;
  initialize $\boldsymbol{\pi} := [1,2,\ldots,n]$;
  while $error > \varepsilon$ **do**
    set $i := \arg\max\{d_{\pi_j} : j = M, M+1,\ldots,n\}$;
    swap π_M and π_i;
    set $\ell_{M,\pi_M} := \sqrt{d_{\pi_M}}$;
    for $M+1 \le i \le n$ **do**
      compute $\ell_{M,\pi_i} := \left(a_{\pi_M,\pi_i} - \sum_{j=1}^{M-1} \ell_{j,\pi_M}\ell_{j,\pi_i}\right) \Big/ \ell_{M,\pi_M}$;
      update $d_{\pi_i} := d_{\pi_i} - \ell_{M,\pi_M}\ell_{M,\pi_i}$;
    end
    compute $error := \sum_{i=M+1}^{n} d_{\pi_i}$;
    increase $M := M+1$;
  end
end

Algorithm 1: Pivoted Cholesky decomposition.

One possibility to obtain a separable expansion (4.2) is to compute the pivoted Cholesky decomposition of the coefficient matrix of $\mathcal{C}_h$ with respect to a basis in $[V_h^s]^d$. To that end, let $\mathbf{\Phi}(\mathbf{x}) := [\varphi_1 \otimes \mathbf{e}_1, \dots, \varphi_N \otimes \mathbf{e}_1, \dots, \varphi_1 \otimes \mathbf{e}_d, \dots, \varphi_N \otimes \mathbf{e}_d]$ denote an orthonormal basis of $[V_h^s]^d$, where $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ denotes the canonical basis in $\mathbb{R}^d$. Then, the coefficient matrix of $\mathcal{C}_h$ with respect to $\mathbf{\Phi}$ is given by the block-matrix

$$\mathbf{C} = [(\mathcal{C}\boldsymbol{\varphi}_j, \boldsymbol{\varphi}_i)_{L^2(D;\mathbb{R}^d)}]_{i,j=1}^{dN} = \begin{bmatrix} \mathbf{C}_{1,1} & \cdots & \mathbf{C}_{1,d} \\ \vdots & & \vdots \\ \mathbf{C}_{d,1} & \cdots & \mathbf{C}_{d,d} \end{bmatrix} \in \mathbb{R}^{dN \times dN}. \tag{4.3}$$

For each finite dimensional ansatz space, the matrix $\mathbf{C}$ is symmetric and positive semidefinite. Thus, $\mathbf{C}$ exhibits a (possibly pivoted) Cholesky decomposition. By pivoting the Cholesky decomposition as seen in Algorithm 1, we achieve numerical stability on the one hand, cf. [Hig90, Hig02], and, if the eigenvalues of $\mathbf{C}$ decay sufficiently fast, a low-rank approximation on the other hand, cf. [HPS12]. Especially, the approximation error of the (truncated) pivoted Cholesky decomposition is a-posteriori controllable in terms of the (discrete) trace, i.e. $\operatorname{trace}(\mathbf{A}) := \sum_{i=1}^n \langle \mathbf{A}\mathbf{e}_i, \mathbf{e}_i \rangle = \sum_{i=1}^n a_{i,i}$, for $\mathbf{A} \in \mathbb{R}^{n \times n}$ and the canonical basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ in $\mathbb{R}^n$.

(4.4) **Remark.** Notice that ACA combined with total pivoting would result for symmetric and positive semidefinite matrix blocks in an algorithm which is quite similar to the pivoted Cholesky decomposition. Nevertheless, for PCD, we do not have to partition the system matrix into far- and near-field, but can directly apply Algorithm 1 to $\mathbf{C}$. In this sense, we may think of PCD as a single-block ACA with total pivoting. Here, the total pivoting is not prohibitive expansive since it is a-priori known that the pivots are located on the main diagonal of $\mathbf{C}$. Furthermore, we have in contrast to ACA, a rigorous stopping criterion based on the quantity $\operatorname{trace}(\mathbf{C} - \mathbf{L}_M \mathbf{L}_M^\intercal)$.

In the following, we establish the connection between the approximation to the random field obtained by the pivoted Cholesky decomposition and the Karhunen-Loève expansion of $(\mathrm{Id} \otimes \mathbf{Q}_h)\mathbf{u}(\omega, \mathbf{x})$. We denote the spectral decomposition related to $\mathbf{C}$ by $\mathbf{C} = \sum_{i=1}^{dN} \lambda_i \mathbf{v}_i \mathbf{v}_i^\intercal$ with $\lambda_i \in [0, \infty)$ and $\mathbf{v}_i \in \mathbb{R}^{dN}$. Therefore, the Karhunen-Loève expansion of $(\mathrm{Id} \otimes \mathbf{Q}_h)\mathbf{u}(\omega, \mathbf{x})$ is given by

$$\mathbf{u}_h(\omega, \mathbf{x}) = \mathbf{Q}_h \mathbb{E}[\mathbf{u}](\mathbf{x}) + \sum_{i=1}^{dN} \sqrt{\lambda_i} X_i(\omega) \mathbf{\Phi}(\mathbf{x}) \mathbf{v}_i$$

with respect to the orthonormal basis $\mathbf{\Phi}$ of $[V_h^s]^d$. This representation can be rewritten in matrix notation as

$$\mathbf{u}_h(\omega, \mathbf{x}) - \mathbf{Q}_h \mathbb{E}[\mathbf{u}](\mathbf{x}) =: \mathbf{\Phi}(\mathbf{x}) \mathbf{V} \mathbf{\Sigma} \mathbf{X}(\omega) \tag{4.5}$$

with $\mathbf{V} := [\mathbf{v}_1, \dots, \mathbf{v}_{dN}]$, $\mathbf{\Sigma} := \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_{dN}})$ and $\mathbf{X}(\omega) := [X_1(\omega), \dots, X_{dN}(\omega)]^\intercal$. The matrix $(\mathbf{V}\mathbf{\Sigma})^\intercal \in \mathbb{R}^{dN \times dN}$ from (4.5) exhibits a QR-decomposition, $\mathbf{Q}\mathbf{L}^\intercal = (\mathbf{V}\mathbf{\Sigma})^\intercal$ or $\mathbf{L}\mathbf{Q}^\intercal = \mathbf{V}\mathbf{\Sigma}$, respectively. Here, $\mathbf{Q}$ denotes an orthogonal matrix, i.e. $\mathbf{Q}^\intercal \mathbf{Q} = \mathbf{I} \in \mathbb{R}^{dN \times dN}$,

and $\mathbf{L} \in \mathbb{R}^{dN \times dN}$ is a lower triangular matrix. We shall next define the transformed random vector $\mathbf{Y}(\omega) := \mathbf{Q}^\intercal \mathbf{X}(\omega)$. Then, $\mathbf{Y}(\omega)$ also consists of dN uncorrelated and centered random variables, since it holds

$$\int_\Omega \mathbf{Y}(\omega)\mathbf{Y}^\intercal(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \mathbf{Q}^\intercal \int_\Omega \mathbf{X}(\omega)\mathbf{X}^\intercal(\omega)\,\mathrm{d}\mathbb{P}(\omega)\mathbf{Q} = \mathbf{Q}^\intercal\mathbf{I}\mathbf{Q} = \mathbf{I}.$$

That the random variables $Y_i(\omega)$ are also centered follows from the fact that they are weighted sums of centered random variables. Thus, we obtain a representation which is equivalent to (4.5) according to

$$\mathbf{\Phi}(\mathbf{x})\mathbf{V}\mathbf{\Sigma}\mathbf{X}(\omega) = \mathbf{\Phi}(\mathbf{x})\mathbf{L}\mathbf{Y}(\omega)$$

where the change of basis $\mathbf{Q}$ only acts on the basis of $L^2_{\mathbb{P}}(\Omega)$. Moreover, we observe

$$\mathbf{C} = \mathbf{V}\mathbf{\Sigma}(\mathbf{V}\mathbf{\Sigma})^\intercal = \mathbf{L}\mathbf{Q}^\intercal\mathbf{Q}\mathbf{L}^\intercal = \mathbf{L}\mathbf{L}^\intercal.$$

Since $\mathbf{L}$ is a lower triangular matrix, we thus end up with the Cholesky decomposition of $\mathbf{C}$. In the following, without loss of generality, we shall assume that $\mathbf{L}\mathbf{L}^\intercal$ corresponds to the pivoted Cholesky decomposition of $\mathbf{C}$.

Using the Cholesky decomposition of $\mathbf{C}$, we obtain the separable representation

$$\mathbf{u}_h(\omega, \mathbf{x}) = \mathbf{Q}_h\mathbb{E}[\mathbf{u}](\mathbf{x}) + \sum_{i=1}^{dN} Y_i(\omega)\mathbf{\Phi}(\mathbf{x})\boldsymbol{\ell}_i \tag{4.6}$$

for $\mathbf{u}_h(\omega, \mathbf{x})$ with $\mathbf{L} = [\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_N]$. Whereas, the related truncated Cholesky decomposition leads to the truncated expansion

$$\mathbf{u}_{h,M}(\omega, \mathbf{x}) = \mathbf{Q}_h\mathbb{E}[\mathbf{u}](\mathbf{x}) + \sum_{i=1}^{M} Y_i(\omega)\mathbf{\Phi}(\mathbf{x})\boldsymbol{\ell}_i.$$

Notice that this is exactly the representation (4.2) with $\boldsymbol{\psi}_{i,h}(\mathbf{x}) = \mathbf{\Phi}(\mathbf{x})\boldsymbol{\ell}_i$.

(4.7) **Remark.** The separable representation (4.6) of the stochastic field is based on the knowledge of an appropriate matrix $\mathbf{R} \in \mathbb{R}^{dN \times dN}$, a square root of the coefficient matrix, such that $\mathbf{C} = \mathbf{R}\mathbf{R}^\intercal$. It is known that for two different square roots, i.e. $\mathbf{C} = \mathbf{R}\mathbf{R}^\intercal = \tilde{\mathbf{R}}\tilde{\mathbf{R}}^\intercal$, there exists an orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{dN \times dN}$ such that $\tilde{\mathbf{R}} = \mathbf{R}\mathbf{Q}^\intercal$. The change of the representation (4.5) due to the application of $\mathbf{Q}$ is then performed by the change of the basis in $L^2_{\mathbb{P}}(\Omega)$, i.e. $\mathbf{Y}(\omega) := \mathbf{Q}^\intercal\mathbf{X}(\omega)$. Thus, any square root of $\mathbf{C}$ yields a separable representation of $\mathbf{u}_h(\omega, \mathbf{x})$. Nevertheless, we focus on the pivoted Cholesky decomposition here.

The approximation error of a given stochastic field which is induced by truncating of the pivoted Cholesky decomposition is controllable in accordance with the following theorem.

(4.8) **Theorem.** Let $\mathbf{C} \in \mathbb{R}^{dN \times dN}$ denote the coefficient matrix given by (4.3). Furthermore, let $\mathbf{C}_M = \mathbf{L}_M\mathbf{L}_M^\intercal \in \mathbb{R}^{dN \times dN}$ denote its (truncated) pivoted Cholesky decomposition computed by Algorithm 1 such that $\operatorname{trace}(\mathbf{C} - \mathbf{C}_M) < \varepsilon$ holds for some $\varepsilon > 0$. Then, for the related stochastic field, we have the error estimate $\|\mathbf{u}_h - \mathbf{u}_{h,M}\|_{L^2_{\mathbb{P}}(\Omega;L^2(D))} < \sqrt{\varepsilon}$.

Proof. Let $\mathbf{C} = \mathbf{L}\mathbf{L}^\intercal$ be the pivoted Cholesky decomposition of $\mathbf{C}$. We define the related integral operator

$$(T^\star \mathbf{v})(\omega) := \int_D \left(\mathbf{\Phi}(\mathbf{x})\mathbf{L}\mathbf{Y}(\omega)\right)^\intercal \mathbf{v}(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

and the orthogonal projection $\mathbf{P}_h \colon [V_h^s]^d \to \operatorname{span}\{\boldsymbol{\varphi}_1(\mathbf{x}), \dots, \boldsymbol{\varphi}_M(\mathbf{x})\}$ onto the space which is spanned by the first M basis functions. Then, it holds in complete analogy to the proof of Theorem (1.3) that

$$\begin{aligned}
\|\mathbf{u}_h - \mathbf{u}_{h,M}\|^2_{L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^d)} &= \sum_{i=1}^{dN} (T^\star \boldsymbol{\varphi}_i, T_h^\star \boldsymbol{\varphi}_i)_{L^2_{\mathbb{P}}(\Omega)} - (T^\star \mathbf{P}_h \boldsymbol{\varphi}_i, T^\star \mathbf{P}_h \boldsymbol{\varphi}_i)_{L^2_{\mathbb{P}}(\Omega)} \\
&= \sum_{i=1}^{dN} \langle \mathbf{L}^\intercal \mathbf{e}_i, \mathbf{L}^\intercal \mathbf{e}_i \rangle - \sum_{i=1}^{M} \langle \mathbf{L}^\intercal \mathbf{e}_i, \mathbf{L}^\intercal \mathbf{e}_i \rangle \\
&= \operatorname{trace}(\mathbf{C} - \mathbf{C}_M) < \varepsilon.
\end{aligned}$$

□

This theorem states that the choice $\varepsilon \eqsim h^{2\min\{s+1,q\}}$ in the pivoted Cholesky decomposition guarantees, together with inequality (1.6), the (optimal) error estimate

$$\|\mathbf{u} - \mathbf{u}_{h,M}\|_{L^2_{\mathbb{P}}(\Omega)\otimes L^2(D;\mathbb{R}^d)} \lesssim h^{\min\{s+1,q\}}.$$

The major advantage of this approach is that at no time the coefficient matrix $\mathbf{C}$ has to be fully assembled. It is sufficient to provide access to single entries of this matrix while processing the pivoted Cholesky decomposition. The error in the approximation of the random field $\mathbf{u}(\omega, \mathbf{x})$ is then a-posteriori controllable by the trace.

Given that the pivoted Cholesky decomposition for $\mathbf{C}$ terminates with $M \ll dN$ terms and $\mathbf{C}_M = \mathbf{L}_M \mathbf{L}_M^\intercal \in \mathbb{R}^{dN \times dN}$, where $\mathbf{L}_M \in \mathbb{R}^{dN \times M}$, the computation of the related Karhunen-Loève expansion is performed with complexity $\mathcal{O}(M^2 dN)$, cf. [HPS12]. This is achieved by computing the eigenvalues of $\mathbf{L}_M^\intercal \mathbf{L}_M \in \mathbb{R}^{M \times M}$ which coincide with those of $\mathbf{C}_M$. Then, if $\mathbf{v}_1, \dots, \mathbf{v}_M$ denote the orthonormal eigenvectors of the small eigen-problem, the eigenvectors of $\mathbf{C}_M$ are given by $\mathbf{L}\mathbf{v}_1, \dots, \mathbf{L}\mathbf{v}_M$ and we have

$$(\mathbf{L}\mathbf{v}_i)^\intercal (\mathbf{L}\mathbf{v}_j) = \mathbf{v}_i \mathbf{L}^\intercal \mathbf{L} \mathbf{v}_j = \lambda_i \delta_{i,j} \quad \text{for all } i, j = 1, \dots, M. \tag{4.9}$$

Thus, the related Karhunen-Loève decomposition is given by

$$\mathbf{u}_{h,M}(\omega, \mathbf{x}) = \mathbf{Q}_h \mathbb{E}[\mathbf{u}](\mathbf{x}) + \sum_{i=1}^{M} \tilde{X}_i(\omega) \mathbf{\Phi}(\mathbf{x}) \mathbf{L} \mathbf{v}_i \tag{4.10}$$

for appropriately chosen random variables $\tilde{X}_1(\omega), \dots, \tilde{X}_M(\omega)$. If the laws of the random variables $X_i(\omega)$ are known, we obtain the relation

$$\tilde{\mathbf{X}}(\omega) = [\mathbf{v}_1, \dots, \mathbf{v}_M]^\intercal \mathbf{Y}(\omega)$$

with $\mathbf{v}_i$ from (4.10). Otherwise, the related random variables $\tilde{X}_i(\omega)$ can be determined by a *maximum likelihood* estimate, cf. [ST06]. Notice that for the important Gaussian case $\mathbf{X}(\omega) \sim [\mathcal{N}(0,1)]^{dN}$, we have for any orthogonal transform $\mathbf{Q}\mathbf{X}(\omega) \sim [\mathcal{N}(0,1)]^{dN}$ and thus $\tilde{\mathbf{X}}$ has the same law as $\mathbf{X}$. Nevertheless, we consider here only distributions with compact range. Therefore, the Gaussian case will be of no interest later on.

5. Numerical results

We have already pointed out that the cluster methods provide a means to approximate single blocks of the matrix-valued covariance function Cov[**u**]. Thus, to keep things simple in the numerical tests, we will consider here only covariance functions related to random scalar fields $u \in L^2_{\mathbb{P}}(\Omega, L^2(D))$. The covariance functions under consideration shall then be given by the Matérn class of covariance functions, cf. Section II.4. The parametric geometries we study here are given by the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ on the one hand and a more complex geometry $D \subset \mathbb{R}^2$, i.e. a rectangular plate with holes, on the other hand.

The implementations of namely ACA, FMM and PCD rely on the same basic routines, which means that they use the same quadrature routines for the assembly of the entries of the matrix $\mathbf{C}$. Especially, we employ here the quadrature proposed in [SS97] for the treatment of singular integrals. Although, all of the kernel functions under consideration are at least continuous, we observed that this quadrature provides more accurate approximations. Exploiting the symmetry of $\mathbf{C}$, we only assemble its lower triangular part of the matrix for ACA and FMM. We refer to the related compressed matrix as $\tilde{\mathbf{C}}$. The constant in the admissibility condition (3.15) is set to $\eta = 1.6$ for both methods. Although, this choice contradicts the requirement $\eta \in (0, 1)$, it is sufficient for the numerical examples at hand. All methods have been implemented in the C-programming language, cf. [KR88]. In the implementation of ACA, we have employed level 1 and 2 BLAS[1] routines in the assembly of the matrix $\tilde{\mathbf{C}}$ and in the matrix-vector product, whenever possible. In the implementation of the FMM, the matrix-vector product is based on the $\mathcal{H}^2$-matrix variant.

In case of ACA and FMM, we end up with a data-sparse representation $\tilde{\mathbf{C}}$ of $\mathbf{C}$. Thus, to obtain a representation of the approximate stochastic field $u_{h,M}$ similar to (4.1), we have still to compute the dominant eigen-pairs of $\tilde{\mathbf{C}}$. The representation of $\tilde{\mathbf{C}}$ provides a fast matrix-vector product. Therefore, we employ ARPACK, cf. [LSY98], to solve the eigen-problem for the compressed matrix $\tilde{\mathbf{C}}$. The size of the Krylov subspace in ARPACK is chosen as twice the number of desired eigenvalues, which is a reasonable choice according to [LSY98].

Notice that a LAPACK[2]-style implementation of the PCD, cf. [Luc04], is not applicable since it relies on the assembly of the entire matrix $\mathbf{C}$, which is not feasible for large values of N.[3]

All computations performed on a computing server with two Intel(R) Xeon(R) X5550 CPUs with a clock rate of 2.67GHz and 48GB of main memory. The computations have been carried out single-threaded, i.e. on a single core. Furthermore, we set the correlation length of the Matérn kernels to $\ell = 1$ in all examples. For the spatial discretization, we choose piecewise constant finite elements, i.e. the ansatz space is V_h^0, cf. Remark (II.6.6). For ACA and PCD, we set the truncation error $\varepsilon \approx h^2$. The FMM works with polynomial degree $p = 3$ which is sufficient to maintain the convergence for all levels of refinement. Thus, we expect for both approaches a linear rate of convergence for

[1] http://www.netlib.org/blas/

[2] http://www.netlib.org/lapack/

[3] For example for $N = 10^5$, the storage of the matrix $\mathbf{C}$ would require about 80GB of memory in 8-Byte double precision.

the stochastic field in terms of the (continuous) traces, i.e.

$$\|u - u_{h,M}\|_{L^2_{\mathbb{P}}(\Omega;L^2(D))} \lesssim h + \sqrt{\operatorname{Tr} \mathcal{C}_h - \operatorname{Tr} \mathcal{C}_{h,M}}.$$

Therefore, we will measure the error by the quantity $\sqrt{\operatorname{Tr} \mathcal{C}_h - \operatorname{Tr} \mathcal{C}_{h,M}}/\sqrt{\operatorname{Tr} \mathcal{C}_h}$. Notice that we consider here the relative error in order to make the error independent of the scaling of the eigenvalues which depends on the size of the domain, cf. (II.4.6).

First example

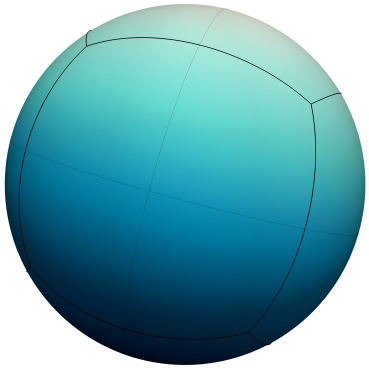

Figure III.4: The unit sphere $\mathbb{S}^2$ represented by 6 patches.

As a benchmark, we consider the three-dimensional unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ represented by 6 congruent patches, see Figure III.4. Thus, with the knowledge from the preceding Section II.4, we can compute the exact eigenvalues and eigenfunctions of the Matérn covariance functions as reference. Furthermore, we can estimate the truncation error due to (II.4.4). For the truncation error related to the Matérn covariance with smoothness parameter ν, it holds that

$$\sqrt{\sum_{m=M}^{\infty} \lambda_m} \lesssim \sqrt{\int_M^{\infty} C x^{-1-\nu} \,\mathrm{d}x} = \sqrt{\frac{1}{\nu} C M^{-\nu}}. \tag{5.1}$$

Notice that the dimension is $d = 2$ here, since we restrict the Matérn kernels to the unit sphere $\mathbb{S}^2$. Thus, to bound the truncation error of the Karhunen-Loève expansion by h, we have to ensure that

$$\sqrt{\frac{1}{\nu} C M^{-\nu}} \leqslant h \quad \Longrightarrow \quad M \geqslant \left(\frac{C}{\nu h^2}\right)^{\frac{1}{\nu}}.$$

With the estimation of the constant C from Figure II.3 at hand, we are able to compute the related length of the Karhunen-Loève expansion. Unfortunately, this approach yields very large numbers of eigen-pairs to be approximated by ACA and FMM. Therefore, we choose another approach. We consider for each respective kernel the sum

j	$\nu = 3/2$	$\nu = 5/2$	$\nu = 7/2$	$\nu = 9/2$
1	6 (9)	4 (4)	4 (4)	4 (4)
2	18 (25)	13 (16)	11 (16)	9 (9)
3	48 (49)	25 (25)	20 (25)	17 (25)
4	120 (121)	45 (49)	33 (36)	26 (36)
5	305 (324)	79 (81)	49 (49)	40 (49)
6	768 (789)	139 (144)	76 (81)	57 (64)
7	1928 (1936)	243 (256)	113 (121)	78 (81)
8	4807 (4900)	423 (441)	166 (169)	107 (121)

Table III.1: Different values of the cut-off parameter M_j on the unit sphere $\mathbb{S}^2$.

j	$\nu = 3/2$	$\nu = 5/2$	$\nu = 7/2$	$\nu = 9/2$
1	5 (6)	5 (6)	4 (5)	4 (5)
2	19 (21)	14 (14)	12 (13)	11 (12)
3	49 (56)	29 (32)	23 (24)	21 (22)
4	137 (158)	53 (58)	38 (41)	32 (35)
5	359 (414)	97 (107)	58 (62)	46 (49)
6	935 (1082)	167 (185)	89 (96)	64 (69)
7	2415 (2812)	295 (327)	132 (143)	90 (96)
8	– (7158)	513 (569)	197 (214)	122 (130)

Table III.2: Ranks determined by PCD on the unit sphere $\mathbb{S}^2$.

of those eigenvalues with magnitude larger than 10^{-10} as an approximation to the actual trace of the kernel, i.e.

$$\int_{\mathbb{S}^2} k_\nu(\mathbf{x},\mathbf{x})\,\mathrm{d}s_{\mathbf{x}} = \sum_{m=1}^{M_{\max}} \lambda_m + \delta \quad \text{with } M_{\max} = \arg\min_m\{\lambda_m > 10^{-10}\},$$

cf. (II.4.6). The resulting truncation error δ is computable due to the exact knowledge of the traces which are equal to 4π for all ν. We have $\delta = 4.18 \cdot 10^{-6}$ for $\nu = 3/2$, $\delta = 2.43 \cdot 10^{-7}$ for $\nu = 5/2$, $\delta = 4.93 \cdot 10^{-8}$ for $\nu = 7/2$, and $\delta = 1.70 \cdot 10^{-8}$ for $\nu = 9/2$. The rank on each level j with mesh width $h \approx 2^{-j}$ is then determined according to

$$M_j = \arg\min_{k\in\{1,\dots,M_{\max}\}} \left\{ \sum_{m=1}^{M_{\max}} \lambda_m - \sum_{m=1}^{k} \lambda_m < h^2 \sum_{m=1}^{M_{\max}} \lambda_m \right\}, \tag{5.2}$$

that is the trace error relative to the scaling of the eigenvalues. The finest level j which we consider here is $j = 8$, resulting in 393216 finite elements. For the levels $j = 1,\dots,8$ and $\nu = 3/2, 5/2, 7/2, 9/2$, the computed cut-off parameters M_j are found in Table III.1. The number in the brackets denotes the size necessary to resolve clusters of eigenvalues by approximating only complete subspaces associated with the multiplicity of the respective eigenvalue. This is proposed in [LSY98] in order to achieve the optimal performance of ARPACK. Nevertheless, in our numerical studies in [HPS14a], we could not observe major differences in neither the computational time nor the precision obtained of the

eigen-pairs. Therefore, we will show here only the results for the number of eigen-pairs which is provided by (5.2).

Table III.2 shows the ranks determined by PCD. The numbers in front of the brackets correspond to the recompressed ranks, the numbers within the brackets denote the original rank. As it turns out, the ranks computed by PCD are rather optimal in the sense that they reflect the estimated length of the Karhunen-Loéve expansion determined by formula (II.4.4). Especially for increasing smoothness of the kernel function, the determined rank gets successively better.

(5.3) **Remark.** We end up with the spectral decomposition of the approximate covariance $\mathcal{C}_{h,M}$ when we solve the eigen-problem (4.9) for PCD. By truncating the obtained decomposition (4.10) with the prescribed relative accuracy h^2, we achieve an a-posteriori recompression of the PCD. This procedure may at most double the approximation error but reduces the rank by up to 10% on average in our computations for this thesis.

The error plots and related computational times for the numerical experiments on the unit sphere are depicted in Figure III.5–III.8. Unfortunately, the computations of ACA and FMM as well as PCD with recompression for $\nu = 3/2$ and level 8, i.e. for 393216 finite elements, could not be carried out since the available main memory has been insufficient.

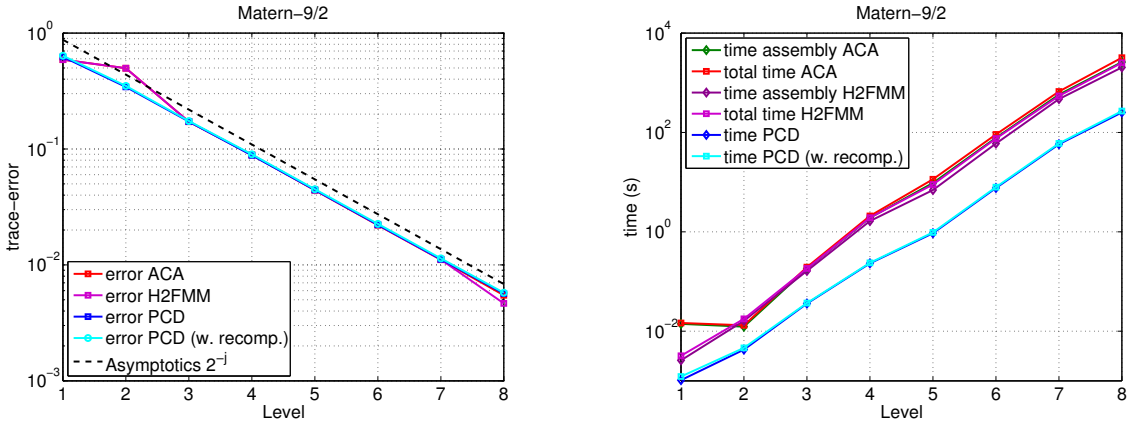

Figure III.5: Trace error (left) and computational times (right) for the $\nu = 9/2$ on $\mathbb{S}^2$.

In the plots on the left hand side of Figures III.5–III.8, the trace error for each particular kernel under consideration is depicted. The expected rate $h \eqsim 2^{-j}$ is indicated in the plots by the dashed black line. The red line shows the error for ACA and the magenta line shows the error for FMM. The error of the PCD is indicated by the blue line and finally the error of PCD with recompression is indicated by the cyan colored line. It turns out that all methods provide at least the expected linear rate of convergence in this example. The rate of FMM is even slightly increased on level 8.

In the plots on the right hand side of Figures III.5–III.8, the computational times for every method and each particular kernel is found. There seems to be no significant difference in the times for ACA and FMM for all kernels under consideration. Furthermore, we observe that the computation time consumed by ACA and FMM for $\nu = 5/2, 7/2, 9/2$ is

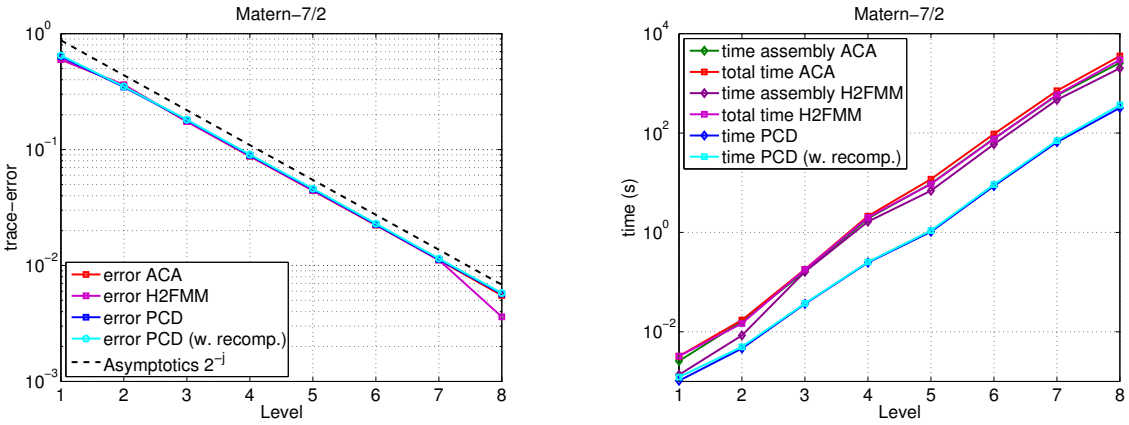

Figure III.6: Trace error (left) and computational times (right) for the $\nu = 7/2$ on $\mathbb{S}^2$.

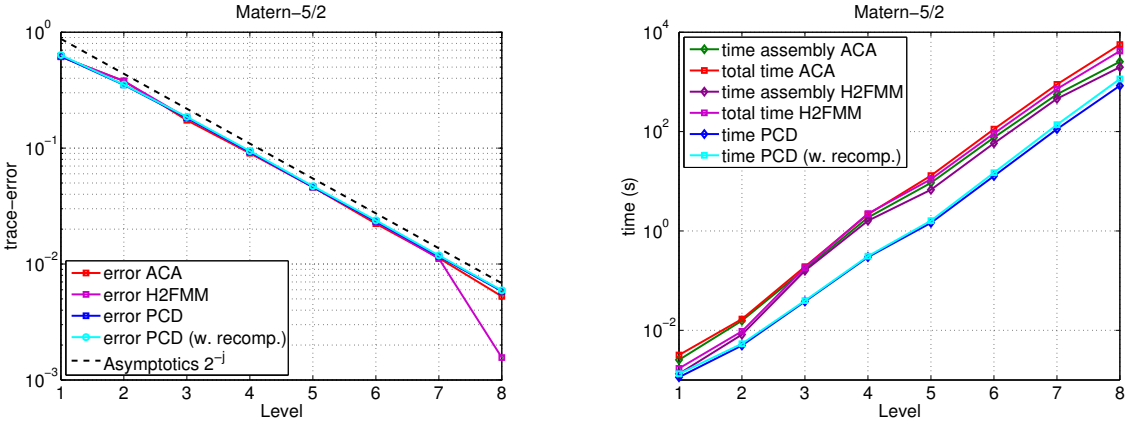

Figure III.7: Trace error (left) and computational times (right) for the $\nu = 5/2$ on $\mathbb{S}^2$.

dominated by the assembly of the matrix $\tilde{\mathbf{C}}$, indicated by the green line for ACA and by the dark magenta line for FMM, whereas the computation time is governed by the eigenvalue computation for $\nu = 3/2$. Nevertheless, we observe that PCD becomes significantly faster than ACA and FMM for increasing smoothness of the kernels. In order to quantify this behaviour, we have plotted in Figure III.9 the average computational times for ACA and FMM relative to the computational time of the PCD with recompression. The bar graph shows that PCD is nearly a factor 3 times faster than ACA and FMM for $\nu = 3/2$ up to factor of 9 and 8 times faster than ACA and FMM for $\nu = 9/2$, respectively.

Second example

In our second example, we consider the plate geometry depicted in Figure III.10. It is a rectangle with 30 inscribed, equi-spaced circular holes which is represented by 120 patches and scaled to a size of 2×2.4. Here, the computations are carried out on levels $j = 1, \ldots, 6$, where level 6 corresponds to 491520 finite elements. Figure III.11 contains a visualization of the four orthonormal eigenfunctions corresponding to the four largest eigenvalues of the

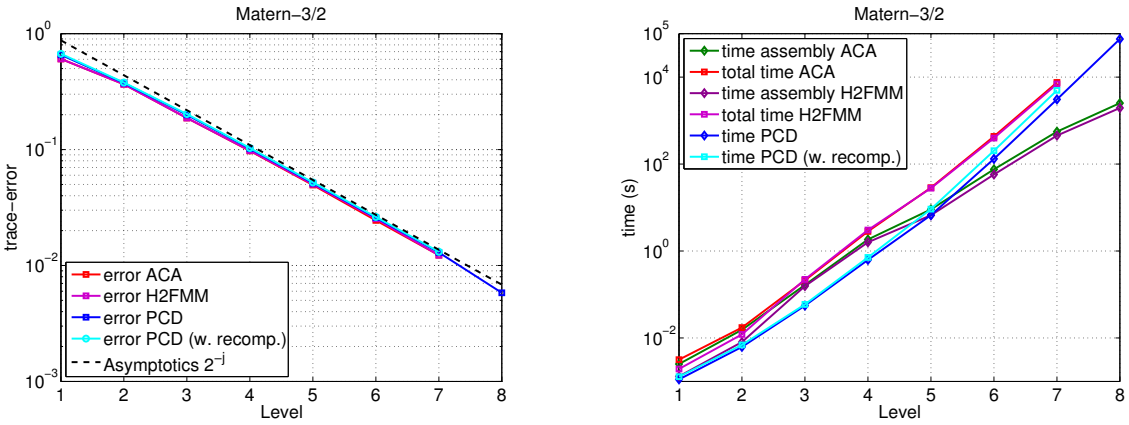

Figure III.8: Trace error (left) and computational times (right) for the $\nu = 3/2$ on $\mathbb{S}^2$.

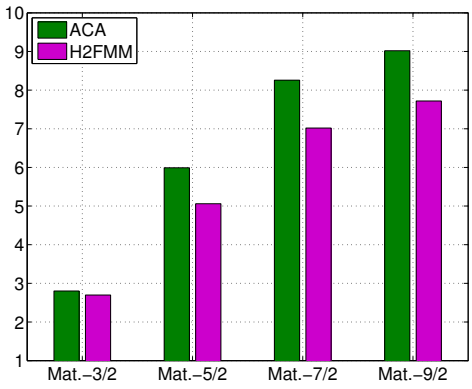

Figure III.9: Computational times relative to PCD on the unit sphere $\mathbb{S}^2$.

Matérn kernel with $\nu = 3/2$. For overview purposes, we have chosen the same colors for each particular method as in the previous visualizations.

In this example, we do not know the number of eigenvalues necessary to achieve the desired precision with ACA and FMM for ARPACK. Therefore, we use here the ranks provided by PCD with recompression as reference. The respective values are found in Table III.3. Again, the numbers in front of the brackets correspond to the recompressed ranks of PCD and the numbers within the brackets denote the original ranks.

The error plots and related computational times for the numerical experiments on the plate geometry are presented in Figures III.12–III.16. The trace error of each particular kernel, i.e. $\nu = 3/2, 5/2, 7/2, 9/2, \infty$, for the different methods is found on the left hand side of Figures III.12–III.16. Again, PCD and FMM provide exactly the expected rate of convergence. Moreover, the FMM exhibits a slightly better constant in the error. The behaviour of ACA is not that monotone as in the previous example. In case of the smoother kernels, i.e. $\nu = 7/2, 9/2, \infty$, the rate of convergence increases on level 4 and deteriorates on levels 5 and 6 and becomes constant. For $\nu = \infty$ the error even increases. Considering $\nu = 5/2$, we have an increased rate on level 4 but also an increasing error on level 5. For $\nu = 3/2$ we observe a rather constant error on levels 3–5 and an increased rate

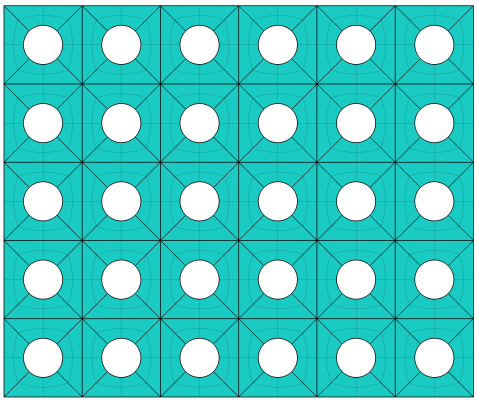

Figure III.10: Plate geometry represented by 120 patches.

j	$\nu = 3/2$	$\nu = 5/2$	$\nu = 7/2$	$\nu = 9/2$	$\nu = \infty$
1	13 (14)	11 (11)	9 (9)	9 (9)	7 (7)
2	34 (36)	19 (20)	15 (15)	14 (14)	10 (10)
3	78 (86)	31 (33)	25 (26)	20 (21)	13 (13)
4	178 (196)	52 (56)	35 (37)	29 (30)	17 (17)
5	416 (459)	87 (93)	49 (52)	38 (39)	21 (21)
6	983(1085)	141 (151)	71 (75)	53 (55)	25 (26)

Table III.3: Ranks determined by PCD on the plate geometry.

on level 6. This results in the correct rate of convergence for this kernel. Possibly, these effects are caused by a failure of the error estimator (3.26) on this geometry which results in a bad approximation of $\tilde{\mathbf{C}}$.

In the plots on the right hand side of Figures III.12–III.16, the computational times for every method and each particular kernel are shown. Notice that the additional time for the recompression of PCD becomes negligible due to the low ranks. Furthermore, the benefit of the recompression is relatively small here, especially for the smoother kernels, cf. Table III.3. Again, PCD outperforms the cluster methods. The related relative computation times are depicted in Figure III.17. In this example, we see the speed-up of PCD more clearly. This is due to the coupling of the number of eigenvalues to be computed to the ranks provided by PCD. The speed-up lies between a factor of 8 (7) for ACA (FMM) for $\nu = 3/2$ and a factor of 14 (12) for ACA (FMM) for $\nu = \infty$. Furthermore, as in the previous example, the computation time for ACA and FMM is governed by the assembly of the matrix for $\nu = 5/2, 7/2, 9/2, \infty$ and by the eigenvalue computation for $\nu = 3/2$.

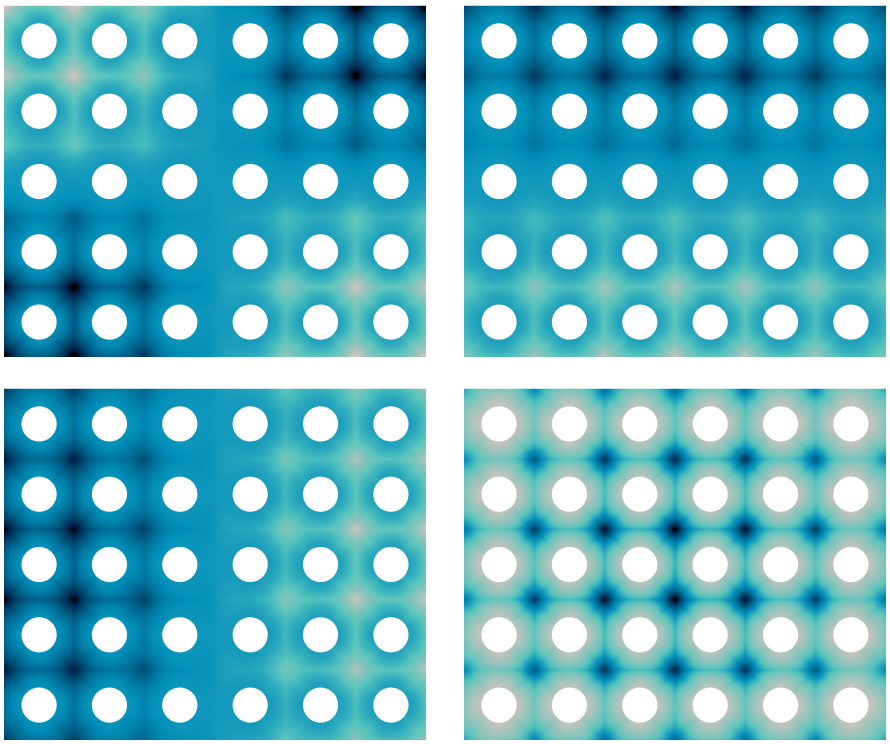

Figure III.11: First four orthonormal eigenfunctions on the plate geometry and Matérn kernel for $\nu = 3/2$.

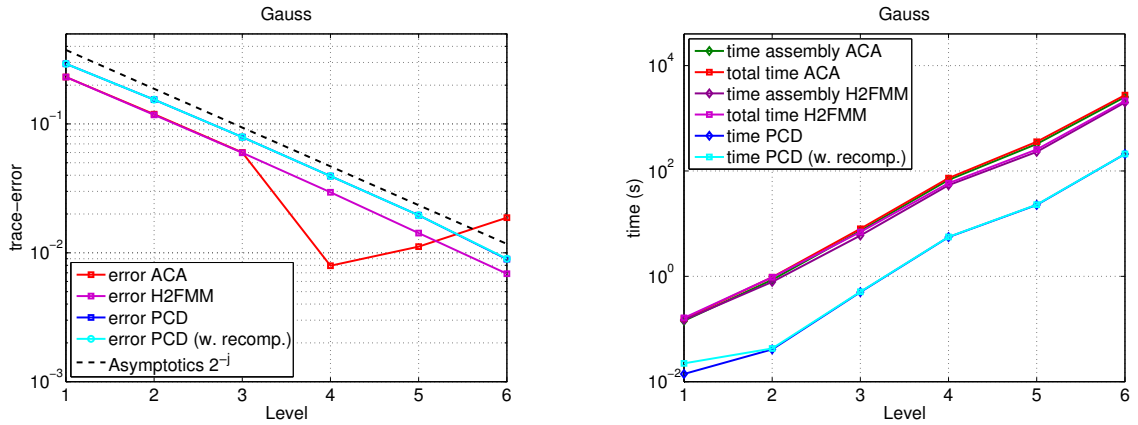

Figure III.12: Trace error (left) and computational times (right) for the $\nu = \infty$ on the plate geometry.

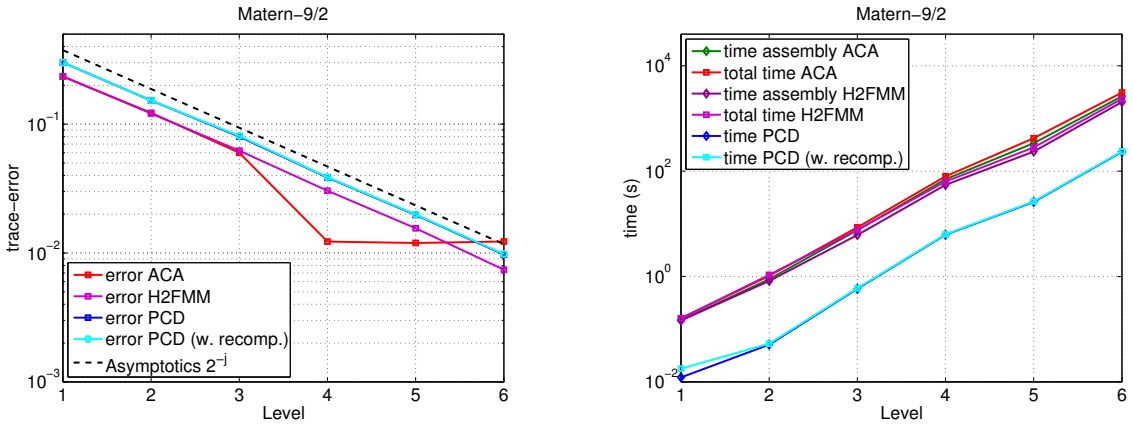

Figure III.13: Trace error (left) and computational times (right) for the $\nu = 9/2$ on the plate geometry.

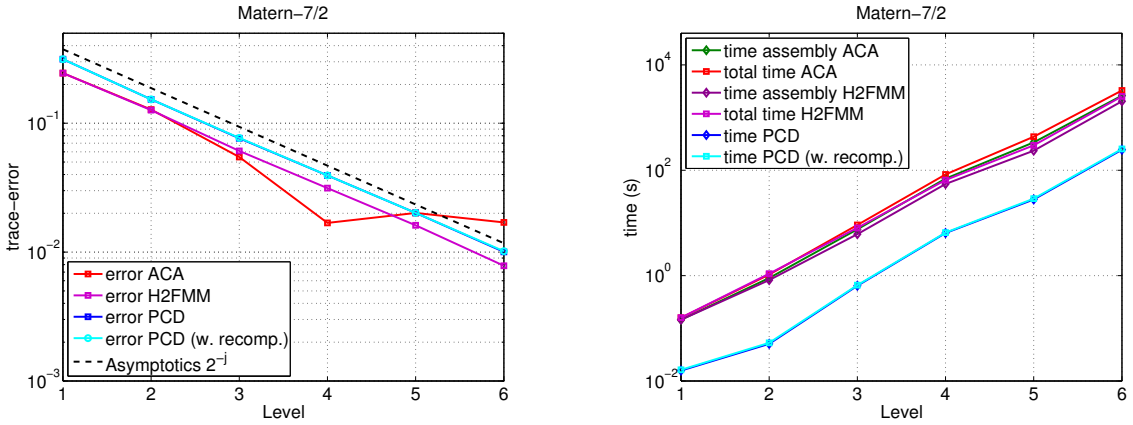

Figure III.14: Trace error (left) and computational times (right) for the $\nu = 7/2$ on the plate geometry.

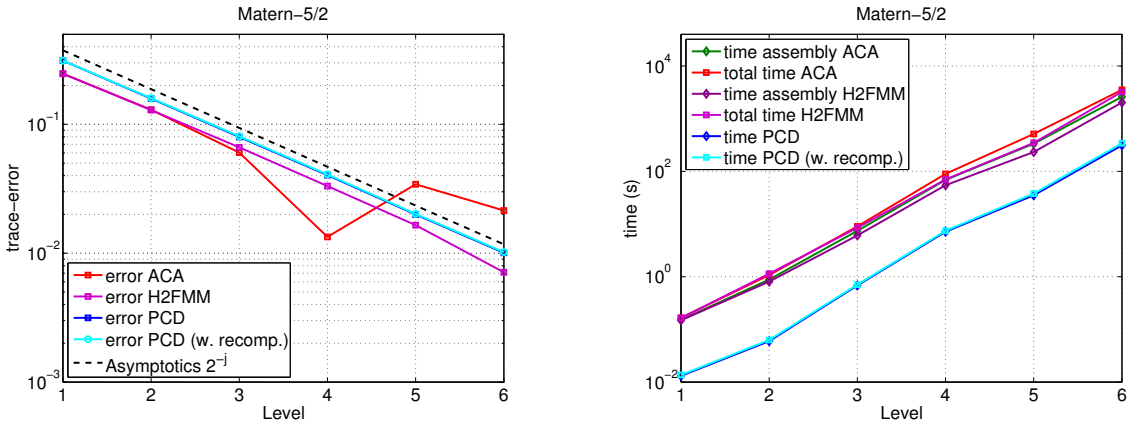

Figure III.15: Trace error (left) and computational times (right) for the $\nu = 5/2$ on the plate geometry.

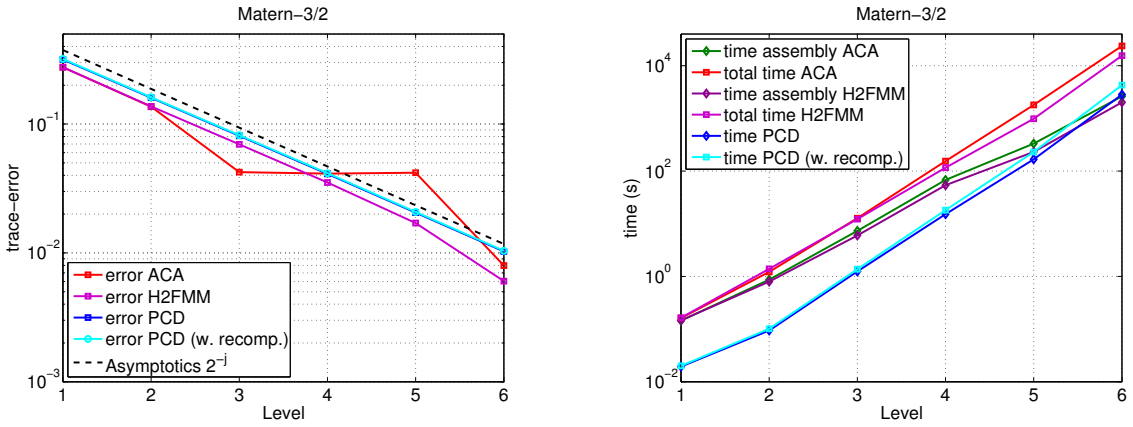

Figure III.16: Trace error (left) and computational times (right) for the $\nu = 3/2$ on the plate geometry.

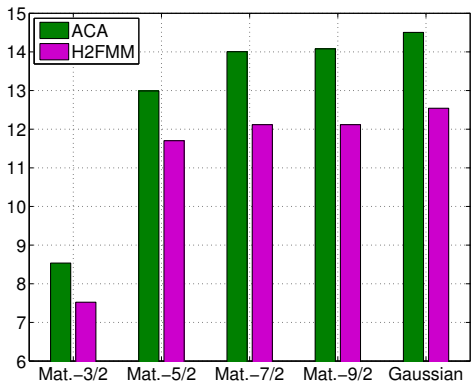

Figure III.17: Computational times relative to PCD on the plate geometry.

Chapter IV

THE DOMAIN MAPPING METHOD

In this chapter, we introduce the domain mapping method for the numerical solution of elliptic diffusion problems on random domains. The randomness in the domain will be described in terms of a random vector field. We have already seen how to make a random vector field feasible for numerical simulations by the computation of the Karhunen-Loève expansion. As main result, we will show how to derive regularity results for the solution to the diffusion problem from the properties of the Karhunen-Loève expansion.

1. Problem formulation

In the following, let $D_{\text{ref}} \subset \mathbb{R}^d$ for $d \in \mathbb{N}$ (of special interest are the cases $d = 2, 3$) be a domain with Lipschitz continuous boundary ∂D_{ref}. Let $\mathbf{V}\colon \Omega \times \overline{D_{\text{ref}}} \to \mathbb{R}^d$ be a C^2-diffeomorphism, i.e., for almost every $\omega \in \Omega$, $\mathbf{V}$ is twice continuously differentiable with respect to the spatial variable $\mathbf{x}$ and has a twice continuously differentiable inverse. Moreover, we impose the uniformity condition $\|\mathbf{V}(\omega)\|_{C^2(\overline{D_{\text{ref}}};\mathbb{R}^d)}, \|\mathbf{V}^{-1}(\omega)\|_{C^2(\overline{D_{\text{ref}}};\mathbb{R}^d)} \leqslant c$ for some $c \in (0, \infty)$ and almost every $\omega \in \Omega$.[1] Thus, $\mathbf{V}$ defines a family of random domains

$$D(\omega) := \mathbf{V}(\omega, D_{\text{ref}}).$$

For the subsequent analysis, we restrict ourselves to the case of the Poisson equation, i.e.

$$-\Delta u(\omega, \mathbf{x}) = f(\mathbf{x}) \text{ in } D(\omega), \quad u(\omega, \mathbf{x}) = 0 \text{ on } \Gamma(\omega) := \partial D(\omega). \tag{1.1}$$

This considerably simplifies the analysis and the extension to non-constant diffusion coefficients is straightforward, cf. Remark (3.22). In order to guarantee solvability for almost every $\omega \in \Omega$, we consider the right hand side to be defined on the *hold-all* domain

$$\mathcal{D} := \bigcup_{\omega \in \Omega} D(\omega). \tag{1.2}$$

From the uniformity condition, we infer for almost every $\omega \in \Omega$ and every $\mathbf{x} \in D$ that the singular-values of the vector field $\mathbf{V}$'s Jacobian $\mathbf{J}(\omega, \mathbf{x})$ satisfy

$$0 < \underline{\sigma} \leqslant \min\left\{\sigma\big(\mathbf{J}(\omega, \mathbf{x})\big)\right\} \leqslant \max\left\{\sigma\big(\mathbf{J}(\omega, \mathbf{x})\big)\right\} \leqslant \overline{\sigma} < \infty. \tag{1.3}$$

[1] Regard that for the analysis it is sufficient to assume that $\mathbf{V}$ is a C^1-diffeomorphism and satisfies the uniformity in $C^1(\overline{D_{\text{ref}}};\mathbb{R}^d)$. Nevertheless, in order to obtain H^2-regularity of the model problem, we make this stronger assumption.

In particular, we assume without loss of generality that $\underline{\sigma} \leqslant 1$ and $\overline{\sigma} \geqslant 1$.

2. Reformulation on the reference domain

In the sequel, we consider the spaces $H_0^1(D(\omega))$ and $H_0^1(D_{\mathrm{ref}})$ to be equipped with the norms

$$\|v\|_{H^1(D(\omega))} := \|\nabla v\|_{L^2(D(\omega);\mathbb{R}^d)} \text{ and } \|v\|_{H^1(D_{\mathrm{ref}})} := \|\nabla v\|_{L^2(D_{\mathrm{ref}};\mathbb{R}^d)},$$

respectively. We assume that the related dual spaces $H^{-1}(D(\omega))$ and $H^{-1}(D_{\mathrm{ref}})$ are also defined with respect to these norms, i.e

$$\|v\|_{H^{-1}(D(\omega))} := \sup_{0\neq u\in H_0^1(D(\omega))} \frac{(u,v)_{L^2(D(\omega))}}{\|\nabla u\|_{L^2(D(\omega);\mathbb{R}^d)}}$$

and

$$\|v\|_{H^{-1}(D_{\mathrm{ref}})} := \sup_{0\neq u\in H_0^1(D_{\mathrm{ref}})} \frac{(u,v)_{L^2(D_{\mathrm{ref}})}}{\|\nabla u\|_{L^2(D_{\mathrm{ref}};\mathbb{R}^d)}}.$$

The main tool we use in the convergence analysis for the model problem (1.1) is the one-to-one correspondence between the problem which is pulled back to the reference domain D_{ref} and the actual realization given by $D(\omega)$. The equivalence between those two problems is described by the vector field $\mathbf{V}(\omega,\mathbf{x})$. For an arbitrary function v on $D(\omega)$, we denote the transported function by

$$\hat{v}(\omega,\mathbf{x}) := (v\circ\mathbf{V})(\omega,\mathbf{x}).$$

According to the chain rule, we have for $v\in C^1(D(\omega))$

$$\nabla v(\mathbf{V}(\omega,\mathbf{x})) = \mathbf{J}(\omega,\mathbf{x})^{-\intercal}\nabla\hat{v}(\omega,\mathbf{x}). \tag{2.1}$$

For given $\omega\in\Omega$, the variational formulation for the model problem (1.1) is given as follows:

$$\begin{aligned}&\text{Find } u(\omega)\in H_0^1(D(\omega)) \text{ such that}\\ &\int_{D(\omega)}\langle\nabla u(\omega),\nabla v\rangle\,\mathrm{d}\mathbf{x} = \int_{D(\omega)} f v\,\mathrm{d}\mathbf{x} \quad \text{for all } v\in H_0^1(D(\omega)).\end{aligned} \tag{2.2}$$

Thus, with

$$\mathbf{A}(\omega,\mathbf{x}) := \big(\mathbf{J}(\omega,\mathbf{x})^{\intercal}\mathbf{J}(\omega,\mathbf{x})\big)^{-1}\det\mathbf{J}(\omega,\mathbf{x}) \tag{2.3}$$

and

$$f_{\mathrm{ref}}(\omega,\mathbf{x}) := \hat{f}(\omega,\mathbf{x})\det\mathbf{J}(\omega,\mathbf{x}), \tag{2.4}$$

we obtain the following variational formulation with respect to the reference domain:

$$\begin{aligned}&\text{Find } \hat{u}(\omega)\in H_0^1(D_{\mathrm{ref}}) \text{ such that}\\ &\int_{D_{\mathrm{ref}}}\langle\mathbf{A}(\omega)\nabla_{\mathbf{x}}\hat{u}(\omega),\nabla_{\mathbf{x}}\hat{v}(\omega)\rangle\,\mathrm{d}\mathbf{x} = \int_{D_{\mathrm{ref}}} f_{\mathrm{ref}}(\omega)\hat{v}(\omega)\,\mathrm{d}\mathbf{x} \quad \text{for all } \hat{v}(\omega)\in H_0^1(D_{\mathrm{ref}}).\end{aligned} \tag{2.5}$$

Remind that $\langle\cdot,\cdot\rangle$ denotes the canonical inner product for $\mathbb{R}^d$.

(2.6) **Remark.** Since $\mathbf{V}$ is assumed to be a C^2-diffeomorphism, we have for almost every $\omega \in \Omega$ that

$$\mathbf{V}^{-1} \circ \mathbf{V} = \mathrm{Id} \quad \Rightarrow \quad \mathbf{J}^{-1}\mathbf{J} = \mathbf{I} \quad \Rightarrow \quad \det \mathbf{J}^{-1} \det \mathbf{J} = 1 \quad \text{for all } \mathbf{x}.$$

Herein, $\mathbf{I} \in \mathbb{R}^{d\times d}$ denotes the identity matrix. Especially, we infer $\det \mathbf{J}^{-1}, \det \mathbf{J} \neq 0$. The continuity of $\mathbf{J}, \mathbf{J}^{-1}$ and of the determinant imply now that either $\det \mathbf{J}^{-1}, \det \mathbf{J} > 0$ or $\det \mathbf{J}^{-1}, \det \mathbf{J} < 0$ for all $\mathbf{x}$. Therefore, without loss of generality, we will assume the positiveness of the determinants.

For given $v \in H_0^1(D(\omega))$, equation (2.5) contains the related transported test function $\hat{v}(\omega)$. This is inconvenient for our analysis. Fortunately, we can show that the test functions may be assumed to be independent of the random parameter $\omega \in \Omega$:

(2.7) **Lemma.** The spaces $H_0^1(D_{\mathrm{ref}})$ and $H_0^1(D(\omega))$ are isomorphic by the isomorphism

$$\mathcal{E}\colon H_0^1(D_{\mathrm{ref}}) \to H_0^1(D(\omega)), \quad v \mapsto v \circ \mathbf{V}(\omega)^{-1}.$$

The inverse mapping is given by

$$\mathcal{E}^{-1}\colon H_0^1(D(\omega)) \to H_0^1(D_{\mathrm{ref}}), \quad v \mapsto v \circ \mathbf{V}(\omega).$$

Proof. The proof of this lemma is a consequence of the chain rule (2.1) and the ellipticity assumption (1.3). □

This lemma implies that the space of test functions is not dependent on $\omega \in \Omega$: Obviously, we have $H_0^1(D(\omega)) = \{\mathcal{E}(v) : v \in H_0^1(D_{\mathrm{ref}})\}$. Thus, for an arbitrary function $\mathcal{E}(v) \in H_0^1(D(\omega))$, it holds that

$$\widehat{\mathcal{E}(v)} = \mathcal{E}(v) \circ \mathbf{V} = v \circ \mathbf{V}^{-1} \circ \mathbf{V} = v \in H_0^1(D_{\mathrm{ref}})$$

independent of $\omega \in \Omega$. In particular, the solutions u to (2.2) and $\hat{u}$ to (2.5) satisfy

(2.8) $$\hat{u}(\omega) = u \circ \mathbf{V}(\omega) \quad \text{and} \quad u(\omega) = \hat{u} \circ \mathbf{V}(\omega)^{-1}.$$

In complete analogy, the one-to-one correspondence between the Galerkin solution $u_j(\omega) \in V_j^s(\omega)$ to (2.2) and the Galerkin solution $\hat{u}_j(\omega) \in V_j^s$ to (2.5) is given by the following

(2.9) **Theorem.** Let $u_j(\omega) \in V_j^s(\omega)$ be the Galerkin solution to the variational formulation (2.2) and $\hat{u}_j(\omega) \in V_j^s$ the Galerkin solution to (2.5), respectively. Then, it holds

$$\hat{u}_j(\omega) = u_j \circ \mathbf{V}(\omega) \quad \text{and} \quad u_j(\omega) = \hat{u}_j \circ \mathbf{V}(\omega)^{-1}.$$

Proof. The proof is a straightforward consequence of the construction of the finite element spaces $V_j^s(\omega)$, cf. (II.6.10), and the equivalence of the problems (2.2) and (2.5), see also (2.8). □

3. Regularity of the solution

In the sequel, we assume that the vector field $\mathbf{V}(\omega, \mathbf{x})$ is given by a finite rank Karhunen-Loève expansion, otherwise it has to be truncated appropriately. Moreover, we replace the random variables by their coordinates $y_i = X_i(\omega)$. Then, we have with

$$\mathbf{y} \in \square := [-1,1]^M, \quad \text{where } \mathbf{y} = [y_1, \dots, y_M],$$

the representation[2]

$$\mathbf{V}(\mathbf{x}, \mathbf{y}) = \mathbb{E}[\mathbf{V}](\mathbf{x}) + \sum_{i=1}^{M} \sigma_i \boldsymbol{\varphi}_i(\mathbf{x}) y_i.$$

Nevertheless, we provide in this section estimates which are independent of $M \in \mathbb{N}$. Thus, we explicitly allow M to become arbitrarily large.

Notice that, due to the independence of the random variables, the related push-forward measure $\mathbb{P}_{\mathbf{X}} := \mathbb{P} \circ \mathbf{X}^{-1}$ where $\mathbf{X}(\omega) := [X_1(\omega), \dots X_M(\omega)]$ is of product structure. Furthermore, we always think of the spaces $L^p(\square)$ for $p \in [1, \infty]$ to be equipped with the measure $\mathbb{P}_{\mathbf{X}}$. Moreover, we set $\boldsymbol{\gamma} = [\gamma_k]_{k=1}^M$, cf. Assumption (II.3.3).

Without loss of generality, we may assume that $\mathbb{E}[\mathbf{V}](\mathbf{x}) = \mathbf{x}$ is the identity mapping. Otherwise, we replace D_{ref} by

$$\widetilde{D}_{\mathrm{ref}} := \mathbb{E}[\mathbf{V}](D_{\mathrm{ref}}) \quad \text{and} \quad \widetilde{\boldsymbol{\varphi}}_k := \sqrt{\det(\mathbb{E}[\mathbf{V}]^{-1})'}\, \boldsymbol{\varphi}_k \circ \mathbb{E}[\mathbf{V}]^{-1}.$$

Therefore, we obtain

$$\mathbf{V}(\mathbf{x}, \mathbf{y}) = \mathbf{x} + \sum_{k=1}^{M} \sigma_i \boldsymbol{\varphi}_i(\mathbf{x}) y_i \quad \text{and} \quad \mathbf{J}(\mathbf{x}, \mathbf{y}) = \mathbf{I} + \sum_{i=1}^{M} \sigma_i \boldsymbol{\varphi}_i'(\mathbf{x}) y_i. \tag{3.1}$$

Here, we mean with $\boldsymbol{\varphi}_i'$ the Jacobian of $\boldsymbol{\varphi}_i$.

In order to show regularity results, we shall also refer to the following Lebesgue-Bochner spaces. We define the space $L^\infty(\square; L^\infty(D_{\mathrm{ref}}; \mathbb{R}^d))$ as the set of all functions $\mathbf{V} \colon \square \to L^\infty(D_{\mathrm{ref}}; \mathbb{R}^d)$ with finite norm

$$|||\mathbf{V}|||_d := \operatorname*{ess\,sup}_{\mathbf{y} \in \square} \|\mathbf{V}(\mathbf{y})\|_{L^\infty(D_{\mathrm{ref}}; \mathbb{R}^d)}.$$

Furthermore, $L^\infty(\square; L^\infty(D_{\mathrm{ref}}; \mathbb{R}^{d \times d}))$ consists of all functions $\mathbf{M} \colon \square \to L^\infty(D_{\mathrm{ref}}; \mathbb{R}^{d \times d})$ with finite norm

$$|||\mathbf{M}|||_{d \times d} := \operatorname*{ess\,sup}_{\mathbf{y} \in \square} \|\mathbf{M}(\mathbf{y})\|_{L^\infty(D_{\mathrm{ref}}; \mathbb{R}^{d \times d})}.$$

The consistency (II.3.6) of the norms in $L^\infty(D_{\mathrm{ref}}; \mathbb{R}^d)$ and $L^\infty(D_{\mathrm{ref}}; \mathbb{R}^{d \times d})$ transfers to the norms $|||\cdot|||_d$ and $|||\cdot|||_{d \times d}$.

We start by providing bounds on the derivatives of $\big(\mathbf{J}(\mathbf{x}, \mathbf{y})^{\intercal} \mathbf{J}(\mathbf{x}, \mathbf{y})\big)^{-1}$.

[2] For notational convenience, we will write from now on $\mathbf{V}(\mathbf{x}, \mathbf{y})$ instead of $\mathbf{V}(\mathbf{y}, \mathbf{x})$, $\mathbf{J}(\mathbf{x}, \mathbf{y})$ instead of $\mathbf{J}(\mathbf{y}, \mathbf{x})$, etc.

(3.2) **Lemma.** Let $\mathbf{J}\colon D_{\mathrm{ref}} \times \square \to \mathbb{R}^{d\times d}$ be defined as in (3.1). Then, it holds for the derivatives of $\big(\mathbf{J}(\mathbf{x},\mathbf{y})^\intercal \mathbf{J}(\mathbf{x},\mathbf{y})\big)^{-1}$ under the conditions of Assumption (II.3.3) that

$$\big|\!\big|\!\big|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} (\mathbf{J}^\intercal \mathbf{J})^{-1}\big|\!\big|\!\big|_{d\times d} \leqslant |\boldsymbol{\alpha}|! \frac{\boldsymbol{\gamma}^{\boldsymbol{\alpha}}}{\underline{\sigma}^2} \left(\frac{2(1+c_\gamma)}{\underline{\sigma}^2 \log 2}\right)^{|\boldsymbol{\alpha}|}.$$

Proof. We define $\mathbf{B}(\mathbf{x},\mathbf{y}) := \mathbf{J}(\mathbf{x},\mathbf{y})^\intercal \mathbf{J}(\mathbf{x},\mathbf{y})$ and $\tilde{\mathbf{A}}(\mathbf{x},\mathbf{y}) := \big(\mathbf{B}(\mathbf{x},\mathbf{y})\big)^{-1}$. Expanding the expression for $\mathbf{B}(\mathbf{x},\mathbf{y})$ yields

$$\mathbf{B}(\mathbf{x},\mathbf{y}) = \mathbf{I} + \sum_{k=1}^{M} \sigma_k \big(\boldsymbol{\varphi}_k'(\mathbf{x}) + \boldsymbol{\varphi}_k'(\mathbf{x})^\intercal\big) y_k + \sum_{k,k'=1}^{M} \sigma_k \sigma_{k'} \boldsymbol{\varphi}_k'(\mathbf{x})^\intercal \boldsymbol{\varphi}_{k'}'(\mathbf{x}) y_k y_{k'}.$$

Thus, the first order derivatives of $\mathbf{B}(\mathbf{x},\mathbf{y})$ are given by

$$\partial_{y_i} \mathbf{B}(\mathbf{x},\mathbf{y}) = \sigma_i \big(\boldsymbol{\varphi}_i'(\mathbf{x}) + \boldsymbol{\varphi}_i'(\mathbf{x})^\intercal\big) + \sum_{k=1}^{M} \sigma_i \sigma_k \big(\boldsymbol{\varphi}_i'(\mathbf{x})^\intercal \boldsymbol{\varphi}_k'(\mathbf{x}) + \boldsymbol{\varphi}_k'(\mathbf{x})^\intercal \boldsymbol{\varphi}_i'(\mathbf{x})\big) y_k \tag{3.3}$$

and the second order derivatives are given by

$$\partial_{y_i} \partial_{y_j} \mathbf{B}(\mathbf{x},\mathbf{y}) = \sigma_i \sigma_j \big(\boldsymbol{\varphi}_i'(\mathbf{x})^\intercal \boldsymbol{\varphi}_j'(\mathbf{x}) + \boldsymbol{\varphi}_j'(\mathbf{x})^\intercal \boldsymbol{\varphi}_i'(\mathbf{x})\big). \tag{3.4}$$

Obviously, all higher order derivatives with respect to $\mathbf{y}$ vanish.

The ellipticity assumption (1.3) now leads to the following bounds:

$$\underline{\sigma}^2 \leqslant |\!|\!|\mathbf{B}|\!|\!|_{d\times d} \leqslant \overline{\sigma}^2 \quad \text{and} \quad \frac{1}{\overline{\sigma}^2} \leqslant |\!|\!|\tilde{\mathbf{A}}|\!|\!|_{d\times d} \leqslant \frac{1}{\underline{\sigma}^2},$$

respectively. Furthermore, we derive from (3.3) that

$$|\!|\!|\partial_{y_i} \mathbf{B}|\!|\!|_{d\times d} \leqslant 2\gamma_i + 2\gamma_i \sum_{k=1}^{M} \gamma_k \leqslant 2(1+c_\gamma)\gamma_i$$

and from (3.4) that $|\!|\!|\partial_{y_j}\partial_{y_i} \mathbf{B}|\!|\!|_{d\times d} \leqslant 2\gamma_i\gamma_j$. Thus, we have

$$|\!|\!|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \mathbf{B}|\!|\!|_{d\times d} \leqslant \begin{cases} 2(1+c_\gamma)\boldsymbol{\gamma}^{\boldsymbol{\alpha}}, & \text{if } |\boldsymbol{\alpha}| = 1,2, \\ 0, & \text{if } |\boldsymbol{\alpha}| > 2. \end{cases} \tag{3.5}$$

Since $\tilde{\mathbf{A}} = v \circ \mathbf{B}$ is a composite function with $v(x) = x^{-1}$, we may employ *Faà di Bruno's formula*, cf. [CS96], which is a generalization of the chain rule, to compute its derivatives. For $n = |\boldsymbol{\alpha}|$ Faà di Bruno's formula formally yields[3]

$$\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \tilde{\mathbf{A}}(\mathbf{x},\mathbf{y}) = \sum_{r=1}^{n} (-1)^r r! \tilde{\mathbf{A}}(\mathbf{x},\mathbf{y})^{r+1} \sum_{P(\boldsymbol{\alpha},r)} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{\big(\partial_{\mathbf{y}}^{\boldsymbol{\beta}_j} \mathbf{B}(\mathbf{x},\mathbf{y})\big)^{k_j}}{k_j! (\boldsymbol{\beta}_j!)^{k_j}}. \tag{3.6}$$

[3]With "formally" we mean that we ignore here the fact that the product of matrices is in general not Abelian. Nevertheless, a differentiation yields exactly the appearing products in a permuted order. The formal representation is justified since we only consider the norm of the representation in the sequel.

Here, the set $P(\boldsymbol{\alpha}, r)$ contains restricted integer partitions of the multiindex $\boldsymbol{\alpha}$ into r non-vanishing multiindices, cf. (A.1.2).

Taking the norm in (3.6), we arrive at the estimate

$$\begin{aligned}
\|\!|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \tilde{\mathbf{A}}|\!\|_{d\times d} &\leqslant \sum_{r=1}^{n} r! \|\!|\tilde{\mathbf{A}}|\!\|_{d\times d}^{r+1} \sum_{P(\boldsymbol{\alpha},r)} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{\|\!|\partial_{\mathbf{y}}^{\boldsymbol{\beta}_j} \mathbf{B}|\!\|_{d\times d}^{k_j}}{k_j!(\boldsymbol{\beta}_j!)^{k_j}} \\
&\leqslant \sum_{r=1}^{n} r! \Big(\frac{1}{\underline{\sigma}^2}\Big)^{r+1} \sum_{P(\boldsymbol{\alpha},r)} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{\big(2(1+c_\gamma)\boldsymbol{\gamma}^{\boldsymbol{\beta}_j}\big)^{k_j}}{k_j!(\boldsymbol{\beta}_j!)^{k_j}} \\
&= \boldsymbol{\gamma}^{\boldsymbol{\alpha}} \sum_{r=1}^{n} r! \Big(\frac{1}{\underline{\sigma}^2}\Big)^{r+1} \big(2(1+c_\gamma)\big)^r \sum_{P(\boldsymbol{\alpha},r)} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{1}{k_j!(\boldsymbol{\beta}_j!)^{k_j}}.
\end{aligned}$$

We know from [CS96] that

$$\sum_{P(\boldsymbol{\alpha},r)} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{1}{k_j!(\boldsymbol{\beta}_j!)^{k_j}} = S_{n,r},$$

where $S_{n,r}$ are the Stirling numbers of the second kind, cf. [AS64]. Thus, we obtain

$$\|\!|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \tilde{\mathbf{A}}|\!\|_{d\times d} \leqslant \frac{\boldsymbol{\gamma}^{\boldsymbol{\alpha}}}{\underline{\sigma}^2} \sum_{r=1}^{n} r! \Big(\frac{2(1+c_\gamma)}{\underline{\sigma}^2}\Big)^r S_{n,r} \leqslant \frac{\boldsymbol{\gamma}^{\boldsymbol{\alpha}}}{\underline{\sigma}^2} \Big(\frac{2(1+c_\gamma)}{\underline{\sigma}^2}\Big)^{|\boldsymbol{\alpha}|} \sum_{r=1}^{n} r! S_{n,r}.$$

The term $\tilde{b}(n) := \sum_{r=0}^{n} r! S_{n,r}$ coincides with the *n-th ordered Bell number*. The ordered Bell numbers satisfy the recurrence relation

$$\tilde{b}(n) = \sum_{r=0}^{n-1} \binom{n}{k} \tilde{b}(r) \quad \text{with } \tilde{b}(0) = 1, \tag{3.7}$$

see [Gro62], and may be estimated as follows[4], cf. [BTNT12],

$$\tilde{b}(n) \leqslant \frac{n!}{(\log 2)^n}. \tag{3.8}$$

This finally proves the assertion. □

The next lemma bounds the derivatives of $\det \mathbf{J}(\mathbf{x}, \mathbf{y})$.

(3.9) **Lemma.** Let $\mathbf{J} \colon \square \to L^\infty(D_{\mathrm{ref}}; \mathbb{R}^{d\times d})$ be defined as in (3.1) and let $\{\gamma_i\}_i \in \ell^p(\mathcal{I})$ for $p < 1/2$. Then, it holds for the derivatives of $\det \mathbf{J}(\mathbf{x}, \mathbf{y})$, under the conditions of Assumption (II.3.3), that

$$\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \det \mathbf{J}\|_{L^\infty(\square; L^\infty(D_{\mathrm{ref}}))} \leqslant C_{\det} |\boldsymbol{\alpha}|! \overline{\sigma}^d \Big(\frac{4}{\underline{\sigma}}\Big)^{|\boldsymbol{\alpha}|} \tilde{\boldsymbol{\gamma}}^{\boldsymbol{\alpha}}$$

with the modified sequence $\tilde{\gamma}_i = \gamma_i i^{1+\varepsilon}/c_\varepsilon$ for arbitrary $\varepsilon > 0$ with a normalization constant $c_\varepsilon > 0$ and a constant $C_{\det}$ depending on the modified sequence and the dimension d.

[4] A more rigorous bound on the ordered Bell numbers is provided by [Wil06]. There, it is shown that

$$\tilde{b}(n) = \frac{n!}{2(\log 2)^{n+1}} + \mathcal{O}\big((0.16)^n n!\big).$$

Nevertheless, for our purposes, the bound from [BTNT12] is sufficient.

Proof. We start from the identity

$$\det \exp(\mathbf{M}) = \exp(\operatorname{trace} \mathbf{M}), \tag{3.10}$$

which holds for any matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$, cf. [HJ94]. From this, we derive

$$\det \mathbf{M} = \exp(\operatorname{trace} \log \mathbf{M}), \tag{3.11}$$

where the matrix logarithm exists whenever $\mathbf{M}$ is non-singular, cf. [HJ94].

Faà di Bruno's formula yields formally with

$$\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \operatorname{trace} \log \mathbf{J}(\mathbf{x},\mathbf{y}) = \operatorname{trace} \partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \log \mathbf{J}(\mathbf{x},\mathbf{y})$$

that

$$\begin{aligned}&\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \operatorname{trace} \log \mathbf{J}(\mathbf{x},\mathbf{y})\\ &\quad= \operatorname{trace}\Bigg(\sum_{r=1}^{n}(-1)^{r-1}(r-1)!\mathbf{J}(\mathbf{x},\mathbf{y})^{-r}\sum_{P(\boldsymbol{\alpha},r)}\boldsymbol{\alpha}!\prod_{j=1}^{n}\frac{\left(\partial_{\mathbf{y}}^{\boldsymbol{\beta}_j}\mathbf{J}(\mathbf{x},\mathbf{y})\right)^{k_j}}{k_j!(\boldsymbol{\beta}_j!)^{k_j}}\Bigg).\end{aligned}$$

Taking into account that $|\operatorname{trace}\mathbf{M}| \leqslant d\max\{\sigma(\mathbf{M})\}$, we obtain

$$\begin{aligned}&\left\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \operatorname{trace} \log \mathbf{J}\right\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))}\\ &\quad\leqslant d\left\|\sum_{r=1}^{n}(-1)^{r-1}(r-1)!\mathbf{J}^{-r}\sum_{P(\boldsymbol{\alpha},r)}\boldsymbol{\alpha}!\prod_{j=1}^{n}\frac{\left(\partial_{\mathbf{y}}^{\boldsymbol{\beta}_j}\mathbf{J}\right)^{k_j}}{k_j!(\boldsymbol{\beta}_j!)^{k_j}}\right\|_{d\times d}.\end{aligned}$$

Furthermore, we may estimate

$$\begin{aligned}&\left\|\sum_{r=1}^{n}(-1)^{r-1}(r-1)!\mathbf{J}^{-r}\sum_{P(\boldsymbol{\alpha},r)}\boldsymbol{\alpha}!\prod_{j=1}^{n}\frac{\left(\partial_{\mathbf{y}}^{\boldsymbol{\beta}_j}\mathbf{J}\right)^{k_j}}{k_j!(\boldsymbol{\beta}_j!)^{k_j}}\right\|_{d\times d}\\ &\quad\leqslant\sum_{r=1}^{n}(r-1)!\left\|\mathbf{J}^{-1}\right\|_{d\times d}^{r}\sum_{P(\boldsymbol{\alpha},r)}\boldsymbol{\alpha}!\prod_{j=1}^{n}\frac{\left\|\partial_{\mathbf{y}}^{\boldsymbol{\beta}_j}\mathbf{J}\right\|_{d\times d}^{k_j}}{k_j!(\boldsymbol{\beta}_j!)^{k_j}} \leqslant (|\boldsymbol{\alpha}|-1)!\left(\frac{1}{\underline{\sigma}}\right)^{|\boldsymbol{\alpha}|}\boldsymbol{\gamma}^{\boldsymbol{\alpha}}.\end{aligned}$$

The last inequality holds due to the fact that the derivatives of the Jacobian $\mathbf{J}(\mathbf{x},\mathbf{y})$ with respect to y_i satisfy $\left\|\partial_{y_i}\mathbf{J}\right\|_{d\times d} \leqslant \gamma_i$ and vanish for $|\boldsymbol{\beta}_j| > 1$. Thus, the related summand does not vanish only if $\boldsymbol{\beta}_j = \mathbf{0}$ or $\boldsymbol{\beta}_j = \mathbf{e}_i$, where $\mathbf{e}_i$ is the i-th unit vector. Due to the definition of $P(\boldsymbol{\alpha},r)$, this choice of $\boldsymbol{\beta}_j$ is only possible if $r = |\boldsymbol{\alpha}|$. Therefore, we arrive at

$$\left\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \operatorname{trace} \log \mathbf{J}\right\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \leqslant d|\boldsymbol{\alpha}|!\left(\frac{1}{\underline{\sigma}}\right)^{|\boldsymbol{\alpha}|}\boldsymbol{\gamma}^{\boldsymbol{\alpha}}. \tag{3.12}$$

By spending a convergent series, i.e. $\{c_\varepsilon/i^{1+\varepsilon}\}_i$, with normalization constant c_ε, we have by Lemma (A.1.4)

$$\left\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \operatorname{trace} \log \mathbf{J}\right\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \leqslant d\tilde{c}\boldsymbol{\alpha}!\left(\frac{1}{\underline{\sigma}}\right)^{|\boldsymbol{\alpha}|}\tilde{\boldsymbol{\gamma}}^{\boldsymbol{\alpha}} \tag{3.13}$$

with $\tilde{c} = 1/(1 - c_\varepsilon)$.

The combination of (3.11) and (3.13) provides

$$\begin{aligned}
\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \det \mathbf{J}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} &= \|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \exp(\operatorname{trace}\log \mathbf{J})\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \\
&= \Big\| \sum_{r=1}^{n} \exp(\operatorname{trace}\log \mathbf{J}) \sum_{P(\boldsymbol{\alpha},r)} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{(\partial_{\mathbf{y}}^{\boldsymbol{\beta}_j} \operatorname{trace}\log \mathbf{J})^{k_j}}{k_j!(\boldsymbol{\beta}_j!)^{k_j}} \Big\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \\
&\leqslant \|\det \mathbf{J}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \sum_{r=1}^{n} \sum_{P(\boldsymbol{\alpha},r)} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{\big(d\tilde{c}\boldsymbol{\beta}_j!(\frac{1}{\underline{\sigma}})^{|\boldsymbol{\beta}_j|}\tilde{\boldsymbol{\gamma}}^{\boldsymbol{\beta}_j}\big)^{k_j}}{k_j!(\boldsymbol{\beta}_j!)^{k_j}} \\
&= \|\det \mathbf{J}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \Big(\frac{1}{\underline{\sigma}}\Big)^{|\boldsymbol{\alpha}|} \tilde{\boldsymbol{\gamma}}^{\boldsymbol{\alpha}} \sum_{r=1}^{n} (d\tilde{c})^r \sum_{P(\boldsymbol{\alpha},r)} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{1}{k_j!}.
\end{aligned}$$

Now, the application of Lemma (A.1.6) gives us

$$\begin{aligned}
&\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \det \mathbf{J}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \\
&\quad\leqslant \|\det \mathbf{J}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \Big(\frac{1}{\underline{\sigma}}\Big)^{|\boldsymbol{\alpha}|} \tilde{\boldsymbol{\gamma}}^{\boldsymbol{\alpha}} \sum_{r=1}^{n} (d\tilde{c})^r \frac{|\boldsymbol{\alpha}|!}{r!} \binom{|\boldsymbol{\alpha}| + r - 1}{r - 1}.
\end{aligned}$$

It holds by the ellipticity assumption (1.3) that

$$\underline{\sigma}^d \leqslant \det \mathbf{J}(\mathbf{x},\mathbf{y}) \leqslant \overline{\sigma}^d \tag{3.14}$$

for almost every $\mathbf{y} \in \square$. Noticing in addition that

$$\sum_{r=1}^{n} (d\tilde{c})^r \frac{1}{r!} \binom{|\boldsymbol{\alpha}| + r}{r} \leqslant \sum_{r=1}^{n} \frac{(d\tilde{c})^r}{r!} \sum_{r=1}^{n} \binom{|\boldsymbol{\alpha}| + r}{r} \leqslant e^{d\tilde{c}} 2^{2|\boldsymbol{\alpha}|},$$

we end up with the assertion due to

$$\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \det \mathbf{J}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \leqslant |\boldsymbol{\alpha}|! \overline{\sigma}^d \Big(\frac{4}{\underline{\sigma}}\Big)^{|\boldsymbol{\alpha}|} e^{d\tilde{c}} \tilde{\boldsymbol{\gamma}}^{\boldsymbol{\alpha}}. \qquad \square$$

The application of the Leibniz rule now results in a regularity estimate for the diffusion matrix $\mathbf{A}(\mathbf{x},\mathbf{y})$.

(3.15) Theorem. The derivatives of the diffusion matrix $\mathbf{A}(\mathbf{x},\mathbf{y})$ defined in (2.3) satisfy

$$\|\!|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \mathbf{A}|\!\|_{d\times d} \leqslant C_{\det}(|\boldsymbol{\alpha}| + 1)! \frac{\overline{\sigma}^d}{\underline{\sigma}^2} \Big(\frac{4(1 + c_\gamma)}{\underline{\sigma}^2 \log 2}\Big)^{|\boldsymbol{\alpha}|} \tilde{\boldsymbol{\gamma}}^{\boldsymbol{\alpha}},$$

where $\tilde{\boldsymbol{\gamma}}$ is the modified sequence from the previous Lemma.

Proof. The Leibniz rule for $\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \mathbf{A}(\mathbf{x},\mathbf{y})$ reads as

$$\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \mathbf{A}(\mathbf{x},\mathbf{y}) = \sum_{\boldsymbol{\alpha}' \leqslant \boldsymbol{\alpha}} \binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'} \partial_{\mathbf{y}}^{\boldsymbol{\alpha}'} (\mathbf{J}(\mathbf{x},\mathbf{y})^\intercal \mathbf{J}(\mathbf{x},\mathbf{y}))^{-1} \partial_{\mathbf{y}}^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'} \det \mathbf{J}(\mathbf{x},\mathbf{y}).$$

Inserting the results of Lemma (3.2) and Lemma (3.9) yields

$$\begin{aligned}
&\left|\!\left|\!\left|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\mathbf{A}\right|\!\right|\!\right|_{d\times d}\\
&\quad\leqslant \sum_{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}|\boldsymbol{\alpha}'|!\frac{\boldsymbol{\gamma}^{\boldsymbol{\alpha}'}}{\underline{\sigma}^2}\left(\frac{2(1+c_\gamma)}{\underline{\sigma}^2\log 2}\right)^{|\boldsymbol{\alpha}'|}|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|!\overline{\sigma}^d C_{\det}\left(\frac{4}{\underline{\sigma}}\right)^{|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|}\tilde{\boldsymbol{\gamma}}^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'}\\
&\quad\leqslant \frac{\overline{\sigma}^d}{\underline{\sigma}^2}\left(\frac{4(1+c_\gamma)}{\underline{\sigma}^2\log 2}\right)^{|\boldsymbol{\alpha}|}\tilde{\boldsymbol{\gamma}}^{\boldsymbol{\alpha}}\sum_{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}|\boldsymbol{\alpha}'|!|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|!.
\end{aligned}$$

Now, we employ the combinatorial identity

$$\sum_{\substack{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}\\ |\boldsymbol{\alpha}'|=j}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}=\binom{|\boldsymbol{\alpha}|}{j} \tag{3.16}$$

and obtain

$$\begin{aligned}
\sum_{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}|\boldsymbol{\alpha}'|!|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|! &= \sum_{j=0}^{|\boldsymbol{\alpha}|} j!(|\boldsymbol{\alpha}|-j)!\sum_{\substack{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}\\ |\boldsymbol{\alpha}'|=j}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}\\
&= \sum_{j=0}^{|\boldsymbol{\alpha}|} j!(|\boldsymbol{\alpha}|-j)!\binom{|\boldsymbol{\alpha}|}{j} = |\boldsymbol{\alpha}|!\sum_{j=0}^{|\boldsymbol{\alpha}|}1 = (|\boldsymbol{\alpha}|+1)!.
\end{aligned}$$

□

In order to prove regularity results for the right hand side f_{ref} in (2.5), we have to assume that f is a smooth function.

(3.17) **Lemma.** Let $f\in C^\infty(\mathcal{D})$ be analytic, i.e. $\|\partial_{\mathbf{x}}^{\boldsymbol{\alpha}}f\|_{L^\infty(\mathcal{D};\mathbb{R}^d)}\leqslant\boldsymbol{\alpha}!\rho^{-|\boldsymbol{\alpha}|}c_f$ for all $\boldsymbol{\alpha}\in\mathbb{N}^d$ and some $\rho\in(0,1]$, and let Assumption (II.3.3) be satisfied. Then, the derivatives of $\hat{f}=f\circ\mathbf{V}$ are bounded by

$$\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{f}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))}\leqslant|\boldsymbol{\alpha}|!c_f\left(\frac{d}{\rho\log 2}\right)^{|\boldsymbol{\alpha}|}\boldsymbol{\gamma}^{\boldsymbol{\alpha}}.$$

Proof. In view of (3.1), differentiation of $\mathbf{V}(\mathbf{x},\mathbf{y})$ yields $\partial_{y_i}\mathbf{V}(\mathbf{x},\mathbf{y})=\sigma_i\boldsymbol{\varphi}_i(\mathbf{x})$. Thus, all higher order derivatives with respect to any direction y_j vanish, i.e. $\partial_{y_j}\partial_{y_i}\mathbf{V}(\mathbf{x},\mathbf{y})=0$. The norm of the first order derivatives is bounded by $\left|\!\left|\!\left|\partial_{y_i}\mathbf{V}\right|\!\right|\!\right|_d\leqslant\gamma_i$.

The rest of the proof is also based on the application of Faà di Bruno's formula. Nevertheless, we have this time to consider its multivariate version. To that end, we define the set $P(\boldsymbol{\alpha},\boldsymbol{\alpha}')$ as in (A.1.3). The application of the multivariate Faà di Bruno formula

induces

$$\begin{aligned}
&\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \hat{f}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \\
&\quad\leqslant \sum_{1\leqslant|\boldsymbol{\alpha}'|\leqslant n} \|\partial_{\mathbf{x}}^{\boldsymbol{\alpha}'} f\|_{L^\infty(\square;L^\infty(\mathcal{D}))} \sum_{P(\boldsymbol{\alpha},\boldsymbol{\alpha}')} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{\|(\partial_{\mathbf{y}}^{\boldsymbol{\beta}_j}\mathbf{V})^{\mathbf{k}_j}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))}}{\mathbf{k}_j!(\boldsymbol{\beta}_j!)^{|\mathbf{k}_j|}} \\
&\quad\leqslant \sum_{1\leqslant|\boldsymbol{\alpha}'|\leqslant n} \boldsymbol{\alpha}'!\rho^{-|\boldsymbol{\alpha}'|}c_f \sum_{P(\boldsymbol{\alpha},\boldsymbol{\alpha}')} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{(\boldsymbol{\gamma}^{\boldsymbol{\beta}_j})^{\mathbf{k}_j}}{\mathbf{k}_j!(\boldsymbol{\beta}_j!)^{|\mathbf{k}_j|}} \\
&\quad= c_f\boldsymbol{\gamma}^{\boldsymbol{\alpha}} \sum_{1\leqslant|\boldsymbol{\alpha}'|\leqslant n} \boldsymbol{\alpha}'!\rho^{-|\boldsymbol{\alpha}'|} \sum_{P(\boldsymbol{\alpha},\boldsymbol{\alpha}')} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{1}{\mathbf{k}_j!(\boldsymbol{\beta}_j!)^{|\mathbf{k}_j|}}.
\end{aligned}$$

From [CS96], we know that

$$\sum_{|\boldsymbol{\alpha}'|=r} \sum_{P(\boldsymbol{\alpha},\boldsymbol{\alpha}')} \boldsymbol{\alpha}! \prod_{j=1}^{n} \frac{1}{\mathbf{k}_j!(\boldsymbol{\beta}_j!)^{|\mathbf{k}_j|}} = d^r S_{n,r},$$

where again $S_{n,r}$ is the Stirling number of the second kind. Hence, we obtain

$$\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \hat{f}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \leqslant c_f\boldsymbol{\gamma}^{\boldsymbol{\alpha}} \sum_{r=1}^{n} \left(\frac{d}{\rho}\right)^r r!S_{n,r} \leqslant c_f\boldsymbol{\gamma}^{\boldsymbol{\alpha}} \left(\frac{d}{\rho}\right)^{|\boldsymbol{\alpha}|} \sum_{r=0}^{n} r!S_{n,r}.$$

Analogously to the proof of Lemma (3.2), we finally arrive at the assertion. □

Now, in complete analogy to Theorem (3.15), we have the following regularity result for the right hand side f_{ref}.

(3.18) **Theorem.** The derivatives of the right hand side $f_{\mathrm{ref}}(\mathbf{x},\mathbf{y})$ defined in (2.4) satisfy

$$\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} f_{\mathrm{ref}}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \leqslant (|\boldsymbol{\alpha}|+1)!c_f C_{\det}\overline{\sigma}^d \left(\frac{4d}{\underline{\sigma}\rho\log 2}\right)^{|\boldsymbol{\alpha}|} \tilde{\boldsymbol{\gamma}}^{\boldsymbol{\alpha}},$$

where $\tilde{\boldsymbol{\gamma}}$ is the modified sequence from Lemma (3.9).

Finally, we shall establish the dependency between the solution $\hat{u}$ to (2.5) and the datum f_{ref}.

(3.19) **Lemma.** Let $\hat{u}(\mathbf{y})$ be the solution to (2.5) and $f_{\mathrm{ref}} \in L^\infty(\square;L^\infty(D_{\mathrm{ref}}))$. Then, there holds

$$\|\hat{u}(\mathbf{y})\|_{H^1(D_{\mathrm{ref}})} \leqslant \frac{\overline{\sigma}^2}{\underline{\sigma}^d} c_D \|f_{\mathrm{ref}}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \tag{3.20}$$

for almost every $\mathbf{y} \in \square$ with a constant c_D only dependent on D_{ref}.

Proof. The bilinear form

$$(\mathbf{A}(\mathbf{y})\nabla\cdot,\nabla\,\cdot\,)_{L^2(D_{\mathrm{ref}};\mathbb{R}^d)} \colon H_0^1(D_{\mathrm{ref}}) \times H_0^1(D_{\mathrm{ref}}) \to \mathbb{R}$$

is coercive and bounded according to (1.3) and (3.14) for almost every $\mathbf{y} \in \square$. It holds

$$\frac{\underline{\sigma}^d}{\overline{\sigma}^2}\|\hat{u}(\mathbf{y})\|^2_{H^1(D_{\mathrm{ref}})} \leqslant (\mathbf{A}(\mathbf{y})\nabla\hat{u}(\mathbf{y}), \nabla\hat{u}(\mathbf{y}))_{\mathbf{L^2(D_{ref};\mathbb{R}^d)}}$$

and

$$(\mathbf{A}(\mathbf{y})\nabla\hat{u}(\mathbf{y}), \nabla\hat{v})_{L^2(D_{\mathrm{ref}};\mathbb{R}^d)} \leqslant \frac{\overline{\sigma}^d}{\underline{\sigma}^2}\|\hat{u}(\mathbf{y})\|_{H^1(D_{\mathrm{ref}})}\|\hat{v}\|_{H^1(D_{\mathrm{ref}})}$$

for all $\hat{u}(\mathbf{y}), \hat{v} \in H^1(D_{\mathrm{ref}})$ and almost every $\mathbf{y} \in \square$. The assertion follows therefore by the application of the Lax-Milgram lemma and the observation that

$$\|f_{\mathrm{ref}}\|_{L^\infty(\square;H^{-1}(D_{\mathrm{ref}}))} \leqslant \sqrt{|D_{\mathrm{ref}}|}c_P\|f_{\mathrm{ref}}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))},$$

where c_P denotes the Poincaré constant of D_{ref}. □

Incorporating the constants provided by Theorem (3.15) and Theorem (3.18) leads to the modified sequence

$$\{\mu_k\}_k := \left\{2C_{\mathrm{det}}\max\left(\frac{4d\overline{\sigma}^d}{\underline{\sigma}\rho\log 2}, \frac{4\overline{\sigma}^d(1+c_\gamma)}{\underline{\sigma}^4\log 2}\right)\tilde{\gamma_k}\right\}_k.$$

Notice that we ignore here the fact that the constant $C_{\mathrm{det}}\overline{\sigma}^d/\underline{\sigma}^2$ in the estimate for the diffusion matrix and the constant $C_{\mathrm{det}}\overline{\sigma}^d$ in the estimate for the right hand side do only occur with multiplicity 1. Moreover, we introduce the additional factor 2 in order to obtain the factor $|\boldsymbol{\alpha}|!$ in the derivatives instead of the factor $(|\boldsymbol{\alpha}|+1)!$. Nevertheless, for the sake of readability, we also insert them into the sequence $\{\mu_k\}_k$.

(3.21) **Theorem.** The derivatives of the solution u to (2.5) satisfy under the assumptions of Lemma (3.2) and (3.18) that

$$\|\partial^{\boldsymbol{\alpha}}_{\mathbf{y}}\hat{u}(\mathbf{y})\|_{H^1(D_{\mathrm{ref}})} \leqslant |\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}}\left(\frac{4\overline{\sigma}^2}{\underline{\sigma}^d}\max\{1, c_f c_D\}\right)^{|\boldsymbol{\alpha}|+1},$$

where c_D denotes the constant from the previous theorem.

Proof. Differentiating the variational formulation (2.5) with respect to $\mathbf{y}$ leads to

$$\left(\partial^{\boldsymbol{\alpha}}_{\mathbf{y}}(\mathbf{A}(\mathbf{y})\nabla_{\mathbf{x}}\hat{u}(\mathbf{y})), \nabla_{\mathbf{x}}\hat{v}\right)_{L^2(D_{\mathrm{ref}};\mathbb{R}^d)} = (\partial^{\boldsymbol{\alpha}}_{\mathbf{y}} f_{\mathrm{ref}}(\mathbf{y}), \hat{v})_{L^2(D_{\mathrm{ref}};\mathbb{R})}.$$

The isomorphism of the spaces $H^1_0(D_{\mathrm{ref}})$ and $H^1_0(D(\mathbf{y}))$ from Lemma (2.7) allows us to consider the test functions v to be independent of $\mathbf{y}$. Furthermore, the application of the Leibniz rule for the expression $\partial^{\boldsymbol{\alpha}}_{\mathbf{y}}(\mathbf{A}(\mathbf{y})\nabla_{\mathbf{x}}\hat{u}(\mathbf{y}))$ results in

$$\partial^{\boldsymbol{\alpha}}_{\mathbf{y}}(\mathbf{A}(\mathbf{y})\nabla_{\mathbf{x}}\hat{u}(\mathbf{y})) = \sum_{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}\partial^{\boldsymbol{\alpha}'}_{\mathbf{y}}\mathbf{A}(\mathbf{y})\partial^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'}_{\mathbf{y}}\nabla_{\mathbf{x}}\hat{u}(\mathbf{y}).$$

Thus, rearranging the preceding expression and using the linearity of the gradient, we arrive at

$$\begin{aligned}&\int_{D_{\mathrm{ref}}} \mathbf{A}(\mathbf{y})\nabla_{\mathbf{x}}\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})\nabla_{\mathbf{x}}v\,\mathrm{d}\mathbf{x}\\&\quad=\int_{D_{\mathrm{ref}}}\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}f_{\mathrm{ref}}(\mathbf{y})v\,\mathrm{d}\mathbf{x}-\sum_{\boldsymbol{\alpha}\neq\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}\int_{D_{\mathrm{ref}}}\partial_{\mathbf{y}}^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'}\mathbf{A}(\mathbf{y})\nabla_{\mathbf{x}}\partial_{\mathbf{y}}^{\boldsymbol{\alpha}'}\hat{u}(\mathbf{y})\nabla_{\mathbf{x}}v\,\mathrm{d}\mathbf{x}.\end{aligned}$$

By choosing $v=\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})$ and by employing the estimates from Theorem (3.15) and Theorem (3.18), it follows that

$$\begin{aligned}&\frac{\underline{\sigma}^d}{\overline{\sigma}^2}\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})\|_{H^1(D_{\mathrm{ref}})}^2\\&\quad\leqslant\int_{D_{\mathrm{ref}}}\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}f_{\mathrm{ref}}(\mathbf{y})\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})\,\mathrm{d}\mathbf{x}-\sum_{\boldsymbol{\alpha}\neq\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}\int_{D_{\mathrm{ref}}}\partial_{\mathbf{y}}^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'}\mathbf{A}(\mathbf{y})\nabla_{\mathbf{x}}\partial_{\mathbf{y}}^{\boldsymbol{\alpha}'}\hat{u}(\mathbf{y})\nabla_{\mathbf{x}}\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})\,\mathrm{d}\mathbf{x}\\&\quad\leqslant|\boldsymbol{\alpha}|!c_f c_D\boldsymbol{\mu}^{\boldsymbol{\alpha}}\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})\|_{H^1(D_{\mathrm{ref}})}\\&\qquad+\sum_{\boldsymbol{\alpha}\neq\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|!\boldsymbol{\mu}^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'}\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}'}\hat{u}(\mathbf{y})\|_{H^1(D_{\mathrm{ref}})}\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})\|_{H^1(D_{\mathrm{ref}})}.\end{aligned}$$

From this, we obtain

$$\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})\|_{H^1(D_{\mathrm{ref}})}\leqslant\frac{C}{4}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}}+\frac{C}{4}\sum_{\boldsymbol{\alpha}\neq\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|!\boldsymbol{\mu}^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'}\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}'}\hat{u}(\mathbf{y})\|_{H^1(D_{\mathrm{ref}})}$$

by setting

$$C:=\frac{4\overline{\sigma}^2}{\underline{\sigma}^d}\max\{1,c_f c_D\}.$$

The proof is now by induction on $|\boldsymbol{\alpha}|$. The induction hypothesis is given by

$$\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})\|_{H^1(D_{\mathrm{ref}})}\leqslant|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}}C^{|\boldsymbol{\alpha}|+1}.$$

For $|\boldsymbol{\alpha}|=0$, we conclude just the stability estimate (3.20), where the right hand side of the inequality is scaled by the factor 4. Therefore, let the assertion hold for all $|\boldsymbol{\alpha}|\leqslant n-1$ for some $n\geqslant 1$. Then, we have

$$\begin{aligned}&\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})\|_{H^1(D_{\mathrm{ref}})}\\&\quad\leqslant\frac{C}{4}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}}+\frac{C}{4}\sum_{\boldsymbol{\alpha}\neq\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|!\boldsymbol{\mu}^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'}|\boldsymbol{\alpha}'|!\boldsymbol{\mu}^{\boldsymbol{\alpha}'}C^{|\boldsymbol{\alpha}'|+1}\\&\quad\leqslant\frac{C}{4}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}}+\frac{C}{4}\boldsymbol{\mu}^{\boldsymbol{\alpha}}\sum_{\boldsymbol{\alpha}\neq\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|!C^{|\boldsymbol{\alpha}'|+1}\\&\quad=\frac{C}{4}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}}+\frac{C}{4}\boldsymbol{\mu}^{\boldsymbol{\alpha}}\sum_{j=0}^{n-1}\sum_{\substack{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}\\|\boldsymbol{\alpha}'|=j}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|!|\boldsymbol{\alpha}'|!C^{|\boldsymbol{\alpha}'|+1}.\end{aligned}$$

Again, we make use of the combinatorial identity (3.16) and obtain the estimate

$$\begin{aligned}\left\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})\right\|_{H^1(D_{\mathrm{ref}})} &\leqslant \frac{C}{4}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}} + \frac{C}{4}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}}\sum_{j=0}^{n-1}\binom{|\boldsymbol{\alpha}|}{j}(|\boldsymbol{\alpha}|-j)!j!C^{|\boldsymbol{\alpha}'|+1}\\ &= \frac{C}{4}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}} + \frac{C}{4}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}}C\sum_{j=0}^{n-1}C^{|\boldsymbol{\alpha}'|}\\ &= \frac{C}{4}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}} + \frac{C}{4}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}}C\frac{C^{|\boldsymbol{\alpha}|}}{C-1}.\end{aligned}$$

Now, the application of Lemma (A.1.7) gives us

$$\frac{C}{2}\frac{C^{|\boldsymbol{\alpha}|}}{C-1} \leqslant C^{|\boldsymbol{\alpha}|}$$

Since $C > 1$, we conclude

$$\left\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}}\hat{u}(\mathbf{y})\right\|_{H^1(D_{\mathrm{ref}})} \leqslant \frac{C}{4}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}} + C\frac{C^{|\boldsymbol{\alpha}|}}{2}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}} \leqslant C^{|\boldsymbol{\alpha}|+1}|\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}}.$$

This completes the proof. □

Taking into account the factor provided by the theorem, we end up with the sequence

$$\{\mu_k\}_k := \left\{\frac{8\overline{\sigma}^2}{\underline{\sigma}^d}C_{\mathrm{det}}\max\{1, c_f c_D\}\max\left(\frac{4d\overline{\sigma}^d}{\underline{\sigma}\rho\log 2}, \frac{4\overline{\sigma}^d(1+c_\gamma)}{\underline{\sigma}^4\log 2}\right)\tilde{\gamma}_k\right\}_k.$$

We are aware, that the constant to be spent could be improved. Nevertheless, we lay here our emphasis on the basic techniques which are used to obtain regularity results and do not focus on deriving optimal constants. Crucial is the fact, that we end up with a fixed constant independent of M. Moreover, we notice that the decay of the derivatives of the solution $\hat{u}(\mathbf{y})$ is only reduced by a single convergent series compared to the original series $\boldsymbol{\gamma}$ provided by the decay of the Karhunen-Loève expansion.

(3.22) **Remark.** The discussion in this section only refers to the case of the Poisson equation. Of course, the analysis presented here immediately applies also to the more general diffusion problem

$$-\operatorname{div}\left(\alpha(\mathbf{x})\nabla u(\mathbf{x},\mathbf{y})\right) = f(\mathbf{x}) \quad \text{for } \mathbf{x} \in D(\mathbf{y}).$$

In this case, one has to impose the restriction that $\alpha(\mathbf{x})$ is an analytic function which is uniformly bounded from above and below away from 0. Then, an estimate analogous to that of Lemma (3.17) applies for $\hat{\alpha}(\mathbf{x},\mathbf{y})$. The proof of a result similar to Theorem (3.15) for $\hat{\alpha}(\mathbf{x},\mathbf{y})\mathbf{A}(\mathbf{x},\mathbf{y})$ then involves an additional application of the Leibniz rule. The same holds for the case of a diffusion matrix which is analytic with respect to $\mathbf{x}$.

In order to apply a Quasi-Monte Carlo quadrature based on Halton points, we have a straightforward modification of [HPS13b, Theorem 4.9].

Let $\{\boldsymbol{\xi}_i\}_{i=1}^N \subset \square$ denote the first N *Halton points*. The related Quasi-Monte Carlo quadrature for $\hat{u}$ is given by

$$(Q\hat{u})(\mathbf{x}) := \frac{1}{N}\sum_{i=1}^{N} \hat{u}(\mathbf{x}, \boldsymbol{\xi}_i).$$

Then, we have the following convergence result for this quadrature.

(3.23) **Corollary.** For all $\delta > 0$, there exists a constant such that the Quasi-Monte Carlo quadrature based on Halton points for approximating the expectation of the solution $\hat{u}$ to (2.5) is strongly tractable if the sequence $\{\boldsymbol{\mu}_k\}_k$ is bounded by $\mu_k \lesssim k^{-3-\varepsilon}$ for arbitrary $\varepsilon > 0$. Especially, we have for the quadrature error based on N points the estimate

$$\|(\mathrm{Id} - Q)\hat{u}\|_{H^1(D_{\mathrm{ref}})} \lesssim N^{\delta-1}$$

with a constant only dependent on δ.

Proof. From Theorem (3.21), we derive for the first order mixed derivatives, i.e. $\|\boldsymbol{\alpha}\|_\infty = 1$, that

$$\left\|\partial_{\mathbf{y}}^{\boldsymbol{\alpha}} \hat{u}(\mathbf{y})\right\|_{H^1(D_{\mathrm{ref}})} \lesssim |\boldsymbol{\alpha}|!\boldsymbol{\mu}^{\boldsymbol{\alpha}} \leqslant \prod_{i=1}^{M} (k\mu_k).$$

Now, the proof is analogously to that of [HPS13b, Theorem 4.9]. □

Notice that we can obtain similar approximation results for the moments of $\hat{u}$, i.e. for $\hat{u}^p$ with $p \in \mathbb{N}$, possibly with worse constants. To that end, one has to bound the derivatives of $\hat{u}^p$ with respect to $\mathbf{y}$, too. This is also achieved by the application of Faà di Bruno's formula. For an idea of the related proofs, we refer to [HPS13b] where this topic is discussed in case of a stochastic diffusion coefficient.

4. Decay of the univariate derivatives

The results from the preceding section can be considerably sharpened if we only consider univariate derivatives $\partial_{y_i}^{\alpha} \hat{u}$ of the solution $\hat{u}$ to (2.5). The major obstruction in deriving estimates without powers of the term $|\boldsymbol{\alpha}|!$ in the estimates is the knowledge of proper bounds on the term $|\boldsymbol{\alpha}|!/\boldsymbol{\alpha}!$. To that end, we have only Lemma (A.1.4) at hand which tells us, that we have to spend a convergent series in order to bound the term $|\boldsymbol{\alpha}|!/\boldsymbol{\alpha}!$. The situation changes if one only considers univariate derivatives since then $|\alpha|! = \alpha!$ holds. This gives rise to a separate discussion of this situation.

We begin by improving the result on the derivatives of the Jacobian's determinant.

(4.1) **Lemma.** For the univariate derivatives of $\det \mathbf{J}(\mathbf{x}, \mathbf{y})$, it holds under the conditions of Assumption (II.3.3) that

$$\left\|\partial_{y_i}^{\alpha} \det \mathbf{J}\right\|_{L^\infty(\square; L^\infty(D_{\mathrm{ref}}))} \leqslant \alpha!\overline{\sigma}^d \left(\frac{2d}{\underline{\sigma}}\right)^{\alpha} \gamma_i^{\alpha}.$$

Proof. The univariate Faà di Bruno formula, cf. [CS96], yields

$$\begin{aligned}&\|\partial_{y_i}^\alpha \det \mathbf{J}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))}\\&\quad= \Big\| \sum_{r=1}^{\alpha} \exp\big(\operatorname{trace}\log \mathbf{J}\big) \sum_{P(\alpha,r)} \alpha! \prod_{j=1}^{\alpha} \frac{\big(\partial_{y_i}^j \operatorname{trace}\log \mathbf{J}\big)^{k_j}}{k_j!(j!)^{k_j}} \Big\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))}\end{aligned}$$

with $P(\alpha,r)$ being the set from (A.1.1). $P(\alpha,r)$ contains the partitions of a set with α elements into k_1 classes of cardinality 1 up to k_α classes of cardinality α, cf. [CS96].

It holds by estimate (3.12) that

$$\begin{aligned}&\Big\| \sum_{r=1}^{\alpha} \exp\big(\operatorname{trace}\log \mathbf{J}\big) \sum_{P(\alpha,r)} \alpha! \prod_{j=1}^{\alpha} \frac{\big(\partial_{y_i}^j \operatorname{trace}\log \mathbf{J}\big)^{k_j}}{k_j!(j!)^{k_j}} \Big\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))}\\&\quad\leqslant \sum_{r=1}^{\alpha} \big\| \exp\big(\operatorname{trace}\log \mathbf{J}\big)\big\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \sum_{P(\alpha,r)} \alpha! \prod_{j=1}^{\alpha} \frac{\Big(dj!\big(\frac{1}{\underline{\sigma}}\big)^j \gamma_i^j\Big)^{k_j}}{k_j!(j!)^{k_j}}\\&\quad\leqslant \overline{\sigma}^d \Big(\frac{d}{\underline{\sigma}}\Big)^\alpha \gamma_i^\alpha \sum_{r=1}^{\alpha} \sum_{P(\alpha,r)} \alpha! \prod_{j=1}^{\alpha} \frac{1}{k_j!}.\end{aligned}$$

Now, the assertion is easily obtained from the identity, cf. [CS96],

$$r! \sum_{P(\alpha,r)} \prod_{j=1}^{\alpha} \frac{1}{k_j!} = \binom{\alpha-1}{r-1}$$

and the estimate

$$\sum_{r=1}^{\alpha} \frac{\alpha!}{r!} \binom{\alpha-1}{r-1} \leqslant 2^\alpha \alpha!.$$

□

The sharpened estimate for the Jacobian's determinant yields together with Lemma (3.2) an improved estimate for the univariate derivatives of the diffusion matrix $\mathbf{A}(\mathbf{x},\mathbf{y})$.

(4.2) **Theorem.** It holds for the univariate derivatives of the diffusion matrix $\mathbf{A}(\mathbf{x},\mathbf{y})$ that

$$\|\!|\partial_{y_i}^\alpha \mathbf{A}|\!\|_{d\times d} \leqslant (\alpha+1)! \frac{\overline{\sigma}^d}{\underline{\sigma}^2} \Big(\frac{2d(1+c_\gamma)}{\underline{\sigma}^2 \log 2}\Big)^\alpha \gamma_i^\alpha.$$

Proof. The Leibniz rule for $\partial_{y_i}^\alpha \mathbf{A}(\mathbf{x},\mathbf{y})$ yields

$$\partial_{y_i}^\alpha \mathbf{A}(\mathbf{x},\mathbf{y}) = \sum_{r=0}^{\alpha} \binom{\alpha}{r} \partial_{y_i}^r \big(\mathbf{J}(\mathbf{x},\mathbf{y})^\intercal \mathbf{J}(\mathbf{x},\mathbf{y})\big)^{-1} \partial_{y_i}^{\alpha-r} \det \mathbf{J}(\mathbf{x},\mathbf{y}).$$

Inserting the estimates from Lemma (3.2) and Lemma (4.1) yields

$$\begin{aligned}\|\!|\partial_{y_i}^\alpha \mathbf{A}|\!\|_{d\times d} &\leqslant \sum_{r=0}^{\alpha} \binom{\alpha}{r} r! \frac{\gamma_i^r}{\underline{\sigma}^2} \Big(\frac{2(1+c_\gamma)}{\underline{\sigma}^2 \log 2}\Big)^r (\alpha-r)! \overline{\sigma}^d \Big(\frac{2d}{\underline{\sigma}}\Big)^{\alpha-r} \gamma_i^{\alpha-r}\\&\leqslant \frac{\overline{\sigma}^d}{\underline{\sigma}^2} \Big(\frac{2d(1+c_\gamma)}{\underline{\sigma}^2 \log 2}\Big)^\alpha \gamma_i^\alpha \sum_{r=0}^{\alpha} \binom{\alpha}{r} r!(\alpha-r)! = (\alpha+1)! \frac{\overline{\sigma}^d}{\underline{\sigma}^2} \Big(\frac{2d(1+c_\gamma)}{\underline{\sigma}^2 \log 2}\Big)^\alpha \gamma_i^\alpha.\end{aligned}$$

□

In complete analogy to the previous theorem, we obtain a bound for the univariate derivatives of the right hand side f_{ref}.

(4.3) **Theorem.** It holds for the univariate derivatives of the right hand side $f_{\mathrm{ref}}(\mathbf{x},\mathbf{y})$ that

$$\|\partial_{y_i}^{\alpha} f_{\mathrm{ref}}\|_{L^\infty(\square;L^\infty(D_{\mathrm{ref}}))} \leqslant (\alpha+1)! c_f \overline{\sigma}^d \left(\frac{2d}{\underline{\sigma}\rho\log 2}\right)^{\alpha} \gamma_i^{\alpha}.$$

The results provided by Theorem (4.2) and Theorem (4.3) are sufficient to show that the solution $\hat{u}$ to (2.5) exhibits an analytic extension into the complex plane with respect to each particular direction y_i. For a proof of this statement, see [BNT07, Lemma 3.2]. This implies the applicability of an anisotropic stochastic collocation method, cf. [BNT07]. Especially, we obtain rates of convergence for the elliptic diffusion problem on random domains in terms of the specific decay of the perturbation field's Karhunen-Loève expansion.

5. Stochastic interface problems

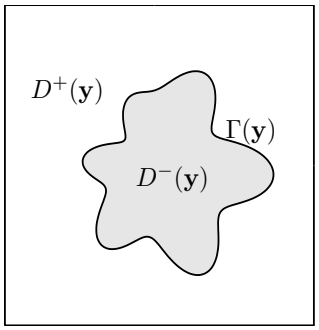

Figure IV.1: Visualization of the domain D and the random interface $\Gamma(\mathbf{y})$.

As a special case of a diffusion problem on random domains, we shall focus on stochastic interface problems as already discussed in e.g. [HL13].

Let the hold-all $\mathcal{D} \subset \mathbb{R}^d$, cf. (1.2), be a simply-connected and convex domain with Lipschitz continuous boundary $\partial\mathcal{D}$. Inscribed into $\mathcal{D}$, we have a randomly varying inclusion $D^-(\mathbf{y}) \subsetneq \mathcal{D}$ for $\mathbf{y} \in \square$ with a C^2-smooth boundary $\Gamma(\mathbf{y}) := \partial D^-(\mathbf{y})$. The complement of $\overline{D^-(\mathbf{y})}$ will be denoted by $D^+(\mathbf{y}) := \mathcal{D} \setminus \overline{D^-(\mathbf{y})}$. A visualization of this setup is found in Figure IV.1.

For given $\mathbf{y} \in \square$, we can state the stochastic elliptic interface problem as follows:

$$\begin{aligned} -\operatorname{div}\left(\alpha(\mathbf{x},\mathbf{y})\nabla u(\mathbf{x},\mathbf{y})\right) &= f(\mathbf{x}) && \text{in } \mathcal{D}\setminus\Gamma(\mathbf{y}),\\ [\![u(\mathbf{x},\mathbf{y})]\!] &= 0 && \text{on } \Gamma(\mathbf{y}),\\ \left[\!\!\left[\alpha(\mathbf{x},\mathbf{y})\frac{\partial u}{\partial \mathbf{n}}(\mathbf{x},\mathbf{y})\right]\!\!\right] &= 0 && \text{on } \Gamma(\mathbf{y}),\\ u(\mathbf{x},\mathbf{y}) &= 0 && \text{on } \partial\mathcal{D}. \end{aligned} \tag{5.1}$$

Here, $\mathbf{n}$ denotes the outward normal vector on $\Gamma(\mathbf{y})$. Furthermore, the diffusion coefficient shall be of the form

$$\alpha(\mathbf{x},\mathbf{y}) := \chi_{D^+(\mathbf{y})}(\mathbf{x})\alpha^+(\mathbf{x}) + \chi_{D^-(\mathbf{y})}(\mathbf{x})\alpha^-(\mathbf{x}) \quad \text{for } \mathbf{x} \in \mathcal{D},$$

where $\chi_{D^-(\mathbf{y})}$ is the characteristic function of $D^-(\mathbf{y})$ and α^+, α^- are smooth deterministic functions with

$$0 < \underline{\alpha} \leqslant \alpha^-(\mathbf{x}), \alpha^+(\mathbf{x}) \leqslant \overline{\alpha} < \infty \quad \text{for almost every } \mathbf{x} \in \mathcal{D}.$$

By $[\![u(\mathbf{x},\mathbf{y})]\!] := u^+(\mathbf{x},\mathbf{y}) - u^-(\mathbf{x},\mathbf{y})$, we denote the jump of the solution u across $\Gamma(\mathbf{y})$, where $u^-(\mathbf{x},\mathbf{y}) := u|_{D^-(\mathbf{y})}$ and $u^+(\mathbf{x},\mathbf{y}) := u|_{D^+(\mathbf{y})}$, respectively. Analogously, we define the jump of the co-normal derivative across $\Gamma(\mathbf{y})$ via

$$\left[\!\!\left[\alpha(\mathbf{x},\mathbf{y})\frac{\partial u}{\partial \mathbf{n}}(\mathbf{x},\mathbf{y})\right]\!\!\right] := \alpha^+(\mathbf{x})\frac{\partial u}{\partial \mathbf{n}}(\mathbf{x},\mathbf{y}) - \alpha^-(\mathbf{x})\frac{\partial u}{\partial \mathbf{n}}(\mathbf{x},\mathbf{y}).$$

(5.2) **Remark.** This formulation of the stochastic interface problem is strongly related with an elliptic equation on a stochastic domain. For example, for $\alpha^+(\mathbf{x}) \equiv 0$ and $\alpha^-(\mathbf{x}) \equiv 1$ (perfect insulation), we have the Poisson equation on $D^-(\mathbf{y})$ with homogeneous Neumann data on $\Gamma(\mathbf{y})$ while, for $\alpha^+(\mathbf{x}) \equiv \infty$ and $\alpha^-(\mathbf{x}) \equiv 1$ (perfect conduction), we have the Poisson equation on $D^-(\mathbf{y})$ with homogeneous Dirichlet data on $\Gamma(\mathbf{y})$.

Modeling the stochastic interface

Instead of solving the stochastic interface problem by a perturbation method by means of shape sensitivity analysis like in [HL13, HSS08b], we propose here to apply the domain mapping approach. To that end, let $\Gamma_{\text{ref}} \subset \mathcal{D}$ denote a reference interface of class C^2 and co-dimension 1 which separates the interior domain D^-_{ref} and the outer domain D^+_{ref}. We assume that $\Gamma(\mathbf{y})$ is prescribed by an application of a vector field $\mathbf{V}\colon \mathcal{D} \times \square \to \mathcal{D}$, i.e.

$$\Gamma(\mathbf{y}) = \mathbf{V}(\Gamma_{\text{ref}}, \mathbf{y}),$$

which is a C^2-diffeomorphism for almost every $\mathbf{y} \in \square$. Furthermore, let the Jacobian of $\mathbf{V}$ satisfy the ellipticity condition (1.3).

As an example, we can consider here an extension of the vector field in [HL13], which only defines the perturbation at the boundary: If Γ_{ref} is of class C^3, then its outward normal $\mathbf{n}$ is of class C^2. Thus, given a stochastic field $\kappa\colon \Gamma_{\text{ref}} \times \square \to \mathbb{R}$ which satisfies $|\kappa(\mathbf{x},\mathbf{y})| \leqslant \overline{\kappa} < 1$ almost surely, we can define $\mathbf{V}(\mathbf{x},\mathbf{y}) := \mathbf{x} + \kappa(\mathbf{x},\mathbf{y})\mathbf{n}(\mathbf{x})$ for $\mathbf{x} \in \Gamma_{\text{ref}}$. A suitable extension of this vector field to the whole domain $\mathcal{D}$ is given by

$$\mathbf{V}(\mathbf{x},\mathbf{y}) := \mathbf{x} + \kappa(P\mathbf{x},\mathbf{y})\mathbf{n}(P\mathbf{x})B(\|\mathbf{x} - P\mathbf{x}\|_2),$$

where $P\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto Γ_{ref} and $B\colon [0,\infty) \to [0,1]$ is a smooth blending function with $B(0) = 1$ and $B(t) = 0$ for all $t \geq c$ for some constant $c \in (0,\infty)$. Notice that, if Γ_{ref} is of class C^3, the orthogonal projection P onto Γ_{ref} and thus $\mathbf{V}(\mathbf{x},\mathbf{y})$ is at least of class C^2, cf. [Hol73].

Reformulation for the reference interface

For $\mathbf{y} \in \square$, the variational formulation of the interface problem (5.1) is given as follows:

$$\text{Find } u \in H_0^1(\mathcal{D}) \text{ such that}$$
$$\int_{D^-(\mathbf{y})\cup D^+(\mathbf{y})} \alpha \langle \nabla u, \nabla v \rangle \, \mathrm{d}\mathbf{x} = \int_{\mathcal{D}} f v \, \mathrm{d}\mathbf{x} \quad \text{for all } v \in H_0^1(\mathcal{D}).$$

As in Section 2, we can reformulate this variational formulation relative to the reference interface. As we have for the transported coefficient

$$\begin{aligned}\hat{\alpha}(\mathbf{x},\mathbf{y}) &= \chi_{\mathbf{V}(D^+_{\mathrm{ref}},\mathbf{y})}(\mathbf{V}(\mathbf{x},\mathbf{y}))\hat{\alpha}^+(\mathbf{x},\mathbf{y}) + \chi_{\mathbf{V}(D^-_{\mathrm{ref}},\mathbf{y})}(\mathbf{V}(\mathbf{x},\mathbf{y}))\hat{\alpha}^-(\mathbf{x},\mathbf{y}) \\ &= \chi_{D^+_{\mathrm{ref}}}(\mathbf{x})\hat{\alpha}^+(\mathbf{x},\mathbf{y}) + \chi_{D^-_{\mathrm{ref}}}(\mathbf{x})\hat{\alpha}^-(\mathbf{x},\mathbf{y}),\end{aligned}$$

we obtain the following variational formulation with the definition (2.3) of the diffusion matrix $\mathbf{A}(\mathbf{x},\mathbf{y})$:

$$\text{Find } \hat{u}(\mathbf{y}) \in H_0^1(\mathcal{D}) \text{ such that} \tag{5.3}$$
$$\int_{D^-_{\mathrm{ref}}\cup D^+_{\mathrm{ref}}} \hat{\alpha}(\mathbf{y}) \langle \mathbf{A}(\mathbf{y}) \nabla_{\mathbf{x}} \hat{u}(\mathbf{y}), \nabla_{\mathbf{x}} v \rangle \, \mathrm{d}\mathbf{x} = \int_{\mathcal{D}} \hat{f}(\mathbf{y}) v \det \mathbf{J}(\mathbf{y}) \, \mathrm{d}\mathbf{x}$$

for all $v \in H_0^1(\mathcal{D})$. Since $\hat{\alpha}(\mathbf{x},\mathbf{y})$ is a smooth function with respect to $\mathbf{y}$, the regularity results from Section 3 remain valid here.

Finite element approximation for the stochastic interface problem

The application of parametric finite elements results especially in an interface-resolved triangulation for the discretization of the interface stochastic problem (5.1). By "interface-resolved" we mean that the vertices of elements around the interface lie exactly on the interface, cf. [CZ98, LMWZ10]. Thus, the approximation error for a particular realization $u(\mathbf{y}_i)$ of the solution $u(\mathbf{y})$ to the stochastic interface problem (5.1) can be quantified by the following theorem adopted from [LMWZ10, Theorem 4.1].

(5.4) **Theorem.** For $\mathbf{y} \in \square$, let $\{\mathcal{T}_j\}_{j>0}$ be a family of interface resolved triangulations for $\mathbf{V}(\mathcal{D},\mathbf{y})$ and $\{V_j(\mathbf{y})\}_{j>0}$ the associated finite element spaces. Let $u_j(\mathbf{y})$ be the finite element solution corresponding to the realization $u(\mathbf{y})$ of the solution to the elliptic problem (5.1). Then, for $s = 0, 1$, there holds that

$$\|u(\mathbf{y}) - u_j(\mathbf{y})\|_{H^s(\mathcal{D})} \lesssim h_j^{2-s} \|u(\mathbf{y})\|_{H^2(D^-(\mathbf{y}))\cup H^2(D^+(\mathbf{y}))}, \tag{5.5}$$

where $H^2(D^-(\mathbf{y})) \cup H^2(D^+(\mathbf{y}))$ is the broken Sobolev space equipped by the norm

$$\|\cdot\|_{H^2(D^-(\mathbf{y}))\cup H^2(D^+(\mathbf{y}))} := \sqrt{\|\cdot\|^2_{H^2(D^-(\mathbf{y}))} + \|\cdot\|^2_{H^2(D^+(\mathbf{y}))}}.$$

In view of Theorem (2.9), the statement of the previous theorem is also valid for the realization of the solution which is pulled back to the domain $\mathcal{D}$ relative to the reference interface Γ_{ref}.

6. Numerical results

In this section, we consider examples for boundary value problems on random domains. On the one hand, we consider a stochastic interface problem and on the other hand, we consider the Poisson equation on two different random domains. In all examples, we employ the pivoted Cholesky decomposition, in order to approximate the Karhunen-Loève expansion of $\mathbf{V}$. The spatial discretization is performed by using piecewise linear parametric finite elements on the mapped domain $\mathbf{V}(D_{\text{ref}}, \mathbf{y}_i)$ for each sample $\mathbf{y}_i \in \square$. It would of course be also possible to perform the computations on the reference domain. In this case, the diffusion matrix $\mathbf{A}$ has to be computed from the Karhunen-Loève expansion of $\mathbf{V}$ for each particular sample. All computations have been carried out on a computing server consisting of four nodes[5] with up to 64 threads.

The stochastic interface problem

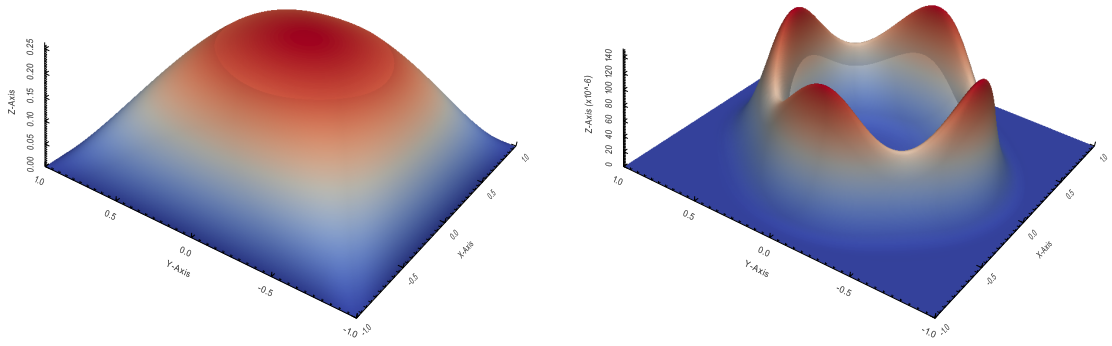

Figure IV.2: Mean (left) and variance (right) of the solution $\hat{u}$ to the stochastic interface problem.

We consider the stochastic interface problem from [HL13] where the hold-all is given as $\mathcal{D} = (-1, 1)^2$ and the reference interface is given as $\Gamma_{\text{ref}} = \{\mathbf{x} \in \mathcal{D} : \|\mathbf{x}\|_2 = 0.5\}$. Thus, the outward normal at Γ_{ref} is $\mathbf{n}(\mathbf{x}) = [\cos(\theta), \sin(\theta)]^\intercal$, where $\mathbf{x} = 0.5[\cos(\theta), \sin(\theta)]^\intercal$ is the representation of $\mathbf{x} \in \Gamma_{\text{ref}}$ in polar coordinates. The random field under consideration reads

$$\kappa(\theta, \omega) = \frac{1}{30} \sum_{k=0}^{5} \cos(k\theta) X_{2k}(\omega) + \sin(k\theta) X_{2k+1}(\omega).$$

Here, $X_1, \ldots, X_{11}$ are independent, uniformly distributed random variables with variance 1, i.e. their range is $[-\sqrt{3}, \sqrt{3}]$. We extend this random field onto $\mathcal{D}$ as described in Section 5 by using the appropriately scaled, quadratic B-spline as blending function, i.e. $B(\mathbf{x}) = \frac{4}{3} B_2(3\|\mathbf{x} - P\mathbf{x}\|_2)$. This yields the covariance

$$\operatorname{Cov}[\mathbf{V}](\mathbf{x}, \mathbf{y}) = B(\mathbf{x}) B(\mathbf{y}) \operatorname{Cov}_\kappa(\theta_{\mathbf{x}}, \theta_{\mathbf{y}}) \begin{bmatrix} \cos(\theta_{\mathbf{x}}) \cos(\theta_{\mathbf{y}}) & \cos(\theta_{\mathbf{x}}) \sin(\theta_{\mathbf{y}}) \\ \sin(\theta_{\mathbf{x}}) \cos(\theta_{\mathbf{y}}) & \sin(\theta_{\mathbf{x}}) \sin(\theta_{\mathbf{y}}) \end{bmatrix}$$

[5] Each node consists of two quad-core Intel(R) Xeon(R) X5550 CPUs with a clock rate of 2.67GHz (hyperthreading enabled) and 48GB of main memory.

with

$$\mathrm{Cov}_\kappa(\theta_{\mathbf{x}}, \theta_{\mathbf{y}}) = \frac{1}{900} \sum_{k=0}^{5} \cos(k\theta_{\mathbf{x}}) \cos(k\theta_{\mathbf{y}}) + \sin(k\theta_{\mathbf{x}}) \sin(k\theta_{\mathbf{y}}).$$

Furthermore, we set $\mathbf{E}[\mathbf{V}](\mathbf{x}) := \mathbf{x}$. A visualization of the reference interface with a particular displacement field $\mathbf{V}_0(\mathbf{x}, \mathbf{y}_i)$ and the resulting perturbed interface is found in Figure IV.3. Finally, the diffusion coefficient is chosen as $\alpha^-(\mathbf{x}) \equiv 2$, $\alpha^+(\mathbf{x}) \equiv 1$ and the right hand side is chosen as $f(\mathbf{x}) \equiv 1$.

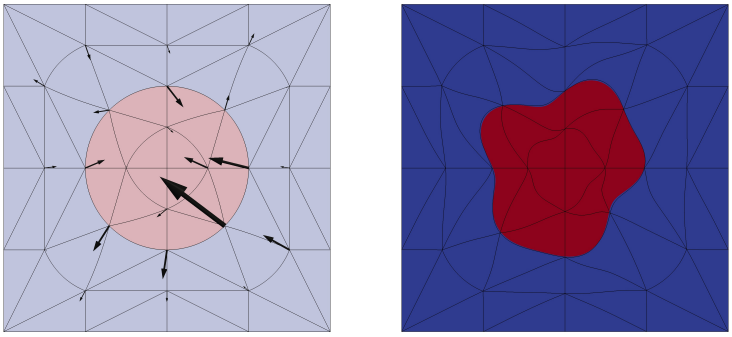

Figure IV.3: Realization of the displacement (left) and the related mapped interface (right).

A visualization of the mean and the variance computed by $N = 10^6$ Quasi-Monte Carlo samples and 1048576 finite elements (level 8) is shown in Figure IV.2. This approximation serves as a reference solution in order to examine the convergence behavior of the Quasi-Monte Carlo method. As a comparison and in order to validate the reference solution, we have also computed the approximate mean and variance on each level by the Monte Carlo method. According to [HPS13b] and our regularity results, the Quasi-Monte Carlo method with N Halton points converges with the rate $N^{\delta-1}$ for any $\delta > 0$. In our experiments, we thus apply $2^{j/(1-\delta)}$ Halton points on the finite element level $j = 1, \ldots, 7$ for the choices $\delta = 0.2, 0.3, 0.4, 0.5$. For the Monte Carlo method, we averaged five approximations each of which being computed with 2^{2j} samples.

In Figure IV.4, a visualization of the number of samples is found for the two methods and each particular choice of δ.

Figure IV.5 depicts the error of the solution's mean measured in the H^1-norm on the left hand side and the error of the solutions variance measured in the $W^{1,1}$-norm on the right hand side. As can be seen, the error of the solution's mean provides the expected linear rate of convergence for each of the choices of δ. For the solution's variance, we observe a certain offset for the choices $\delta = 0.2$ and $\delta = 0.3$ until the asymptotic rate of convergence is achieved. The choices $\delta = 0.4$ and $\delta = 0.5$ as well as the Monte Carlo approximation achieve even better rates of convergence than the expected linear rate. At least the error for the solution's mean seems to be dominated by the finite element discretization. Therefore, we found it instructive to present also the respective errors measured in the L^2-norm. They are plotted in Figure IV.6. Here, the error is dominated

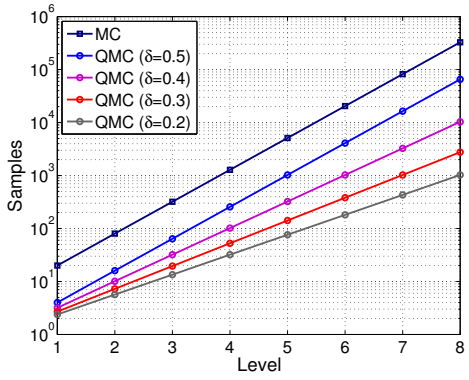

Figure IV.4: Number of samples for the two methods and the different choices of δ.

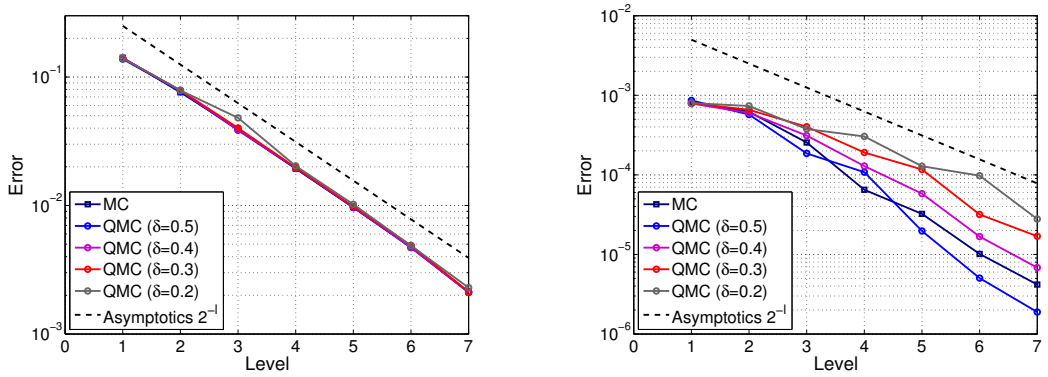

Figure IV.5: Error in the mean measured in H^1 (left) and in the variance measured in $W^{1,1}$ (right) for the stochastic interface problem.

by the stochastic discretization and we can observe that the rate of convergence increases as δ increases even for the solution's mean.

The Poisson equation on random domains

For our second example, we consider an infinite dimensional random field described by its mean $\mathbb{E}[\mathbf{V}](\mathbf{x}) = \mathbf{x}$ and its covariance function

$$\mathrm{Cov}[\mathbf{V}](\mathbf{x},\mathbf{y}) = \frac{1}{100}\begin{bmatrix} 5\exp(-4\|\mathbf{x}-\mathbf{y}\|_2^2) & \exp(-0.1\|2\mathbf{x}-\mathbf{y}\|_2^2) \\ \exp(-0.1\|\mathbf{x}-2\mathbf{y}\|_2^2) & 5\exp(-\|\mathbf{x}-\mathbf{y}\|_2^2) \end{bmatrix}.$$

Furthermore, we consider the random variables in the Karhunen-Loève expansion to be uniformly distributed. The unit disc $D_{\mathrm{ref}} = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 < 1\}$ serves as reference domain and the load is set to $f(\mathbf{x}) \equiv 1$. Figure IV.8 shows the reference domain with a particular displacement field and the resulting perturbed domain.

In Figure IV.7, a visualization of the mean and the variance computed by $N = 10^6$ Quasi-Monte Carlo samples and 1048576 finite elements (level 9) is found. Here, the

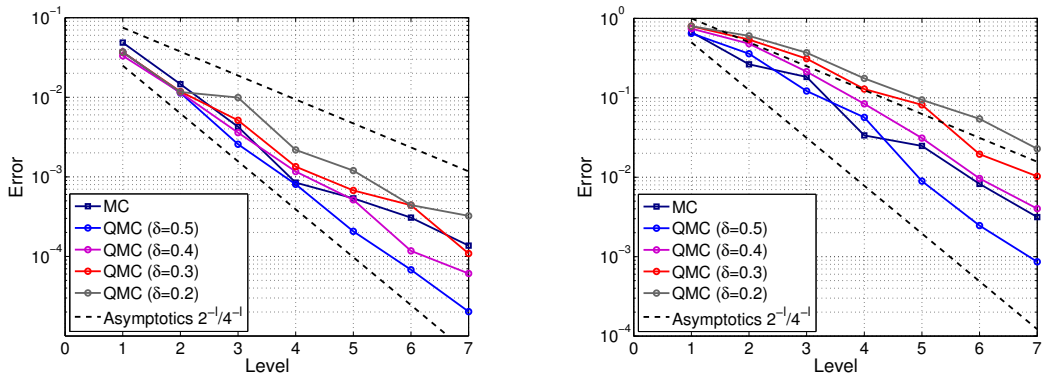

Figure IV.6: Error in the mean (left) and in the variance (right) measured in L^2 for the stochastic interface problem.

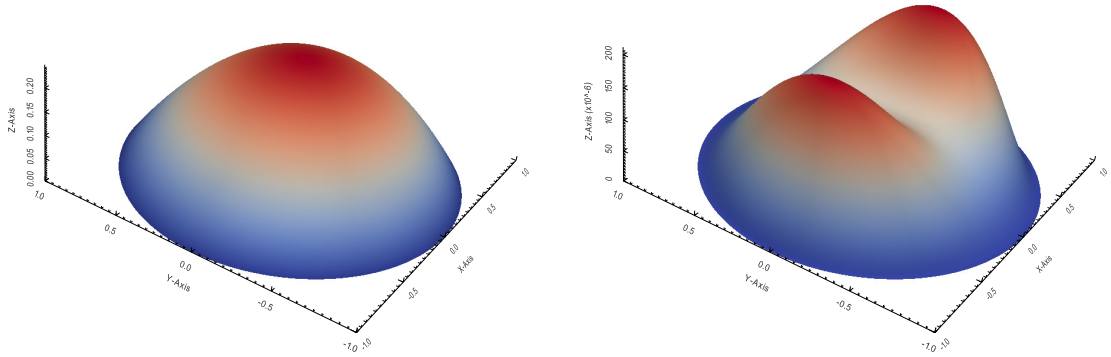

Figure IV.7: Mean (left) and variance (right) of the solution $\hat{u}$ to the Poisson equation on the unit disc.

Karhunen-Loève expansion has been truncated after $M = 303$ terms which leads to a truncation error that is smaller than 10^{-6}. For the convergence study, however, we have coupled the truncation error of the Karhunen-Loève expansion to the spatial discretization error of order 2^{-j} on the finite element level j. It is observed that the truncation rank M linearly grows in the level j, namely, it holds $M = 10, 23, 37, 49, 64, 79, 91, 108$ for $j = 1, 2, 3, 4, 5, 6, 7, 8$.

The number of samples of the quadrature methods under consideration has been chosen in dependence on the finite element level j as in the previous example. Figure IV.9 displays the error of the solution's mean and variance measured in the H^1-norm and the $W^{1,1}$-norm, respectively. For the mean, we observe again the expected linear rate of convergence with a slight deterioration for $\delta = 0.3$ and $\delta = 0.2$ on level 4. For the variance, we observe linear and even better rates of convergence except for $\delta = 0.2$. Again, we have also provided the respective errors with respect to the L^2-norm. The related plots are found in Figure IV.10. Here, the error is also dominated by the stochastic approximation. For larger values of δ, we again observe successively better rates of convergence.

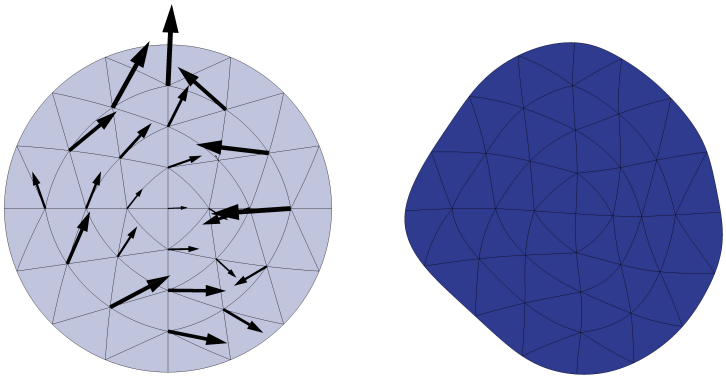

Figure IV.8: Realization of the displacement $\mathbf{V}_0(\mathbf{x}, \mathbf{y}_i)$ (left) and the related mapped unit disc (right).

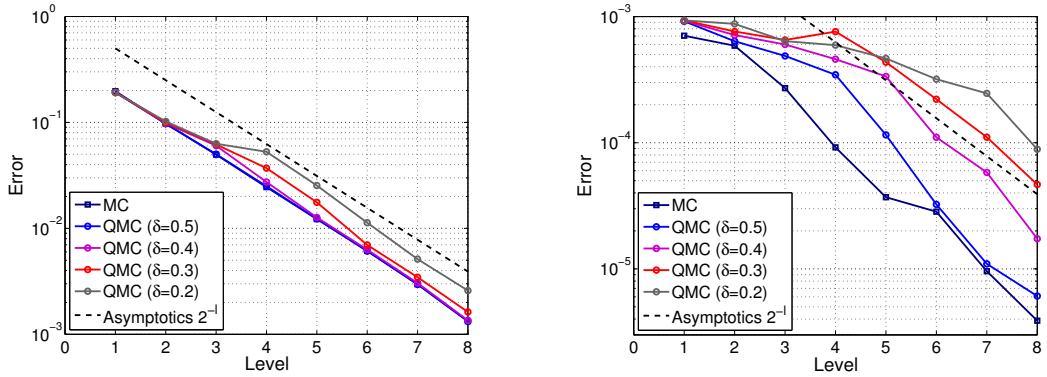

Figure IV.9: Error in the mean measured in H^1 (left) and in the variance measured in $W^{1,1}$ (right) on the unit disc.

Finally, for our last example, we consider an infinite dimensional random field described by its mean $\mathbb{E}[\mathbf{V}](\mathbf{x}) = \mathbf{x}$ and its covariance function

$$\mathrm{Cov}[\mathbf{V}](\mathbf{x}, \mathbf{y}) = \frac{1}{25} \begin{bmatrix} 2\exp(-4\|\mathbf{x} - \mathbf{y}\|_2^2) & 0 \\ 0 & \exp(-\|\mathbf{x} - \mathbf{y}\|_2^2) \end{bmatrix}.$$

Again, the random variables in the Karhunen-Loève expansion are assumed to be uniformly distributed. The reference domain in this example is given by the L-shape, i.e. $D_{\mathrm{ref}} = (-1, 1)^2 \setminus ([0, 1] \times [-1, 0])$. Figure IV.12 depicts a realization of the displacement for this reference domain. We use $f(\mathbf{x}) = 2\pi^2 \sin(\pi x_1) \sin(\pi x_2)$ as right hand side. In Figure IV.11, a visualization of the mean and the variance computed by $N = 10^6$ Quasi-Monte Carlo samples and 786432 finite elements (level 8) is found. Here, the Karhunen-Loève expansion has been truncated after $M = 343$ terms which yields a truncation error smaller than 10^{-6}. For the convergence study, we have again coupled the truncation error of the Karhunen-Loève expansion to the spatial accuracy 2^{-j} of the discretization on the finite element level j. The truncation rank M grows here also linearly in the level j, namely it holds $M = 14, 28, 40, 57, 70, 85, 101$ for $j = 1, 2, 3, 4, 5, 6, 7, 8$. The number of samples

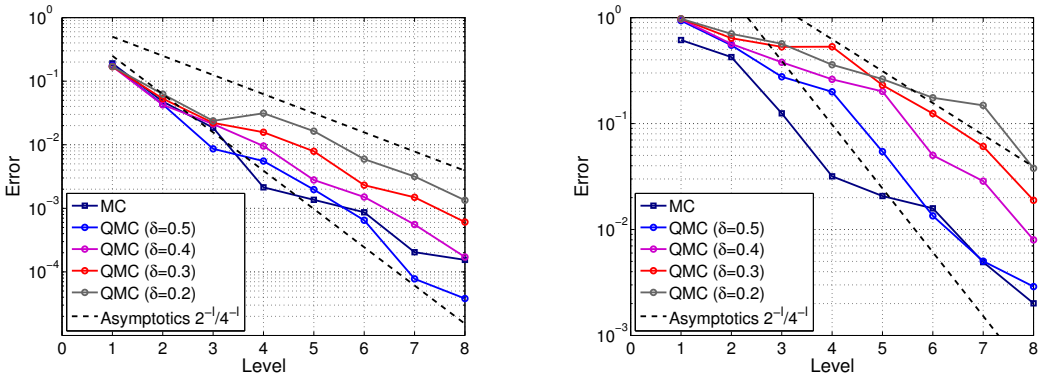

Figure IV.10: Error in the mean (left) and in the variance (right) measured in L^2 on the unit disc.

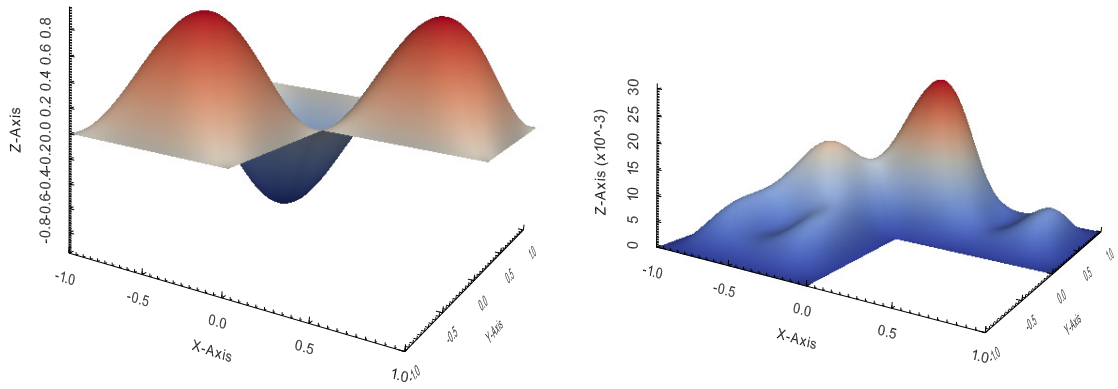

Figure IV.11: Mean (left) and variance (right) of the solution $\hat{u}$ to the Poisson equation on the L-shape.

of the quadrature methods under consideration has been chosen in dependence on the finite element level j as in the previous examples, cf. Figure IV.4. Figure IV.13 shows the error of the solution's mean and variance measured in the H^1-norm and the $W^{1,1}$-norm, respectively. For the mean, we observe also the expected linear rate of convergence with a slight deterioration for $\delta = 0.3$ on levels 3–4 and for $\delta = 0.2$ on level 3–5. For the variance, we observe an offset in the convergence for $\delta = 0.2, 0.3, 0.4$. Moreover, the choice $\delta = 0.2$ seems not sufficient here to achieve convergence for the computed levels. Again, we have also provided the respective errors with respect to the L^2-norm. The related plots are found in Figure IV.14. As it turns out, the error in this example is also dominated by the stochastic approximation. For larger values of δ, we again observe successively better rates of convergence. Unsurprisingly, $\delta = 0.2$ seems to be insufficient also with respect to the L^2-error.

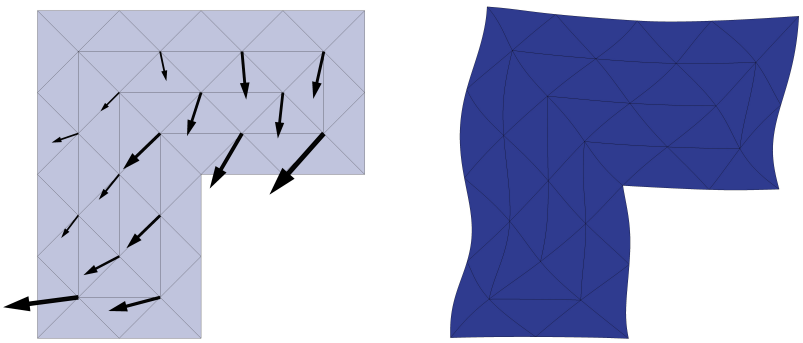

Figure IV.12: Realization of the displacement $\mathbf{V}_0(\mathbf{x}, \mathbf{y}_i)$ (left) and the related mapped L-shape (right).

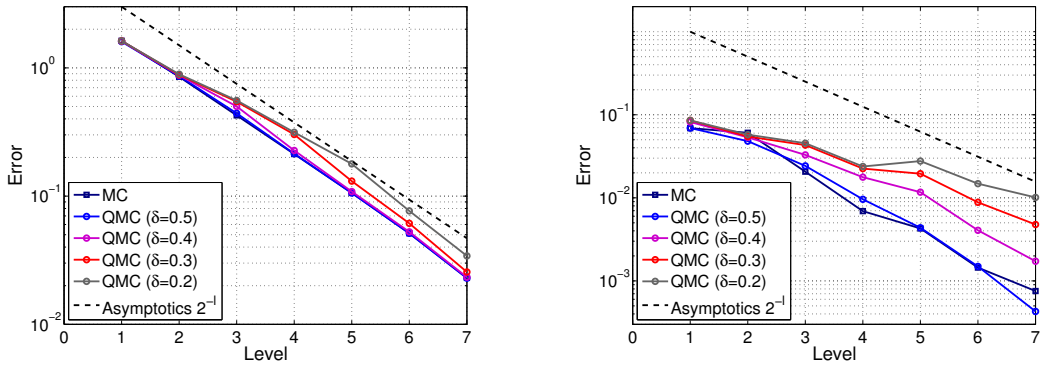

Figure IV.13: Error in the mean measured in H^1 (left) and in the variance measured in $W^{1,1}$ (right) on the L-shape.

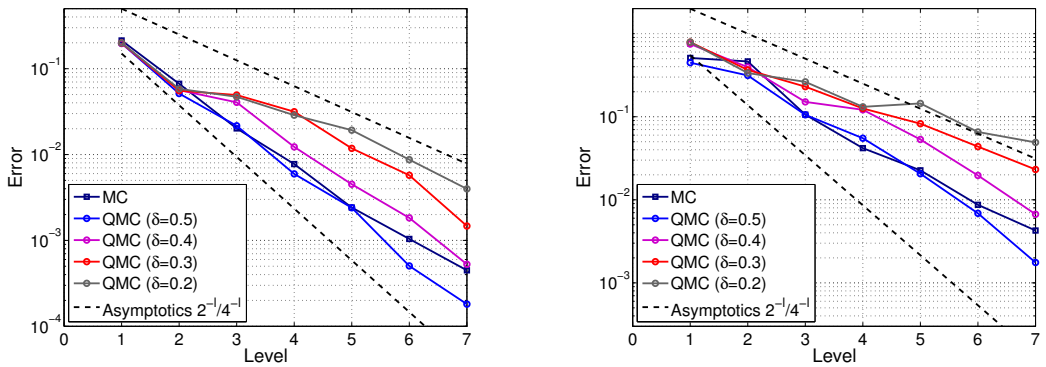

Figure IV.14: Error in the mean (left) and in the variance (right) measured in L^2 on the L-shape.

Chapter V

THE PERTURBATION METHOD

In this chapter, we propose another approach to the solution of elliptic diffusion equations on random domains. As already in Chapter IV, we restrict ourselves to the Poisson equation. Nevertheless, the linearization technique employed here also extends to the situation of elliptic diffusion problems with arbitrary diffusion matrix $\mathbf{A}(\mathbf{x})$, cf. [Har10b]. The present linearization technique has already been applied to random diffusion coefficients or even to elliptic equations on random domains in [Har10b, HSS08a, HSS08b]. In these works, however, either multilevel frames or wavelets have been used for the numerical realization of the approach. We intend to use the *combination technique* as proposed in [HPS13a] for elliptic equations with random diffusion.

1. Shape sensitivity analysis

In this section, we briefly summarize results on shape sensitivity analysis for our model problem. For a more general framework and the details on this topic, we refer to the works [DZ11, Epp07, SZ92] and the references therein.

Assume that D_{ref} is of class C^2. This smoothness assumption guarantees the H^2-regularity of the model problem, cf. [SZ92, Proposition 2.83]. Moreover, let $\mathbf{V} \in C^2(\mathbb{R}^d; \mathbb{R}^d)$ be a vector field. We may define the family of transformations $\{T_\varepsilon\}_{\varepsilon>0}$ by the perturbations of identity

$$T_\varepsilon(\mathbf{x}) = \mathrm{Id}(\mathbf{x}) + \varepsilon \mathbf{V}(\mathbf{x}).$$

Then, there exists an $\varepsilon_0 > 0$ such that the transformations T_ε are C^2-diffeomorphisms for all $\varepsilon \in [0, \varepsilon_0]$. This is a consequence from the results in [Sim80, Section 1.1], see also Remark (II.6.11). The related family of domains will be denoted by $D_\varepsilon := T_\varepsilon(D_{\text{ref}})$. We shall consider the Poisson equation on these domains, i.e.

$$-\Delta u_\varepsilon = f \text{ in } D_\varepsilon, \quad u_\varepsilon = 0 \text{ on } \Gamma_\varepsilon := \partial D_\varepsilon. \tag{1.1}$$

Here, in order to guarantee the well-posedness of the equation, we assume that the right hand side is defined on the holdall

$$\mathcal{D} = \bigcup_{0 \leqslant \varepsilon \leqslant \varepsilon_0} D_\varepsilon.$$

Now, we have for the weak solution $u_\varepsilon \in H^s(D_\varepsilon)$ with $s \in [0,2]$ that

$$u^\varepsilon = u_\varepsilon \circ T_\varepsilon \in H^s(D_{\mathrm{ref}})$$

for all $\varepsilon \in [0, \varepsilon_0]$, see e.g. [SZ92]. Especially, we set $\overline{u} := u_0 \in H^s(D_{\mathrm{ref}})$. Then, we may define the *material derivative* of u as in [SZ92, Definition 2.71].

(1.2) **Definition.** The function $\dot{u}[\mathbf{V}] \in H^s(D_{\mathrm{ref}})$ is called the strong (weak) *material derivative* of $u \in H^s(D_{\mathrm{ref}})$ in the direction $\mathbf{V}$ if the strong (weak) limit

$$\dot{u}[\mathbf{V}] = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(u^\varepsilon - \overline{u})$$

exists.

The shape sensitivity analysis considered in this section is based on the notion of the *local shape derivative*. To this end, we consider for $\overline{u} \in H^s(D_{\mathrm{ref}})$ and $u_\varepsilon \in H^s(D_\varepsilon)$ the expression

$$\frac{1}{\varepsilon}(u_\varepsilon(\mathbf{x}) - \overline{u}(\mathbf{x})).$$

Obviously, this expression is only meaningful if $\mathbf{x} \in D_\varepsilon \cap D_{\mathrm{ref}}$. Nevertheless, according to [Epp07, Remark 2.2.12], there exists an $\varepsilon(\mathbf{x}, \mathbf{V})$ due to the regularity of T_ε such that $\mathbf{x} \in D_\varepsilon \cap D_{\mathrm{ref}}$ for all $0 \leqslant \varepsilon \leqslant \varepsilon(\mathbf{x}, \mathbf{V})$. Moreover, in order to define a meaningful functional analytic framework for the limit $\varepsilon \to 0$, one has to consider compact subsets $K \Subset D_{\mathrm{ref}}$, cf. [Sim80]. Hence, we have from [Epp07, Definition 2.2.13] the following

(1.3) **Definition.** For $K \Subset D_{\mathrm{ref}}$, the function $\delta u[\mathbf{V}] \in H^s(K)$ is called the strong (weak) *local* $H^s(K)$ shape derivative of u in direction $\mathbf{V}$, if the strong (weak) $H^s(K)$ limit

$$\delta u[\mathbf{V}] = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(u_\varepsilon - \overline{u})$$

exists. It holds $\delta u \in H^s_{\mathrm{loc}}(D_{\mathrm{ref}})$ strongly (weakly) if the limit exists for arbitrary $K \Subset D_{\mathrm{ref}}$.

Notice that the definition of $\delta u[\mathbf{V}]$ has no meaning on Γ_{ref} in general, cf. [Epp07, Remark 2.2.14]. Nevertheless, since boundary values for $\dot{u}[\mathbf{V}]$ are obtained via the trace operator, cf. [SZ92, Proposition 2.75], we may define the boundary values for $\delta u[\mathbf{V}]$ employing the relation

$$\dot{u}[\mathbf{V}] = \delta u[\mathbf{V}] + \langle \nabla u, \mathbf{V} \rangle,$$

cf. [SZ92, Definition 2.85]. Therefore, if $f \in H^1(\mathcal{D})$, the local shape derivative for the Poisson equation (1.1) satisfies the boundary value problem

$$\Delta \delta u = 0 \text{ in } D_{\mathrm{ref}}, \quad \delta u = -\langle \mathbf{V}, \mathbf{n} \rangle \frac{\partial \overline{u}}{\partial \mathbf{n}} \text{ on } \Gamma_{\mathrm{ref}},$$

cf. [SZ92, Proposition 3.1]. Here, $\mathbf{n}(\mathbf{x})$ denotes the outward normal at the boundary Γ_{ref}. This representation of $\delta u[\mathbf{V}]$ indicates that it is already sufficient to consider vector fields $\mathbf{V}$ which are compactly supported in a neighbourhood of Γ_{ref}, i.e. $\mathbf{V}|_K \equiv 0$ for all

$K \Subset D_{\mathrm{ref}}$, cf. [Epp07, Remark 2.1.6]. More precisely, it holds for two perturbation fields $\mathbf{V}$ and $\tilde{\mathbf{V}}$ that

$$\delta u[\mathbf{V}] = \delta u[\tilde{\mathbf{V}}] \quad \text{if } \mathbf{V}|_{\Gamma_{\mathrm{ref}}} = \tilde{\mathbf{V}}|_{\Gamma_{\mathrm{ref}}},$$

cf. [SZ92, Proposition 2.90]. For example, it is quite common to consider (normal) perturbations of the boundary, see e.g. [Epp07, Kir93, Pir82, Pot94a].

Having the local shape derivative of the solution u to (1.1) at hand, we can linearize the solution in a neighbourhood of D_{ref} in terms of a *shape Taylor expansion*, cf. [Epp00a, Epp00b, HSS08b, Pot94b], according to

$$u_\varepsilon(\mathbf{x}) = \overline{u}(\mathbf{x}) + \varepsilon \delta u(\mathbf{x}) + \varepsilon^2 C(\mathbf{x}) \quad \text{for } \mathbf{x} \in K \Subset (D_{\mathrm{ref}} \cap D_\varepsilon),$$

where the function $C(\mathbf{x}) > 0$ depends on the distance $\operatorname{dist}(K, \Gamma_{\mathrm{ref}})$ and the load f.

2. Approximation of mean and covariance

Now, we get back to our model problem, the Poisson equation on a random domain, i.e.

$$-\Delta u(\omega, \mathbf{x}) = f(\mathbf{x}) \text{ in } D(\omega), \quad u(\omega, \mathbf{x}) = 0 \text{ on } \Gamma(\omega). \tag{2.1}$$

Here, as in Chapter IV, the random domain is described by a random vector field. We have

$$D(\omega) := \mathbf{V}(\omega, D_{\mathrm{ref}}).$$

According to the discussion in the beginning of Section IV.3, it is reasonable to assume that $\mathbf{V}$ is a perturbation of identity. More precisely, we assume that

$$\mathbf{V}(\omega, \mathbf{x}) = \mathrm{Id}(\mathbf{x}) + \mathbf{V}_0(\omega, \mathbf{x})$$

for a vector field $\mathbf{V}_0(\omega) \in C^2(D_{\mathrm{ref}}; \mathbb{R}^d)$ for almost every $\omega \in \Omega$ with $\mathbb{E}[\mathbf{V}_0] = \mathbf{0}$. We shall assume here the uniformity condition $\|\mathbf{V}_0(\omega)\|_{C^2(\overline{D_{\mathrm{ref}}}; \mathbb{R}^d)} \leqslant c$ for some $c \in (0, \infty)$ and almost every $\omega \in \Omega$. Then, with respect to the transformation

$$T_\varepsilon(\mathbf{x}) = \mathrm{Id}(\mathbf{x}) + \varepsilon \mathbf{V}_0(\omega, \mathbf{x}), \tag{2.2}$$

the first-order shape Taylor expansion for the solution to the model problem (2.1) is given by

$$u(\omega, \mathbf{x}) = \overline{u}(\mathbf{x}) + \varepsilon \delta u(\mathbf{x})[\mathbf{V}_0(\omega)] + \mathcal{O}(\varepsilon^2).$$

In this expansion, $\overline{u}$ is the solution to

$$-\Delta \overline{u} = f \text{ in } D_{\mathrm{ref}}, \quad \overline{u} = 0 \text{ on } \Gamma_{\mathrm{ref}} \tag{2.3}$$

while $\delta u[\mathbf{V}_0(\omega)]$ is the solution to

$$\Delta \delta u[\mathbf{V}_0(\omega)] = 0 \text{ in } D_{\mathrm{ref}}, \quad \delta u[\mathbf{V}_0(\omega)] = -\langle \mathbf{V}_0(\omega), \mathbf{n} \rangle \frac{\partial \overline{u}}{\partial \mathbf{n}} \text{ on } \Gamma_{\mathrm{ref}}. \tag{2.4}$$

As already pointed out in the end of the preceding section, it is sufficient to know $\mathbf{V}_0(\omega, \mathbf{x})$ only in a neighbourhood of the boundary Γ_{ref} of D_{ref}. This is in contrast to the domain mapping method where we always have to know the perturbation field for the whole domain Γ_{ref}.

In order to simplify the notation, we will write $\delta u(\omega)$ instead of $\delta u[\mathbf{V}_0(\omega)]$ in the sequel. Having the first-order shape Taylor expansion (2.2) of $u(\omega)$ at hand, we can approximate the related moments from this.

(2.5) **Theorem.** For $\varepsilon > 0$ sufficiently small, it holds for $K \Subset (D_{\mathrm{ref}} \cap D_\varepsilon)$ that

$$\mathbb{E}[u] = \overline{u} + \mathcal{O}(\varepsilon^2) \quad \text{in } K \tag{2.6}$$

with $\overline{u} \in H^1_0(D_{\mathrm{ref}})$. The covariance of u satisfies

$$\mathrm{Cov}[u] = \varepsilon^2 \,\mathrm{Cov}[\delta u] + \mathcal{O}(\varepsilon^3) \quad \text{in } K \times K \tag{2.7}$$

with the covariance $\mathrm{Cov}[\delta u] \in H^1(D_{\mathrm{ref}}) \otimes H^1(D_{\mathrm{ref}})$. The covariance is given as the solution to the following boundary value problem

$$\begin{aligned}
(\Delta \otimes \Delta)\,\mathrm{Cov}[\delta u] &= 0 \quad \text{in } D_{\mathrm{ref}} \times D_{\mathrm{ref}},\\
(\Delta \otimes \gamma_0^{\mathrm{int}})\,\mathrm{Cov}[\delta u] &= 0 \quad \text{in } D_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}},\\
(\gamma_0^{\mathrm{int}} \otimes \Delta)\,\mathrm{Cov}[\delta u] &= 0 \quad \text{in } \Gamma_{\mathrm{ref}} \times D_{\mathrm{ref}},\\
(\gamma_0^{\mathrm{int}} \otimes \gamma_0^{\mathrm{int}})\,\mathrm{Cov}[\delta u] &= \langle \mathbf{n}(\mathbf{x}), \mathrm{Cov}[\mathbf{V}]\mathbf{n}(\mathbf{y})\rangle \left(\frac{\partial \overline{u}}{\partial \mathbf{n}} \otimes \frac{\partial \overline{u}}{\partial \mathbf{n}}\right) \quad \text{on } \Gamma_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}}.
\end{aligned} \tag{2.8}$$

Here, $\gamma_0^{\mathrm{int}} \colon H^1(D_{\mathrm{ref}}) \to H^{1/2}(\Gamma_{\mathrm{ref}})$ denotes the trace operator.

Proof. The equation for the mean is easily obtained by exploiting the linearity of the mean. It remains to show that

$$\mathbb{E}[\delta u] = 0.$$

By Theorem (II.2.6), we know that we may interchange the Bochner integral with the Laplace operator. Thus, we obtain the following boundary value problem for $\mathbb{E}[\delta u]$. It holds

$$\Delta \mathbb{E}[\delta u] = 0 \text{ in } D_{\mathrm{ref}}, \quad \mathbb{E}[\delta u] = -\mathbb{E}\left[\langle \mathbf{V}_0, \mathbf{n}\rangle \frac{\partial \overline{u}}{\partial \mathbf{n}}\right] \text{ on } \Gamma_{\mathrm{ref}}.$$

By the linearity of the Bochner integral, the boundary condition can be written as

$$-\mathbb{E}\left[\langle \mathbf{V}_0, \mathbf{n}\rangle \frac{\partial \overline{u}}{\partial \mathbf{n}}\right] = \langle \mathbb{E}[\mathbf{V}_0], \mathbf{n}\rangle \frac{\partial \overline{u}}{\partial \mathbf{n}} = 0,$$

since $\mathbb{E}[\mathbf{V}_0] = \mathbf{0}$. Thus, $\mathbb{E}[\delta u]$ solves the Laplace equation with homogeneous boundary condition. From this, we infer $\mathbb{E}[\delta u] = 0$.

For the covariance $\mathrm{Cov}[u]$, we obtain

$$\begin{aligned}
\mathrm{Cov}[u] &= \mathbb{E}\big[(u - \mathbb{E}[u]) \otimes (u - \mathbb{E}[u])\big]\\
&= \mathbb{E}\Big[\big(\overline{u} + \varepsilon \delta u(\omega) + \mathcal{O}(\varepsilon^2) - \mathbb{E}[u]\big) \otimes \big(\overline{u} + \varepsilon \delta u(\omega) + \mathcal{O}(\varepsilon^2) - \mathbb{E}[u]\big)\Big].
\end{aligned}$$

Since we can estimate $\mathbb{E}[u] - \overline{u} = \mathcal{O}(\varepsilon^2)$ in K due to (2.6), we arrive at

$$\begin{aligned}\mathrm{Cov}[u] &= \mathbb{E}\Big[\big(\varepsilon\delta u(\omega) + \mathcal{O}(\varepsilon^2)\big) \otimes \big(\varepsilon\delta u(\omega) + \mathcal{O}(\varepsilon^2)\big)\Big] \\ &= \varepsilon^2 \mathbb{E}[\delta u(\omega) \otimes \delta u(\omega)] + \mathcal{O}(\varepsilon^3).\end{aligned}$$

In view of $\mathrm{Cov}[\delta u] = \mathbb{E}[\delta u(\omega) \otimes \delta u(\omega)]$, we conclude (2.7). Finally, by tensorization of (2.4) and application of the mean together with Theorem (II.2.6), one infers that $\mathrm{Cov}[\delta u] \in H^1(D_{\mathrm{ref}}) \otimes H^1(D_{\mathrm{ref}})$ is given by (2.8). □

In the sequel, for $t \geqslant 0$, we set

$$H^t_{\mathrm{mix}}(D_{\mathrm{ref}} \times D_{\mathrm{ref}}) := H^t(D_{\mathrm{ref}}) \otimes H^t(D_{\mathrm{ref}}),$$

$$H^t_{\mathrm{mix}}(\Gamma_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}}) := H^t(\Gamma_{\mathrm{ref}}) \otimes H^t(\Gamma_{\mathrm{ref}}).$$

(2.9) **Remark.** The technique for the approximation error of the covariance of u can be straightforwardly applied to obtain a similar result for the k-th moment, i.e.

$$\mathbb{E}\big[\underbrace{(u - \mathbb{E}[u]) \otimes \ldots \otimes (u - \mathbb{E}[u])}_{k\text{-times}}\big].$$

In this case, we end up with the expression

$$\mathbb{E}\big[(\varepsilon\delta u + \mathcal{O}(\varepsilon^2)) \otimes \ldots \otimes (\varepsilon\delta u + \mathcal{O}(\varepsilon^2))\big] = \varepsilon^k \mathbb{E}[\delta u \otimes \ldots \otimes \delta u] + \mathcal{O}(\varepsilon^{k+1}),$$

where the constant depends exponentially on k, cf. also [CS13].

3. Discretization of tensor product Dirichlet problems

In the previous section, we have seen that we end up solving the tensor product Dirichlet problem (2.8) in order to approximate the covariance of the model problem's solution. The treatment of the non-homogenous tensor product Dirichlet boundary conditions is not trivial. Therefore, we think that it is justified to consider here the discretization by finite elements in detail. We start with the discretization of univariate Dirichlet problems and then generalize the approach towards the tensor product case.

At first, we aim at solving the Dirichlet boundary value problem

(3.1) $$\Delta u = 0 \text{ in } D_{\mathrm{ref}}, \quad u = g \text{ on } \Gamma_{\mathrm{ref}}.$$

By the inverse trace theorem, see e.g. [Wlo82], there exists an extension $u_g \in H^1(D_{\mathrm{ref}})$ with $\gamma_0^{\mathrm{int}} u_g = g$ provided that $g \in H^{1/2}(\Gamma_{\mathrm{ref}})$. Therefore, it remains to determine the function $u_0 = u - u_g \in H^1_0(D_{\mathrm{ref}})$ such that there holds

(3.2) $$a^D(u_0, v) = -a^D(u_g, v) \quad \text{for all } v \in H^1_0(D_{\mathrm{ref}}).$$

Here and in the sequel, the elliptic bilinear form related to the Laplace operator is given by

$$a^D(u, v) := (\nabla u, \nabla v)_{L^2(D_{\mathrm{ref}})} \quad \text{for } u, v \in H^1_0(D_{\mathrm{ref}}).$$

The question arises how to numerically determine a suitable extension u_g of the Dirichlet data. We follow here the approach from [BK94], see also [FGP83], where the extension is generated by means of an L^2-projection of the given boundary data. To that end, we distinguish between basis functions $\{\varphi_{j,k} \in V_j^s : k \in \mathcal{I}_j^D\}$ which are supported in the interior of the reference domain, i.e. $\varphi_{j,k}|_{\Gamma_{\mathrm{ref}}} \equiv 0$, and boundary functions $\{\varphi_{j,k} \in V_j^s : k \in \mathcal{I}_j^\Gamma\}$ with $\varphi_{j,k}|_{\Gamma_{\mathrm{ref}}} \not\equiv 0$. Notice that $\mathcal{I}_j = \mathcal{I}_j^D \cup \mathcal{I}_j^\Gamma$ and $\mathcal{I}_j^D \cap \mathcal{I}_j^\Gamma = \varnothing$. The related finite element spaces are then given by

$$V_j^D := \operatorname{span}\{\varphi_{j,k} \in V_j^s : k \in \mathcal{I}_j^D\} \quad \text{and} \quad V_j^\Gamma := \operatorname{span}\{\varphi_{j,k}|_{\Gamma_{\mathrm{ref}}} : \varphi_{j,k} \in V_j^s, k \in \mathcal{I}_j^\Gamma\}.$$

Moreover, we denote the L^2-inner product on Γ_{ref} by

$$a^\Gamma(u,v) := (u,v)_{L^2(\Gamma_{\mathrm{ref}})} \quad \text{for } u, v \in L^2(\Gamma_{\mathrm{ref}}).$$

Then, the L^2-orthogonal projection of the Dirichlet data is given by the solution to the following variational formulation:

$$\begin{aligned} &\text{Find } g_j \in V_j^\Gamma \text{ such that} \\ &\qquad a^\Gamma(g_j, v) = a^\Gamma(g, v) \quad \text{for all } v \in V_j^\Gamma. \end{aligned} \tag{3.3}$$

We are now prepared to formulate the Galerkin discretization of (3.2). To that end, we introduce the stiffness matrices

$$\mathbf{S}_j^\Lambda := \left[a^D(\varphi_{j,\ell}, \varphi_{j,k})\right]_{k \in \mathcal{I}_j^D,\, \ell \in \mathcal{I}_j^\Lambda}, \quad \Lambda \in \{D, \Gamma\} \tag{3.4}$$

and the mass matrices with respect to the boundary

$$\mathbf{G}_j := \left[a^\Gamma(\varphi_{j,\ell}, \varphi_{j,k})\right]_{k \in \mathcal{I}_j^\Gamma,\, \ell \in \mathcal{I}_{j'}^\Gamma}. \tag{3.5}$$

The related data vector is given by

$$\mathbf{g}_j = \left[a^\Gamma(g, \varphi_{j,k})\right]_{k \in \mathcal{I}_j^\Gamma}.$$

In order to compute an approximate solution to this boundary value problem in the finite element space $V_J^s \subset H^1(D_{\mathrm{ref}})$ for $J \in \mathbb{N}$, we make the ansatz

$$u_J = \sum_{k \in \mathcal{I}_J} u_{J,k}\varphi_{J,k} = \sum_{k \in \mathcal{I}_J^D} u_{J,k}\varphi_{J,k} + \sum_{k \in \mathcal{I}_J^\Gamma} u_{J,k}\varphi_{J,k} = u_J^D + u_J^\Gamma.$$

We determine first the boundary part $u_J^\Gamma \in H^1(D)$ such that

$$\mathbf{G}_J \mathbf{u}_J^\Gamma = \mathbf{g}_J. \tag{3.6}$$

Thus, $u_J^\Gamma|_{\Gamma_{\mathrm{ref}}}$ is the L^2-orthogonal projection of the Dirichlet datum g onto the discrete trace space V_J^Γ. Having u_J^Γ at hand, we can compute the domain part $u_J^D \in H_0^1(D)$ from

$$\mathbf{S}_J^D \mathbf{u}_J^D = -\mathbf{S}_J^\Gamma \mathbf{u}_J^\Gamma. \tag{3.7}$$

We use the conjugate gradient method to iteratively solve the systems of linear equations (3.6) and (3.7). Using a nested iteration, combined with the BPX-preconditioner [BPX90] in case of (3.7), results in a linear over-all complexity [Hac85]. Moreover, from [BK94, Theorem 1], we obtain the following convergence result.

(3.8) **Theorem.** Let $g \in H^t(\Gamma_{\mathrm{ref}})$ for $0 \leqslant t \leqslant 3/2$. Then, if $g_J \in V_J^\Gamma$ is given by (3.3), the Galerkin solution u_J to (3.1) satisfies

$$\|u - u_J\|_{L^2(D_{\mathrm{ref}})} \lesssim 2^{-J(t+1/2)} \|g\|_{H^t(\Gamma_{\mathrm{ref}})}.$$

Proof. For a proof of this result, see [BK94]. □

Next, we consider the tensor product boundary value problem (2.8) and discretize it in $V_J^s \otimes V_J^s$. Here we make the ansatz

$$\begin{aligned} \mathrm{Cov}[\delta u]_J &= \sum_{k \in \mathcal{I}_J} \sum_{k' \in \mathcal{I}_J} u_{J,k,k'} (\varphi_{J,k} \otimes \varphi_{J,k'}) \\ &= \mathrm{Cov}[\delta u]_J^{D,D} + \mathrm{Cov}[\delta u]_J^{D,\Gamma} + \mathrm{Cov}[\delta u]_J^{\Gamma,D} + \mathrm{Cov}[\delta u]_J^{\Gamma,\Gamma} \end{aligned} \tag{3.9}$$

with

$$\mathrm{Cov}[\delta u]_J^{\Lambda,\Lambda'} = \sum_{k \in \mathcal{I}_J^\Lambda} \sum_{k \in \mathcal{I}_J^{\Lambda'}} u_{J,k,k'} (\varphi_{J,k} \otimes \varphi_{J,k'}) \quad \text{for } \Lambda, \Lambda' \in \{D, \Gamma\}.$$

In complete analogy to the non-tensor product case, we obtain the solution scheme

(1) Solve $(\mathbf{G}_J \otimes \mathbf{G}_J)\mathbf{u}_J^{\Gamma,\Gamma} = \mathbf{g}_J$.

(2) Solve $(\mathbf{G}_J \otimes \mathbf{S}_J^D)\mathbf{u}_J^{\Gamma,D} = -(\mathbf{G}_J \otimes \mathbf{S}_J^\Gamma)\mathbf{u}_J^{\Gamma,\Gamma}$ and $(\mathbf{S}_J^D \otimes \mathbf{G}_J)\mathbf{u}_J^{D,\Gamma} = -(\mathbf{S}_J^\Gamma \otimes \mathbf{G}_J)\mathbf{u}_J^{\Gamma,\Gamma}$.

(3) Solve $(\mathbf{S}_J^D \otimes \mathbf{S}_J^D)\mathbf{u}_J^{D,D} = -(\mathbf{S}_J^\Gamma \otimes \mathbf{S}_J^\Gamma)\mathbf{u}_J^{\Gamma,\Gamma} - (\mathbf{S}_\Gamma \otimes \mathbf{S}_J^D)\mathbf{u}_J^{\Gamma,D} - (\mathbf{S}_D \otimes \mathbf{S}_J^\Gamma)\mathbf{u}_J^{D,\Gamma}$.

Herein, we set $\mathbf{u}_J^{\Lambda,\Lambda'} = [u_{J,k,k'}]_{k \in \mathcal{I}_J^\Lambda, k' \in \mathcal{I}_J^{\Lambda'}}$ for $\Lambda, \Lambda' \in \{D, \Gamma\}$ and

$$\mathbf{g}_J = \left[\left(\langle \mathbf{n}(\mathbf{x}), \mathrm{Cov}[\mathbf{V}]\mathbf{n}(\mathbf{y}) \rangle \left(\frac{\partial \overline{u}}{\partial \mathbf{n}} \otimes \frac{\partial \overline{u}}{\partial \mathbf{n}} \right), \varphi_{J,k} \otimes \varphi_{J,k'} \right)_{L^2(\Gamma_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}})} \right]_{k,k' \in \mathcal{I}_J^\Gamma}.$$

The different tensor products of mass matrices and stiffness matrices in this formulation arise from the related tensor products of the bilinear forms $a^D(\cdot,\cdot)$ and $a^\Gamma(\cdot,\cdot)$. Namely, these are

$$\begin{aligned} a^{\Gamma,\Gamma}(u,v) &:= (u,v)_{L^2(\Gamma_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}})} && \text{for } u,v \in L^2(\Gamma_{\mathrm{ref}}) \otimes L^2(\Gamma_{\mathrm{ref}}), \\ a^{\Gamma,D}(u,v) &:= ((\mathrm{Id} \otimes \nabla)u, (\mathrm{Id} \otimes \nabla)v)_{L^2(\Gamma_{\mathrm{ref}} \times D_{\mathrm{ref}})} && \text{for } u,v \in L^2(\Gamma_{\mathrm{ref}}) \otimes H_0^1(D_{\mathrm{ref}}), \\ a^{D,\Gamma}(u,v) &:= ((\nabla \otimes \mathrm{Id})u, (\nabla \otimes \mathrm{Id})v)_{L^2(D_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}})} && \text{for } u,v \in H_0^1(D_{\mathrm{ref}}) \otimes L^2(\Gamma_{\mathrm{ref}}), \\ a^{D,D}(u,v) &:= ((\nabla \otimes \nabla)u, (\nabla \otimes \nabla)v)_{L^2(D_{\mathrm{ref}} \times D_{\mathrm{ref}})} && \text{for } u,v \in H_0^1(D_{\mathrm{ref}}) \otimes H_0^1(D_{\mathrm{ref}}). \end{aligned} \tag{3.10}$$

For the approximation error of the Galerkin solution in $V_J^s \otimes V_J^s$, there holds a result similar to Theorem (3.8).

(3.11) **Theorem.** Let $g \in H^t_{\mathrm{mix}}(\Gamma_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}})$ for $0 \leqslant t \leqslant 3/2$. Then, if $g_J \in V_J^\Gamma \otimes V_J^\Gamma$ is the L^2-orthogonal projection of the Dirichlet data, the Galerkin solution u_J to the tensor product Dirichlet problem satisfies

$$\|u - u_J\|_{L^2(D_{\mathrm{ref}} \times D_{\mathrm{ref}})} \lesssim 2^{-J(t+1/2)} \|g\|_{H^t_{\mathrm{mix}}(\Gamma_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}})}.$$

Proof. By a tensor product argument, the proof of this theorem is obtained by summing up the uni-directional error estimates provided by Theorem (3.8). □

Unfortunately, the computational complexity of this approximation is of order $(\dim V_J^s)^2$, which may become prohibitive for increasing level J. A possibility to overcome this obstruction is given by the discretization in *sparse tensor product spaces*. In the following we shall focus on this approach.

4. Sparse second moment analysis

According to Section 2, to leading order, the mean of the solution of the random boundary value problem (2.1) satisfies the deterministic equation (2.3). This equation can be discretized straightforwardly by means of finite elements. The resulting system of linear equations may then be solved in optimal complexity e.g. by a multigrid solver. The solution of the tensor product structured problem (2.8) is a little more involved and requires another approach in order to maintain the overall complexity.

Instead of discretizing the tensor product boundary value problem (2.8) in the space $V_J^s \otimes V_J^s$, we consider here the discretization in the *sparse tensor product space*

$$\widehat{V_J^s \otimes V_J^s} := \sum_{j+j' \leqslant J} V_j^s \otimes V_{j'}^s = \sum_{j+j'=J} V_j^s \otimes V_{j'}^s \subset H^1_{\mathrm{mix}}(D_{\mathrm{ref}} \times D_{\mathrm{ref}}). \tag{4.1}$$

For the dimension of the sparse tensor product space, we have

$$\dim \widehat{V_J^s \otimes V_J^s} \eqsim \dim V_J^s \log(\dim V_J^s)$$

instead of $(\dim V_J^s)^2$, which is the dimension of $V_J^s \otimes V_J^s$, cf. [BG04]. Thus, the dimension of the sparse tensor product space is substantially smaller than that of the full tensor product space.

The following lemma, proven in [vPS06, ST03b], tells us that the approximation power in the sparse tensor product space is nearly as good as in the full tensor product space, provided that the given function has some extra regularity in terms of bounded mixed derivatives.

(4.2) **Lemma.** For $0 \leq t < 3/2$, $t \leq q \leq s+1$ there holds the error estimate

$$\inf_{\widehat{v}_J \in \widehat{V_J^s \otimes V_J^s}} \|v - \widehat{v}_J\|_{H^t_{\mathrm{mix}}(D_{\mathrm{ref}} \times D_{\mathrm{ref}})} \lesssim \begin{cases} 2^{J(q-t)} \sqrt{J} \|v\|_{H^q_{\mathrm{mix}}(D_{\mathrm{ref}} \times D_{\mathrm{ref}})}, & \text{if } q = s+1, \\ 2^{J(q-t)} \|v\|_{H^q_{\mathrm{mix}}(D_{\mathrm{ref}} \times D_{\mathrm{ref}})}, & \text{otherwise,} \end{cases}$$

provided that $v \in H^q_{\mathrm{mix}}(D_{\mathrm{ref}} \times D_{\mathrm{ref}})$.

This lemma gives rise to an estimate for the Galerkin approximation $\widehat{\mathrm{Cor}[\delta u]}_J$ of (2.8) in the sparse tensor product space $\widehat{V_J^s \otimes V_J^s}$, see e.g. [Har10b, Proposition 5]. We state it here for the case $s = 1$, i.e. pice-wise linear finite elements, which will also be considered in the numerical examples.

(4.3) **Corollary.** The Galerkin approximate $\widehat{\operatorname{Cor}[\delta u]}_J \in \widehat{V_J^1 \otimes V_J^1}$ to (2.8) satisfies the error estimate

$$\|\operatorname{Cor}[\delta] - \widehat{\operatorname{Cor}[\delta u]}_J\|_{L^2(D_{\mathrm{ref}}\times D_{\mathrm{ref}})} \lesssim 2^{-2J} J \|\operatorname{Cor}[\delta u]\|_{H^2_{\mathrm{mix}}(D_{\mathrm{ref}}\times D_{\mathrm{ref}})}$$

provided that the given data are sufficiently smooth.

The Galerkin discretization of (2.8) in the sparse tensor product space is now rather similar to the approach in [Har10b], where *sparse multilevel frames*, cf. [HSS08a], have been employed for the discretization. We can considerably improve this approach by combining it with the idea from [HPS13a]: Instead of dealing with all combinations which occur in the discretization by a sparse frame for each of the four subproblems on $\Gamma_{\mathrm{ref}}\times\Gamma_{\mathrm{ref}}$, on $D_{\mathrm{ref}}\times\Gamma_{\mathrm{ref}}$, on $\Gamma_{\mathrm{ref}}\times D_{\mathrm{ref}}$ and in $D_{\mathrm{ref}}\times D_{\mathrm{ref}}$, we employ here the *combination technique*, cf. [GSZ92]. In this situation, we have only to compute combinations of the solution on two consecutive levels instead of all combinations.

The analogue to the ansatz (3.9) for the Galerkin approximation in the sparse tensor product space reads

$$\begin{aligned}\widehat{\operatorname{Cov}[\delta u]}_J &= \sum_{j+j'\leqslant J}\sum_{k\in\mathcal{I}_j}\sum_{k'\in\mathcal{I}_{j'}} \widehat{u}_{j,j',k,k'}(\varphi_{j,k}\otimes\varphi_{j',k'}) \\ &= \widehat{\operatorname{Cov}[\delta u]}_J^{D,D} + \widehat{\operatorname{Cov}[\delta u]}_J^{D,\Gamma} + \widehat{\operatorname{Cov}[\delta u]}_J^{\Gamma,D} + \widehat{\operatorname{Cov}[\delta u]}_J^{\Gamma,\Gamma}\end{aligned} \tag{4.4}$$

with

$$\widehat{\operatorname{Cov}[\delta u]}_J^{\Lambda,\Lambda'} = \sum_{j+j'\leqslant J}\sum_{k\in\mathcal{I}_J^\Lambda}\sum_{k\in\mathcal{I}_J^{\Lambda'}} \widehat{u}_{j,j',k,k'}(\varphi_{j,k}\otimes\varphi_{j',k'}) \in \widehat{V_J^\Lambda \otimes V_J^{\Lambda'}} \text{ for } \Lambda,\Lambda' \in \{D,\Gamma\}. \tag{4.5}$$

The basic idea of our approach is to define *detail spaces* with respect to Galerkin projections and remove by this the redundancy in the ansatz for the subproblems (4.5). We need thus the Galerkin projection $P_j\colon H_0^1(D_{\mathrm{ref}}) \to V_j^D$, $w \mapsto P_j w$ defined via

$$(\nabla(w - P_j w), \nabla v_j,)_{L^2(D_{\mathrm{ref}})} = 0 \quad \text{for all } v_j \in V_j^D$$

and the L^2-orthogonal projection $Q_j\colon L^2(\Gamma_{\mathrm{ref}}) \to V_j^\Gamma$, $w \mapsto Q_j w$ defined via

$$((w - Q_j w), v_j,)_{L^2(\Gamma_{\mathrm{ref}})} = 0 \quad \text{for all } v_j \in V_j^\Gamma.$$

Furthermore, we introduce the related *detail projections*

$$\Theta_j^D := P_j - P_{j-1}, \text{ with } P_{-1} := 0$$

and

$$\Theta_j^\Gamma := Q_j - Q_{j-1}, \text{ with } Q_{-1} := 0.$$

With the detail projections at hand, we define the related *detail spaces*

$$W_j^D := \Theta_j^D H_0^1(D_{\mathrm{ref}}) = (P_j - P_{j-1})H_0^1(D_{\mathrm{ref}}) \subset V_j^D$$

and

$$W_j^\Gamma := \Theta_j^\Gamma L^2(\Gamma) = (Q_j - Q_{j-1})L^2(\Gamma) \subset V_j^\Gamma.$$

Obviously, it holds $V_j = V_{j-1}^\Lambda \oplus W_j^\Lambda$ for $\Lambda \in \{D, \Gamma\}$. This gives rise to the decompositions

$$V_J^\Lambda = W_0^\Lambda \oplus W_1^\Lambda \oplus \ldots \oplus W_J^\Lambda \text{ for } \Lambda \in \{D, \Gamma\}.$$

Especially, these decompositions are orthogonal with respect to their defining inner products.

(4.6) **Lemma.** It holds

$$(\nabla w_j, \nabla w_{j'})_{L^2(D_{\mathrm{ref}})} = 0 \quad \text{for } w_j \in W_j^D,\ w_{j'} \in W_{j'}^D \text{ and } j' \neq j$$

as well as

$$(w_j, w_{j'})_{L^2(D_{\mathrm{ref}})} = 0 \quad \text{for } w_j \in W_j^\Gamma,\ w_{j'} \in W_{j'}^\Gamma \text{ and } j \neq j'.$$

Proof. We show the assertion for the spaces W_j^D. The proof for the spaces W_j^Γ is analogous. Without loss of generality, let $j > j'$. Otherwise, due to the symmetry of the inner products, we may interchange the roles of j and j'. Let $w_j = \Theta_j v \in W_j^D$ for some $v \in H_0^1(D_{\mathrm{ref}})$ and $w_{j'} \in W_{j'}^D \subset V_{j'}^D$. Then, we have since $j - 1 \geqslant j'$ that

$$(\nabla P_j v, \nabla w_{j'})_{L^2(D_{\mathrm{ref}})} = (\nabla v, \nabla w_{j'})_{L^2(D_{\mathrm{ref}})}$$

and

$$(\nabla P_{j-1} v, \nabla w_{j'})_{L^2(D_{\mathrm{ref}})} = (\nabla v, \nabla w_{j'})_{L^2(D_{\mathrm{ref}})}.$$

Thus, we obtain

$$\begin{aligned}(\nabla w_j, \nabla w_{j'})_{L^2(D_{\mathrm{ref}})} &= (\nabla P_j v, \nabla w_{j'})_{L^2(D_{\mathrm{ref}})} - (\nabla P_{j-1} v, \nabla w_{j'})_{L^2(D_{\mathrm{ref}})} \\ &= (\nabla v, \nabla w_{j'})_{L^2(D_{\mathrm{ref}})} - (\nabla v, \nabla w_{j'})_{L^2(D_{\mathrm{ref}})} = 0.\end{aligned}$$

□

Galerkin discretization of the second moment

Now, we shall rewrite the sparse tensor product spaces given by (4.1) according to

$$\widehat{V_J^\Lambda \otimes V_J^{\Lambda'}} = \sum_{j+j'=J} V_j^\Lambda \otimes V_{j'}^{\Lambda'} = \sum_{j+j'=J} \left(\bigoplus_{i=0}^{j} W_i^\Lambda \right) \otimes V_{j'}^{\Lambda'} = \bigoplus_{j=0}^{J} W_j^\Lambda \otimes V_{J-j}^{\Lambda'}.$$

By exploiting the symmetry in this expression, we have also

$$\widehat{V_J^\Lambda \otimes V_J^{\Lambda'}} = \bigoplus_{j=0}^{J} W_j^\Lambda \otimes V_{J-j}^{\Lambda'} = \bigoplus_{j=0}^{J} V_j^\Lambda \otimes W_{J-j}^{\Lambda'}.$$

Thus, fixing a basis $\psi_{j,k} \in W_j^\Lambda$ for $\Lambda \in \{D, \Gamma\}$, we have for the subproblems (4.5) the formulation

$$\widehat{\mathrm{Cov}[\delta u]}_J^{\Lambda,\Lambda'} = \bigoplus_{j=0}^{J} \sum_{k \in \mathcal{I}_J^\Lambda} \sum_{k \in \mathcal{I}_{J-j}^{\Lambda'}} \widehat{u}_{j,J-j,k,k'} (\psi_{j,k} \otimes \varphi_{J-j,k'}) \quad \text{for } \Lambda, \Lambda' \in \{D, \Gamma\}. \tag{4.7}$$

Taking further into account the orthogonality described by Lemma (4.6), we can show that the computation of $\widehat{\mathrm{Cov}[\delta u]}_J^{\Lambda,\Lambda'}$ for $\Lambda, \Lambda' \in \{D, \Gamma\}$ yields to decoupled subproblems.

(4.8) **Lemma.** Let $\Lambda, \Lambda' \in \{D, \Gamma\}$. For $\widehat{v}_j \in W_j^\Lambda \otimes V_{J-j}^{\Lambda'}$ and $\widehat{v}_{j'} \in W_{j'}^\Lambda \otimes V_{J-j'}^{\Lambda'}$, there holds

$$a^{\Lambda,\Lambda'}(\widehat{v}_j, \widehat{v}_{j'}) = 0 \quad \text{if } j \neq j'.$$

Proof. We show the proof for the case $\Lambda = \Gamma$ and $\Lambda' = D$. The other cases are analogous, see also [HPS13a, Lemma 6]. Assume that

$$\widehat{v}_j = \sum_{i \in \mathcal{I}} \alpha_i (\psi_{j,i} \otimes \varphi_{J-j,i}) \quad \text{and} \quad \widehat{v}_{j'} = \sum_{i \in \mathcal{I}'} \beta_i (\psi_{j',i} \otimes \varphi_{J-j',i})$$

are representations of $\widehat{v}_j \in W_j^\Gamma \otimes V_{J-j}^D$ and $\widehat{v}_{j'} \in W_{j'}^\Gamma \otimes V_{J-j'}^D$, respectively, for some finite index sets $\mathcal{I}, \mathcal{I}' \subset \mathbb{N}$. Then, we obtain

$$\begin{aligned} &a^{\Gamma,D}(\widehat{v}_j, \widehat{v}_{j'}) \\ &\quad = \Big((\mathrm{Id} \otimes \nabla) \sum_{i \in \mathcal{I}} \alpha_i (\psi_{j,i} \otimes \varphi_{J-j,i}), (\mathrm{Id} \otimes \nabla) \sum_{i' \in \mathcal{I}'} \beta_i (\psi_{j',i'} \otimes \varphi_{J-j',i'}) \Big)_{L^2(\Gamma_{\mathrm{ref}} \times D_{\mathrm{ref}})} \\ &\quad = \sum_{i \in \mathcal{I}} \sum_{i' \in \mathcal{I}'} \alpha_i \beta_{i'} (\psi_{j,i}, \psi_{j',i'})_{L^2(\Gamma_{\mathrm{ref}})} (\nabla \varphi_{J-j,i}, \nabla \varphi_{J-j',i'})_{L^2(D_{\mathrm{ref}})} = 0 \end{aligned}$$

whenever $j \neq j'$ due to Lemma (4.6). □

This lemma tells us that, given $\widehat{\mathrm{Cov}[\delta u]}_J^{\Gamma,\Gamma}$, the computation of $\widehat{\mathrm{Cov}[\delta u]}_J^{\Lambda,\Lambda'}$ for $\Lambda, \Lambda' \in \{D, \Gamma\}$ decouples into $J+1$ subproblems. It holds

$$\widehat{\mathrm{Cov}[\delta u]}_J^{\Lambda,\Lambda'} = \sum_{j=0}^{J} \widehat{v}_j,$$

where $\widehat{v}_j \in W_j^\Lambda \otimes V_{J-j}^{\Lambda'}$ is the solution to the following Galerkin formulation:

$$\begin{aligned} &\text{Find } \widehat{v}_j \in W_j^\Lambda \otimes V_{J-j}^{\Lambda'} \text{ such that} \\ &\qquad a^{\Lambda,\Lambda'}(\widehat{v}_j, \widehat{w}) = \mathrm{rhs}^{\Lambda,\Lambda'}(\widehat{w}) \quad \text{for all } \widehat{w}_j \in W_j^\Lambda \otimes V_{J-j}^{\Lambda'}. \end{aligned}$$

Herein, we set

$$\mathrm{rhs}^{\Lambda,\Lambda'}(\widehat{w}) := \begin{cases} -a^{D,\Gamma}(\widehat{\mathrm{Cov}[\delta u]}_J^{\Gamma,\Gamma}, \widehat{w}), & \Lambda = D,\ \Lambda' = \Gamma, \\ -a^{\Gamma,D}(\widehat{\mathrm{Cov}[\delta u]}_J^{\Gamma,\Gamma}, \widehat{w}), & \Lambda = \Gamma,\ \Lambda' = D, \\ -a^{D,D}(\widehat{\mathrm{Cov}[\delta u]}_J^{D,\Gamma} + \widehat{\mathrm{Cov}[\delta u]}_J^{\Gamma,D} + \widehat{\mathrm{Cov}[\delta u]}_J^{\Gamma,\Gamma}, \widehat{w}), & \Lambda = D,\ \Lambda' = D. \end{cases} \tag{4.9}$$

By taking into account the definition of the detail spaces, we end up with the following representation of the solution to (2.8) in the sparse tensor product space, which is known as the combination technique.

(4.10) **Theorem.** Given $\widehat{\mathrm{Cov}[\delta u]}_J^{\Gamma,\Gamma}$, the computation of $\widehat{\mathrm{Cov}[\delta u]}_J^{\Lambda,\Lambda'}$ for $\Lambda, \Lambda' \in \{D, \Gamma\}$ decouples as follows. It holds

$$\widehat{\mathrm{Cov}[\delta u]}_J^{\Lambda,\Lambda'} = \sum_{j=0}^{J} p_{j,J-j} - p_{j-1,J-j}, \tag{4.11}$$

where $p_{j,J-j} \in V_j^\Lambda \otimes V_{J-j}^\Lambda$ and $p_{j-1,J-j} \in V_{j-1}^\Lambda \otimes V_{J-j}^{\Lambda'}$ satisfy the following subproblems which are defined relative to full tensor product spaces:

$$\begin{aligned}&\text{Find } p_{j,j'} \in V_j^\Lambda \otimes V_{j'}^{\Lambda'} \text{ such that}\\ &\quad a^{\Lambda,\Lambda'}(p_{j,j'}, q_{j,j'}) = \mathrm{rhs}^{\Lambda,\Lambda'}(q_{j,j'}) \quad \text{for all } q_{j,j'} \in V_j^\Lambda \otimes V_{j'}^{\Lambda'}.\end{aligned}$$

Here, the right hand side is given according to (4.9).

Proof. The proof of this theorem is a consequence of the previous lemma together with the definition of the detail spaces W_j^Λ for $\Lambda \in \{D, \Gamma\}$. □

Numerical implementation

Our numerical realization heavily relies on the sparse frame discretization of the model problem as presented in [Har10b]. Nevertheless, in contrast to this work, we make here use of the fact, that we already obtain a sparse tensor product representation of the solution if we have the representations in the spaces $V_j^s \otimes V_{J-j}^s$ and $V_{j-1}^s \otimes V_{J-j}^s$. This means that it is sufficient to compute the diagonal $(j, J-j)$ for $j = 0, \dots, J$ and the subdiagonal $(j, J-j-1)$ for $j = 0, \dots, J-1$ of a sparse frame representation. Moreover, each block in this representation corresponds to the solution of a tensor product subproblem as stated in Theorem (4.10). The corresponding right hand sides are obtained by means of the matrix-vector product in the frame representation. Therefore, in this context, the combination technique can be considered as an improved solver for the approach presented in [Har10b], which results in a remarkable speed-up. In the sequel, we describe this approach in detail.

We focus here on the discretization by piecewise linear finite elements and start by discretizing the Dirichlet data. The proceeding is as considered in [Har10a]. Setting $\mathcal{J}_0 := \mathcal{I}_0^\Gamma$ and $\mathcal{J}_j := \mathcal{I}_j^\Gamma \setminus \mathcal{I}_{j-1}^\Gamma$ for $j > 0$, the hierarchical basis in $\operatorname{span}\{\varphi_{j,k} \in V_j^s : k \in \mathcal{I}_j^\Gamma\}$ is given by $\bigcup_{j=0}^J \{\varphi_{j,k}\}_{k \in \mathcal{J}_j}$. We replace the normal part of the covariance by its piecewise linear sparse grid interpolant, cf. [BG04],

$$\langle \mathbf{n}(\mathbf{x}), \mathrm{Cov}[\mathbf{V}]\mathbf{n}(\mathbf{y})\rangle \approx \Bigg(\sum_{j+j' \le J} \sum_{k \in \mathcal{J}_j} \sum_{k' \in \mathcal{J}_{j'}} \gamma_{(j,k),(j',k')} (\varphi_{j,k} \otimes \varphi_{j',k'}) \Bigg)\Bigg|_{\Gamma_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}}}.$$

Thus, the coefficient vector $\mathbf{g}_{j,j'}$ of the Dirichlet data becomes

$$\mathbf{g}_{j,j'} = \sum_{\ell+\ell' \le J} (\mathbf{B}_{j,\ell} \otimes \mathbf{B}_{j',\ell'})[\gamma_{(\ell,k),(\ell',k')}]_{k \in \mathcal{J}_j, k' \in \mathcal{J}_{j'}}, \tag{4.12}$$

where the matrices $\mathbf{B}_{j,j'}$, $0 \le j, j' \le J$, are given by

$$\mathbf{B}_{j,j'} = \left[a^\Gamma \left(\frac{\partial \overline{u}}{\partial \mathbf{n}} \varphi_{j,k}, \varphi_{j',k'} \right) \right]_{k \in \mathcal{I}_j, k' \in \mathcal{J}_{j'}}.$$

The expression (4.12) can be evaluated in optimal complexity by applying the matrix-vector multiplication from [Zei11]. Nevertheless, for the sake of an easier implementation, we employ here the matrix-vector multiplication from [HSS08a], which is optimal

up to logarithmic factors. In particular, by using prolongations and restrictions, the matrices $\mathbf{B}_{j,j'}$ have to be provided only for the case $j = j'$. Thus, having all right hand sides at hand, we can solve now

$$(\mathbf{G}_j \otimes \mathbf{G}_{j'})\mathbf{p}_{j,j'}^{\Gamma,\Gamma} = \mathbf{g}_{j,j'},$$

for all indices satisfying $j' = J - j$ or $j' = J - j - 1$. With these coefficients, we determine the right hand sides for the problems on $D_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}}$ and $\Gamma_{\mathrm{ref}} \times D_{\mathrm{ref}}$ according to

$$\mathbf{f}_{j,j'}^{D,\Gamma} = -\sum_{\ell+\ell' \le J} (\mathbf{S}_{j,\ell}^{\Gamma} \otimes \mathbf{G}_{j',\ell'})\mathbf{p}_{\ell,\ell'}^{\Gamma,\Gamma} \quad \text{and} \quad \mathbf{f}_{j,j'}^{\Gamma,D} = -\sum_{\ell+\ell' \le J} (\mathbf{G}_{j,\ell} \otimes \mathbf{S}_{j',\ell'}^{\Gamma})\mathbf{p}_{\ell,\ell'}^{\Gamma,\Gamma},$$

where the matrices $\mathbf{S}_{j,j'}^{\Gamma}$, $\mathbf{G}_{j,j'}$, $0 \le j, j' \le J$, are given by

$$\mathbf{S}_{j,j'}^{\Gamma} = \big[a^D(\varphi_{j',\ell}, \varphi_{j,k})\big]_{k\in\mathcal{I}_j^D,\, \ell\in\mathcal{I}_{j'}^{\Gamma}} \quad \text{and} \quad \mathbf{G}_{j,j'} = \big[a^{\Gamma}(\varphi_{j',\ell}, \varphi_{j,k})\big]_{k\in\mathcal{I}_j^{\Gamma},\, \ell\in\mathcal{I}_{j'}^{\Gamma}}.$$

Notice that we have $\mathbf{S}_{j,j}^{\Gamma} = \mathbf{S}_j^{\Gamma}$ and $\mathbf{G}_{j,j} = \mathbf{G}_j$, cf. (3.4) and (3.5). Now, we can solve

$$(\mathbf{S}_j^D \otimes \mathbf{G}_{j'})\mathbf{p}_{j,j'}^{D,\Gamma} = \mathbf{f}_{j,j'}^{D,\Gamma} \quad \text{and} \quad (\mathbf{G}_j \otimes \mathbf{S}_{j'}^D)\mathbf{p}_{j,j'}^{\Gamma,D} = \mathbf{f}_{j,j'}^{\Gamma,D}$$

for all indices satisfying $j' = J - j$ or $j' = J - j - 1$.

From the solutions $\mathbf{p}_{j,j'}^{D,\Gamma}$ and $\mathbf{p}_{j,j'}^{\Gamma,D}$, we can finally determine the right hand sides

$$\mathbf{f}_{j,j'}^{D,D} = -\sum_{\ell+\ell' \le J} (\mathbf{S}_{j,\ell}^{\Gamma} \otimes \mathbf{S}_{j',\ell'}^{\Gamma})\mathbf{p}_{\ell,\ell'}^{\Gamma,\Gamma} + (\mathbf{S}_{j,\ell}^{D} \otimes \mathbf{S}_{j',\ell'}^{\Gamma})\mathbf{p}_{\ell,\ell'}^{D,\Gamma} + (\mathbf{S}_{j,\ell}^{\Gamma} \otimes \mathbf{S}_{j',\ell'}^{D})\mathbf{p}_{\ell,\ell'}^{\Gamma,D},$$

where the matrices $\mathbf{S}_{j,j'}^{D}$, $0 \le j, j' \le J$, are given by

$$\mathbf{S}_{j,j'}^{D} = \big[a^D(\varphi_{j',\ell}, \varphi_{j,k})\big]_{k\in\mathcal{I}_j^D,\, \ell\in\mathcal{I}_{j'}^{D}}.$$

It remains to compute the solutions to

$$(\mathbf{S}_j^D \otimes \mathbf{S}_{j'}^D)\mathbf{p}_{j,j'}^{D,D} = \mathbf{f}_{j,j'}^{D,D}$$

for all indices satisfying $j' = J - j$ or $j' = J - j - 1$.

Appropriate tensorization of the BPX-preconditioner [BPX90] yields an asymptotically optimal preconditioning for each of the preceding linear systems, cf. [HSS08a, Theorem 7]. Consequently, the computational complexity for their solution is linear, i.e. of the order $\mathcal{O}(2^{(j+j')d})$. Moreover, the right hand sides $\mathbf{f}_{j,j'}^{\Lambda,\Lambda'}$ for $\Lambda, \Lambda' \in \{D, \Gamma\}$ can be computed by the algorithm proposed in [Zei11] with an effort of $\mathcal{O}(J \log N_J)$. We thus obtain the following result:

(4.13) **Theorem.** The cost of computing the Galerkin solution $\widehat{\mathrm{Cor}}_{\delta u,J}$ via the expansion (4.11) is of optimal order $\mathcal{O}(N_J \log N_J)$.

Proof. For each $0 \le j \le J$ and $\Lambda, \Lambda' \in \{D, \Gamma\}$, the cost to determine $p_{j,J-j}^{\Lambda,\Lambda'}$ and $p_{j-1,J-j}^{\Lambda,\Lambda'}$ is of order $\mathcal{O}(N_J)$. Summing over j yields immediately the assertion due to $J \eqsim \log N_J$.□

5. Numerical results

To demonstrate the described method, we consider an analytical example on the one hand and a stochastic example on the other hand. There, for a given random domain perturbation described by the vector field $\mathbf{V}$, we compute the function $\overline{u}$ and the covariance $\mathrm{Cov}[\delta u]$, cf. (2.8). As in the previous numerical results, all computations are carried out on a computing server with two Intel(R) Xeon(R) X5550 CPUs with a clock rate of 2.67GHz and 48GB of main memory. The computations have been performed single-threaded, i.e. on a single core.

An analytical example

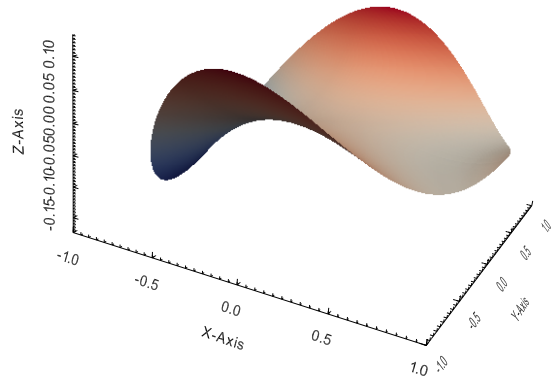

Figure V.1: Trace $u|_{\mathbf{x}=\mathbf{y}}$ of the solution u to (5.1).

In this analytical example, we want to validate the convergence rates of the combination technique for the sparse tensor product solution of tensor product Dirichlet problems. To that end, consider the tensor product boundary value problem

$$\begin{aligned}
(\Delta \otimes \Delta)u &= 0 && \text{in } D_{\mathrm{ref}} \times D_{\mathrm{ref}},\\
(\Delta \otimes \gamma_0^{\mathrm{int}})u &= 0 && \text{in } D_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}},\\
(\gamma_0^{\mathrm{int}} \otimes \Delta)u &= 0 && \text{in } \Gamma_{\mathrm{ref}} \times D_{\mathrm{ref}},\\
(\gamma_0^{\mathrm{int}} \otimes \gamma_0^{\mathrm{int}})u &= g_1 \otimes g_2 && \text{on } \Gamma_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}},
\end{aligned} \tag{5.1}$$

where $D_{\mathrm{ref}} = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 < 1\}$ is the two-dimensional unit disk. We choose g_1 and g_2 to be the traces of harmonic functions. More precisely, we set

$$g_1(\mathbf{x}) = x_1^2 - x_2^2 \quad \text{and} \quad g_2(\mathbf{x}) = -\frac{1}{2\pi} \log\left(\sqrt{(x_1-2)^2 + (x_2-2)^2}\right) \quad \text{for } \mathbf{x} \in \Gamma_{\mathrm{ref}}.$$

Then, the solution u is simply given by the product

$$u(\mathbf{x},\mathbf{y}) = -\frac{1}{2\pi}(x_1^2 - x_2^2) \log\left(\sqrt{(y_1-2)^2 + (y_2-2)^2}\right).$$

A visualization of the trace $u|_{\mathbf{x}=\mathbf{y}}$ of this function is depicted in Figure V.1.

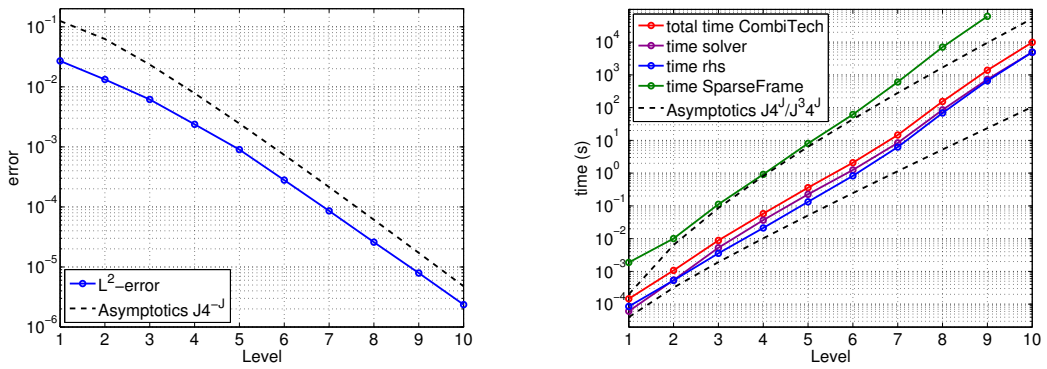

Figure V.2: Relative L^2-error (left) and computation times (right) of the combination technique in case of the analytic example.

The convergence plot on the left of Figure V.2 shows that the relative L^2-error, indicated by the blue line, exhibits almost the convergence rate predicted in Corollary (4.3), indicated by the black dashed line. On level 10, there are about 2.1 million degrees of freedom in each variable, which is, up to a logarithmic factor, the number of degrees of freedom appearing in the discretization by the combination technique. Vice versa, a full tensor product discretization on this level would result in about $4.4 \cdot 10^{12}$ degrees of freedom, which is no more feasible.

The plot on the right hand side of Figure V.2 depicts the related computational times. For comparison, we have added here the computational times for the sparse tensor product frame discretization from [Har10b]. The related curve is indicated in green. The computational time by the combination technique is represented by the red curve. Notice that we have set up both methods such that they provide similar accuracies for the approximation of the solution. From level 3 to 9, the combination technique is in average a factor 30 faster than the frame discretization, where the speed-up is growing when the level increases. Nevertheless, it seems that, from level 7 on, both methods do not achieve the theoretical rate of $J^3 4^J$. Furthermore, we present in this plot the time consumed for exclusively computing the appropriate right hand sides for the combination technique, indicated by the blue line. As can be seen, on the higher levels, this computation takes nearly half of the total computational time. A potential improvement could here be made by using the matrix-vector product from [Zei11]. Finally, we have plotted the time which is needed for exclusively solving the linear systems by the tensor product solver. Here, it seems that we have the optimal behavior of order $J4^J$ up to level 7. Then, also this rate deteriorates.

The Poisson equation on a random domain

In order to understand the differences between the domain mapping method and the perturbation method, we shall consider also the second example from Section IV.6. This means, we consider again $D_{\text{ref}} = \{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 < 1\}$ as reference domain and the load is set to $f(\mathbf{x}) \equiv 1$. The random field $\mathbf{V}$ is given by its mean $\mathbb{E}[\mathbf{V}](\mathbf{x}) = \mathbf{x}$ and its covariance

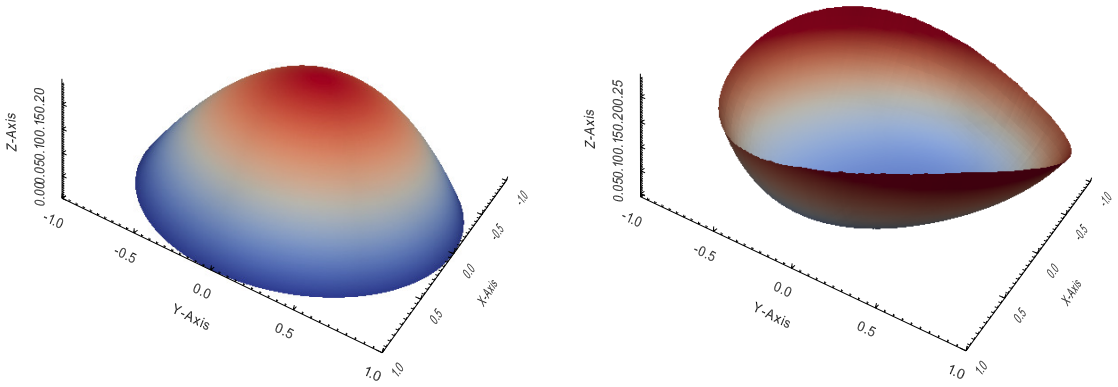

Figure V.3: Solution $\overline{u}$ (left) and $\mathbb{V}[\delta u]$ (right) on the unit disc.

function

$$\mathrm{Cov}[\mathbf{V}](\mathbf{x},\mathbf{y}) = \frac{1}{100}\begin{bmatrix} 5\exp(-4\|\mathbf{x}-\mathbf{y}\|_2^2) & \exp(-0.1\|2\mathbf{x}-\mathbf{y}\|_2^2) \\ \exp(-0.1\|\mathbf{x}-2\mathbf{y}\|_2^2) & 5\exp(-\|\mathbf{x}-\mathbf{y}\|_2^2) \end{bmatrix}.$$

In Figure V.3, a visualization of the solution $\overline{u}$ to (2.3) (left) and the variance $\mathbb{V}[\delta u]$ of the solution to (2.8) (right) is depicted. As can be seen, we compute an approximation for $\overline{u}$ which is similar to the mean depicted on the left in Figure IV.7. The situation changes dramatically if we have a look at the related variances. For the computation of $\mathbb{V}[\delta u]$, we solve a single Poisson equation with homogenous load and non-homogenous tensor product Dirichlet data on $\Gamma_{\mathrm{ref}} \times \Gamma_{\mathrm{ref}}$. By the maximum principle, the related solution cannot have a maximum inside the domain. In contrast to this, in the domain mapping approach, we solve many Poisson equations with non-homogenous load and homogenous Dirichlet data. Thus, again by the maximum principle, each solution attains its maximum inside the domain. Since the variance on the reference domain is a weighted average of these solutions, it also attains its maximum inside the domain while preserving the homogenous boundary conditions.

This indicates that the two approaches in fact compute two different things. In the domain mapping approach, we keep track of each point in the reference domain D_{ref} and map it with the vector field $\mathbf{V}$. The statistical data are then obtained by taking into account each value which is assigned to the transported point. In the perturbation approach, we fix a set of points inside the reference domain D_{ref} and observe how the values assigned to this points alter if we perturb the boundary of this domain. The resulting qualitative differences are clearly demonstrated by this numerical example.

Appendix

1. Multivariate combinatorics

We start this section by defining the arithmetic for multi-indices. To that end, let $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^M$ for some $M \in \mathbb{N}$ with $M \geqslant 1$. The set of natural numbers is always supposed to include the element 0, i.e. $0 \in \mathbb{N}$. We define the addition and subtraction of two multi-indices in the canonical way. Moreover, we define

$$\boldsymbol{\alpha}^{\boldsymbol{\beta}} := \alpha_1^{\beta_1} \cdots \alpha_M^{\beta_M}$$

with the convention $0^0 = 1$. The modulus of $\boldsymbol{\alpha}$ is given by

$$|\boldsymbol{\alpha}| := \sum_{i=1}^{M} \alpha_i$$

and its factorial is defined according to

$$\boldsymbol{\alpha}! := \alpha_1! \cdots \alpha_M!.$$

Then, we can also define the multivariate binomial coefficient

$$\binom{\boldsymbol{\alpha}}{\boldsymbol{\beta}} := \frac{\boldsymbol{\alpha}!}{(\boldsymbol{\alpha} - \boldsymbol{\beta})!\boldsymbol{\beta}!},$$

where we assume $\boldsymbol{\beta} \leqslant \boldsymbol{\alpha}$ and the relation $\leqslant$ has to be understood component-wise.

According to [CS96], we introduce the following sets

$$P(\alpha, r) := \left\{ (k_1, \ldots, k_\alpha) \in \mathbb{N}^\alpha : \sum_{i=1}^{\alpha} k_i = r,\ \sum_{i=1}^{\alpha} i k_i = \alpha \right\} \tag{1.1}$$

and

$$\begin{aligned} P(\boldsymbol{\alpha}, r) := \Bigg\{ & ((k_1, \boldsymbol{\beta}_1), \ldots, (k_n, \boldsymbol{\beta}_n)) \in (\mathbb{N} \times \mathbb{N}^M)^n : \sum_{i=1}^{n} k_i \boldsymbol{\beta}_i = \boldsymbol{\alpha},\ \sum_{i=1}^{n} k_i = r, \\ & \text{and } \exists\, 1 \leq s \leq n : k_i = 0 \text{ and } \boldsymbol{\beta}_i = \mathbf{0} \text{ for all } 1 \leq i \leq n - s, \\ & k_i > 0 \text{ for all } n - s + 1 \leq i \leq n \text{ and } \mathbf{0} \prec \boldsymbol{\beta}_{n-s+1} \prec \cdots \prec \boldsymbol{\beta}_n \Bigg\}, \end{aligned} \tag{1.2}$$

where $n = |\boldsymbol{\alpha}|$. Here and later on, the relation $\boldsymbol{\nu} \prec \boldsymbol{\nu}'$ means either $|\boldsymbol{\nu}| < |\boldsymbol{\nu}'|$ or, if $|\boldsymbol{\nu}| = |\boldsymbol{\nu}'|$, it denotes the lexicographical order. Furthermore, we define for $d \in \mathbb{N}$ the set

$$
\begin{aligned}
P(\boldsymbol{\alpha}, \boldsymbol{\alpha}') := \Big\{ &((\mathbf{k}_1, \boldsymbol{\beta}_1), \dots, (\mathbf{k}_n, \boldsymbol{\beta}_n)) \in (\mathbb{N}^d \times \mathbb{N}^M)^n : \sum_{i=1}^n |\mathbf{k}_i| \boldsymbol{\beta}_i = \boldsymbol{\alpha},\ \sum_{i=1}^n \mathbf{k}_i = \boldsymbol{\alpha}', \\
&\text{and } \exists 1 \leq s \leq n : |\mathbf{k}_j| = |\boldsymbol{\beta}_a| = 0 \text{ for all } 1 \leq i \leq n-s, \\
&|\mathbf{k}_i| \neq 0 \text{ for all } n-s+1 \leq i \leq n \text{ and } \mathbf{0} \prec \boldsymbol{\beta}_{n-s+1} \prec \cdots \prec \boldsymbol{\beta}_n \Big\},
\end{aligned} \tag{1.3}
$$

where $n = |\boldsymbol{\alpha}|$.

The following lemma is a special case of formula (7.4) in [CDS10].

(1.4) **Lemma.** Let $\boldsymbol{\gamma} = \{\gamma_i\}_i \in \ell^1(\mathcal{I})$ for an index set $\mathcal{I} \subset \mathbb{N}$ with $\gamma_i \geqslant 0$. Morever, assume that $c_\gamma := \sum_{i \in \mathcal{I}} \gamma_i < 1$ and that $\boldsymbol{\gamma}$ has a finite support. Then, it holds

$$
\sum_{\boldsymbol{\alpha}} \frac{|\boldsymbol{\alpha}|!}{\boldsymbol{\alpha}!} \boldsymbol{\gamma}^{\boldsymbol{\alpha}} = \frac{1}{1 - c_\gamma}
$$

and therefore there exists a constant with $|\boldsymbol{\alpha}|!/\boldsymbol{\alpha}!\boldsymbol{\gamma}^{\boldsymbol{\alpha}} \leqslant c$ for all $\boldsymbol{\alpha} \in \mathbb{N}^M$, where $M := \#\mathcal{I}$.

Proof. It holds

$$
\sum_{\boldsymbol{\alpha}} \frac{|\boldsymbol{\alpha}|!}{\boldsymbol{\alpha}!} \boldsymbol{\gamma}^{\boldsymbol{\alpha}} = \sum_{i=0}^\infty \sum_{|\boldsymbol{\alpha}|=i} \frac{i!}{\boldsymbol{\alpha}!} \boldsymbol{\gamma}^{\boldsymbol{\alpha}} = \sum_{i=0}^\infty \Big(\sum_{j=1}^M \gamma_j \Big)^i = \sum_{i=0}^\infty c_\gamma^i = \frac{1}{1 - c_\gamma}
$$

by the multinomial theorem and the limit of the geometric series. □

(1.5) **Lemma.** For all $\alpha, \beta, r \in \mathbb{N}$ with $r > 0$ it holds

$$
\binom{\alpha + r - 1}{r - 1} \binom{\beta + r - 1}{r - 1} \leqslant \frac{(\alpha + \beta)!}{\alpha! \beta!} \binom{\alpha + \beta + r - 1}{r - 1}.
$$

Proof. It holds

$$
\begin{aligned}
& \binom{\alpha + r - 1}{r - 1} \binom{\beta + r - 1}{r - 1} \leqslant \frac{(\alpha + \beta)!}{\alpha! \beta!} \binom{\alpha + \beta + r - 1}{r - 1} \\
\iff\ & \binom{\alpha + r - 1}{r - 1} \frac{(\beta + r - 1)!}{\beta! (r - 1)!} \leqslant \frac{(\alpha + \beta)!}{\alpha! \beta!} \frac{(\alpha + \beta + r - 1)!}{(\alpha + \beta)! (r - 1)!} \\
\iff\ & \binom{\alpha + r - 1}{r - 1} (\beta + r - 1)! \leqslant \frac{(\alpha + \beta + r - 1)!}{\alpha!} \\
\iff\ & \binom{\alpha + r - 1}{r - 1} \leqslant \binom{\alpha + \beta + r - 1}{\beta + r - 1}.
\end{aligned}
$$

The last inequality is true due to the monotonically increasing diagonals in Pascal's triangle. This proves the assertion. □

(1.6) **Lemma.** It holds for $\boldsymbol{\alpha} \in \mathbb{N}^M$ and $r \in \mathbb{N}$ with $r \leqslant |\boldsymbol{\alpha}|$ that

$$\boldsymbol{\alpha}! \sum_{P(\boldsymbol{\alpha},r)} \prod_{i=1}^{|\boldsymbol{\alpha}|} \frac{1}{k_i!} \leqslant \frac{|\boldsymbol{\alpha}|!}{r!} \binom{|\boldsymbol{\alpha}| + r - 1}{r - 1}.$$

Proof. From [CS96], we have the identity

$$r! \sum_{P(\boldsymbol{\alpha},r)} \prod_{i=1}^{|\boldsymbol{\alpha}|} \frac{1}{k_i!} = |s^+(\boldsymbol{\alpha}, r)|,$$

where

$$s^+(\boldsymbol{\alpha}, r) := \left\{ (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_r) : |\boldsymbol{\beta}_i| \neq 0 \text{ and } \sum_{i=1}^{r} \boldsymbol{\beta}_i = \boldsymbol{\alpha} \right\}.$$

To bound the cardinality of the set $s^+(\boldsymbol{\alpha}, r)$ we use the identity for the number of weak integer compositions, see e.g. [HM09]: It holds

$$|\{(\beta_1, \ldots, \beta_r) : \beta_i \in \mathbb{N} \text{ and } \beta_1 + \ldots + \beta_r = \alpha\}| = \binom{\alpha + r - 1}{r - 1}.$$

Thus, we estimate $|s^+(\boldsymbol{\alpha}, r)|$, by the product of the number of weak compositions in each component. This yields

$$|s^+(\boldsymbol{\alpha}, r)| \leqslant \prod_{i=1}^{M} \binom{\alpha_i + r - 1}{r - 1}.$$

The proof is now by induction on M. The induction hypothesis is given by

$$\prod_{i=1}^{M} \binom{\alpha_i + r - 1}{r - 1} \leqslant \frac{|\boldsymbol{\alpha}|!}{\boldsymbol{\alpha}!} \binom{|\boldsymbol{\alpha}| + r - 1}{r - 1}.$$

For $M = 1$, we have

$$\binom{\alpha_1 + r - 1}{r - 1} = \frac{\alpha_1!}{\alpha_1!} \binom{\alpha_1 + r - 1}{r - 1},$$

which holds with equality. Let the induction hypothesis be valid for $M - 1$ and set $\boldsymbol{\alpha}_{M-1} = [\alpha_1, \ldots, \alpha_{M-1}]$. Then, we derive with the previous lemma that

$$\begin{aligned}
\prod_{i=1}^{M} \binom{\alpha_i + r - 1}{r - 1} &\leqslant \frac{|\boldsymbol{\alpha}_{M-1}|!}{\boldsymbol{\alpha}_{M-1}!} \binom{|\boldsymbol{\alpha}_{M-1}| + r - 1}{r - 1} \binom{\alpha_M + r - 1}{r - 1} \\
&\leqslant \frac{|\boldsymbol{\alpha}_{M-1}|!}{\boldsymbol{\alpha}_{M-1}!} \frac{(|\boldsymbol{\alpha}_{M-1}| + \alpha_M)!}{|\boldsymbol{\alpha}_{M-1}|! \alpha_M!} \binom{|\boldsymbol{\alpha}_{M-1}| + \alpha_M + r - 1}{r - 1} \\
&= \frac{|\boldsymbol{\alpha}|!}{\boldsymbol{\alpha}!} \binom{|\boldsymbol{\alpha}| + r - 1}{r - 1}.
\end{aligned}$$

We therefore arrive at

$$r! \sum_{P(\boldsymbol{\alpha},r)} \prod_{i=1}^{|\boldsymbol{\alpha}|} \frac{1}{k_i!} \leqslant \frac{|\boldsymbol{\alpha}|!}{\boldsymbol{\alpha}!} \binom{|\boldsymbol{\alpha}| + r - 1}{r - 1}.$$

Rearranging this expression yields the assertion. □

(1.7) **Lemma.** Let $c, m \in \mathbb{R}$ with $m \geqslant 2$ and $c \geqslant m/(m-1)$. It holds for $n \in \mathbb{N}$ that

$$\frac{c}{m} \frac{c^n - 1}{c - 1} \leqslant c^n.$$

Proof. It holds

$$\begin{aligned} & \frac{c}{m} \frac{c^n - 1}{c - 1} && \leqslant c^n \\ \iff \quad & c^{n+1} - c && \leqslant m(c^{n+1} - c^n) \\ \iff \quad & mc^n && \leqslant (m-1)c^{n+1} + c \\ \iff \quad & \frac{m}{m-1} && \leqslant c + \frac{1}{(m-1)c^{n-1}} \end{aligned}$$

Omitting the second summand together with the condition $c \geqslant m/(m-1)$ yields the assertion. □

2. Interpolation of function spaces

We employ the K-method for real interpolation, cf. [BL76], as it is presented in [BS08]. To that end, let $\mathscr{B}_1 \subset \mathscr{B}_0$ be two Banach spaces. We introduce

(2.1) $$K(t,u) = \inf_{v \in \mathscr{B}_1} (\|u - v\|_{\mathscr{B}_0} + t\|v\|_{\mathscr{B}_1}) \quad \text{for } t > 0.$$

Obviously, it holds $K(t,u) \leqslant t\|u\|_{\mathscr{B}_1}$ by the choice $u = v$ for $u \in \mathscr{B}_1$ and $K(t,u) \leqslant \|u\|_{\mathscr{B}_0}$ by the choice $v = 0$. For $1 \leqslant p < \infty$ and $0 < \theta < 1$, we define the norm

$$\|u\|_{[\mathscr{B}_0,\mathscr{B}_1]_{\theta,p}} := \left(\int_0^\infty (t^{-\theta} K(t,u))^p \frac{\mathrm{d}t}{t} \right)^{\frac{1}{p}}.$$

For $p = \infty$, we set

$$\|u\|_{[\mathscr{B}_0,\mathscr{B}_1]_{\theta,\infty}} := \sup_{0<t<\infty} t^{-\theta} K(t,u).$$

The space

$$[\mathscr{B}_0,\mathscr{B}_1]_{\theta,p} := \{u \in \mathscr{B}_0 : \|u\|_{[\mathscr{B}_0,\mathscr{B}_1]_{\theta,p}} < \infty\}$$

then becomes together with the norm $\|\cdot\|_{[\mathscr{B}_0,\mathscr{B}_1]_{\theta,p}}$ a Banach space, cf. [BL76]. If $\mathscr{B}_1$ is dense in $\mathscr{B}_0$, it holds for the related dual space

(2.2) $$([\mathscr{B}_0,\mathscr{B}_1]_{\theta,p})' = [\mathscr{B}_1',\mathscr{B}_0']_{1-\theta,p'}$$

for $0 < \theta < 1$, $1 \leqslant p < \infty$ and $1/p + 1/p' = 1$, cf. [BS08, p. 373].

We have the following interpolation theorem for operators on Banach spaces, cf. [BL76, Theorem 3.1.2] and [BS08, Proposition 14.1.5].

(2.3) **Theorem.** Let $\mathscr{A}_i, \mathscr{B}_i$ for $i \in \{0,1\}$ be two pairs of Banach spaces. Moreover, let $T\colon \mathscr{A}_i \to \mathscr{B}_i$. Then, it holds also $T\colon [\mathscr{A}_0, \mathscr{A}_1]_{\theta,p} \to [\mathscr{B}_0, \mathscr{B}_1]_{\theta,p}$, where the operator norm of T satisfies

$$\|T\|_{[\mathscr{A}_0,\mathscr{A}_1]_{\theta,p}\to[\mathscr{B}_0,\mathscr{B}_1]_{\theta,p}} \leqslant \|T\|^{1-\theta}_{\mathscr{A}_0\to\mathscr{B}_0}\|T\|^{\theta}_{\mathscr{A}_1\to\mathscr{B}_1}.$$

Proof. For a proof of this theorem, see [BS08, Proposition 14.1.5]. □

Especially for the Sobolev spaces $H^s(D)$ and $H^p(D)$, it holds

$$[H^s(D), H^p(D)]_{\theta,2} = H^{(1-\theta)s+\theta p}(D) \tag{2.4}$$

for all $s, p \in \mathbb{R}$ with an equivalent norm, i.e. the interpolation norm is equivalent to the classical Sobolev norm, cf. [BS08, Theorem 14.2.7].

Next, for $d_1, d_2 \in \mathbb{N}$, let $D \subset \mathbb{R}^{d_1}$ be a Lipschitz domain. We consider the pair $L^2(D;\mathbb{R}^{d_2})$ and $H^p(D;\mathbb{R}^{d_2})$. For $d_2 = 1$, we observe

$$[L^2(D), H^p(D)]_{\theta,2} = H^{\theta p}(D).$$

Thus, with the redefinition $\theta := s/p$, we obtain $[L^2(D), H^p(D)]_{s/p,2} = H^s(D)$. On the one hand, the related K-functional, cf. (2.1), satisfies

$$\begin{aligned}
K(t,\mathbf{u})^2 &= \Big(\inf_{\mathbf{v}\in H^p(D;\mathbb{R}^{d_2})} \|\mathbf{u}-\mathbf{v}\|_{L^2(D;\mathbb{R}^{d_2})} + t\|\mathbf{v}\|_{H^p(D;\mathbb{R}^{d_2})}\Big)^2 \\
&\leqslant \Big(\inf_{\mathbf{v}\in H^p(D;\mathbb{R}^{d_2})} \sum_{i=1}^{d_2} \|u_i - v_i\|_{L^2(D)} + t\|v_i\|_{H^p(D)}\Big)^2 \\
&= \Big(\sum_{i=1}^{d_2} \inf_{v_i\in H^p(D)} \|u_i - v_i\|_{L^2(D)} + t\|v_i\|_{H^p(D)}\Big)^2 \\
&\leqslant d_2 \sum_{i=1}^{d_2} \big(\inf_{v_i\in H^p(D)} \|u_i - v_i\|_{L^2(D)} + t\|v_i\|_{H^p(D)}\big)^2 = d_2 \sum_{i=1}^{d_2} K(t,u_i)^2.
\end{aligned}$$

Here, the infimum can be interchanged with the summation since the infimum is taken with respect to independent sets. The last step holds by the discrete Cauchy-Schwarz

inequality. On the other hand, it holds

$$\begin{aligned}
K(t,\mathbf{u})^2 &= \Big(\inf_{\mathbf{v}\in H^p(D;\mathbb{R}^{d_2})} \|\mathbf{u}-\mathbf{v}\|_{L^2(D;\mathbb{R}^{d_2})} + t\|\mathbf{v}\|_{H^p(D;\mathbb{R}^{d_2})}\Big)^2 \\
&= \inf_{\mathbf{v}\in H^p(D;\mathbb{R}^{d_2})} \big(\|\mathbf{u}-\mathbf{v}\|_{L^2(D;\mathbb{R}^{d_2})} + t\|\mathbf{v}\|_{H^p(D;\mathbb{R}^{d_2})}\big)^2 \\
&= \inf_{\mathbf{v}\in H^p(D;\mathbb{R}^{d_2})} \sum_{i=1}^{d_2} \|u_i - v_i\|_{L^2(D)}^2 + 2t\Big(\sum_{i=1}^{d_2} \|u_i - v_i\|_{L^2(D)}^2 \sum_{i=1}^{d_2} \|v_i\|_{L^2(D)}^2\Big)^{\frac{1}{2}} \\
&\quad + t^2 \sum_{i=1}^{d_2} \|v_i\|_{H^p(D)}^2 \\
&\geqslant \inf_{\mathbf{v}\in H^p(D;\mathbb{R}^{d_2})} \sum_{i=1}^{d_2} \|u_i - v_i\|_{L^2(D)}^2 + 2t\sum_{i=1}^{d_2} \|u_i - v_i\|_{L^2(D)}\|v_i\|_{L^2(D)} \\
&\quad + t^2 \sum_{i=1}^{d_2} \|v_i\|_{H^p(D)}^2 \\
&= \sum_{i=1}^{d_2} K(t,u_i)^2,
\end{aligned}$$

again by the discrete Cauchy-Schwarz inequality. The combination of both estimates shows the norm equivalence

$$\|\mathbf{u}\|_{H^s(D;\mathbb{R}^{d_2})} \lesssim \|\mathbf{u}\|_{[L^2(D;\mathbb{R}^{d_2}),H^p(D;\mathbb{R}^{d_2})]_{s/p,2}} \lesssim \sqrt{d_2}\|\mathbf{u}\|_{H^s(D;\mathbb{R}^{d_2})}, \tag{2.5}$$

where the hidden constant is induced by the norm equivalence between $[L^2(D), H^p(D)]_{s/p,2}$ and $H^s(D)$.

Now let $\boldsymbol{\varphi} \in H^p(D;\mathbb{R}^{d_2})$. Then, due to the definition of the dual norm, we deduce

$$\|\boldsymbol{\varphi}\|_{H^p(D;\mathbb{R}^{d_2})} = \sup_{\mathbf{0}\neq\mathbf{v}\in\tilde{H}^{-p}(D;\mathbb{R}^{d_2})} \frac{(\boldsymbol{\varphi},\mathbf{v})_{L^2(D;\mathbb{R}^d)}}{\|\mathbf{v}\|_{\tilde{H}^{-p}(D;\mathbb{R}^{d_2})}}.$$

Hence, the operator $(\boldsymbol{\varphi},\cdot)_{L^2(D;\mathbb{R}^{d_2})}\colon \tilde{H}^{-p}(D;\mathbb{R}^{d_2}) \to \mathbb{R}$ is bounded with norm $\|\boldsymbol{\varphi}\|_{H^p(D;\mathbb{R}^{d_2})}$. Moreover, we have

$$\|\boldsymbol{\varphi}\|_{L^2(D;\mathbb{R}^{d_2})} = \sup_{\mathbf{0}\neq\mathbf{v}\in L^2(D;\mathbb{R}^{d_2})} \frac{(\boldsymbol{\varphi},\mathbf{v})_{L^2(D;\mathbb{R}^{d_2})}}{\|\mathbf{v}\|_{L^2(D;\mathbb{R}^{d_2})}}.$$

Thus, the operator $(\boldsymbol{\varphi},\cdot)_{L^2(D;\mathbb{R}^{d_2})}\colon L^2(D;\mathbb{R}^{d_2}) \to \mathbb{R}$ is bounded with norm $\|\boldsymbol{\varphi}\|_{L^2(D;\mathbb{R}^d)}$.

As a consequence, we obtain by Theorem (2.3) that

$$\begin{aligned}
&\|(\boldsymbol{\varphi},\cdot)_{L^2(D;\mathbb{R}^{d_2})}\|_{[\tilde{H}^{-p}(D;\mathbb{R}^{d_2}),L^2(D;\mathbb{R}^{d_2})]_{1-s/p,2}\to\mathbb{R}} \\
&\quad\lesssim \|(\boldsymbol{\varphi},\cdot)_{L^2(D;\mathbb{R}^{d_2})}\|_{\tilde{H}^{-p}(D)\to\mathbb{R}}^{s/p} \|(\boldsymbol{\varphi},\cdot)_{L^2(D;\mathbb{R}^{d_2})}\|_{L^2(D)\to\mathbb{R}}^{1-s/p},
\end{aligned}$$

where the hidden constant comes from the norm equivalence between $\mathbb{R}$ and $[\mathbb{R},\mathbb{R}]_{\theta,2}$. As a consequence, we have due to (2.5) that

$$\begin{aligned}
\|\boldsymbol{\varphi}\|_{H^s(D;\mathbb{R}^{d_2})} &\lesssim \|\boldsymbol{\varphi}\|_{[L^2(D;\mathbb{R}^{d_2}),H^p(D;\mathbb{R}^{d_2})]_{s/p,2}} \\
&= \|(\boldsymbol{\varphi},\cdot)_{L^2(D;\mathbb{R}^{d_2})}\|_{[\tilde{H}^{-p}(D;\mathbb{R}^{d_2}),L^2(D;\mathbb{R}^{d_2})]_{1-s/p,2}\to\mathbb{R}}.
\end{aligned}$$

Therefore, we have just proven the norm estimate

$$\|\boldsymbol{\varphi}\|_{H^s(D;\mathbb{R}^{d_2})} \lesssim \|\boldsymbol{\varphi}\|^{1-s/p}_{L^2(D;\mathbb{R}^{d_2})} \|\boldsymbol{\varphi}\|^{s/p}_{H^p(D;\mathbb{R}^{d_2})}. \tag{2.6}$$

3. Asymptotically smooth kernels

Usually, for fast boundary element methods, it is assumed that the kernel is asymptotically smooth in the *space*. This means

$$|\partial_{\mathbf{x}}^{\boldsymbol{\alpha}} \partial_{\mathbf{y}}^{\boldsymbol{\beta}} k(\mathbf{x},\mathbf{y})| \leq c_k \frac{(|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|)!}{r_k^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}} \|\mathbf{x}-\mathbf{y}\|_2^{-2-2q-|\boldsymbol{\alpha}|-|\boldsymbol{\beta}|} \tag{3.1}$$

for some constants $c_k > 0$ and $r_k > 0$ which are independent of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

We show in the following theorem that the decay estimate (3.1) implies the condition (III.3.6) provided that the parameterization is piecewise analytic.

(3.2) **Theorem.** Let the kernel function $k(\mathbf{x},\mathbf{y})$ fulfill the decay estimate (3.1) and let the parameterizations $\boldsymbol{\gamma}_i$ and $\boldsymbol{\gamma}_{i'}$ be analytic functions. Then, for all $i, i' = 1, 2, \ldots, K$, there exist constants $R_{i,i'} > 0$ and $c > 0$ such that the transported kernel (III.3.4) satisfies the estimate

$$|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}} \partial_{\mathbf{t}}^{\boldsymbol{\beta}} k_{i,i'}(\mathbf{s},\mathbf{t})| \leqslant c_k \frac{(|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|)!}{R_{i,i'}^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}} \frac{\| |\boldsymbol{\kappa}_i| \|_{L^\infty(\square)} \| |\boldsymbol{\kappa}_{i'}| \|_{L^\infty(\square)}}{\|\boldsymbol{\kappa}_i(\mathbf{s}) - \boldsymbol{\kappa}_{i'}(\mathbf{t})\|_2^{2(1+q)+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}} \tag{3.3}$$

uniformly for all $\boldsymbol{\alpha}, \boldsymbol{\beta}$ provided that $2 + 2q + |\boldsymbol{\alpha}| + |\boldsymbol{\beta}| > 0$.

Proof. For the following, it is convenient to compute an estimate of

$$\tilde{k}_{i,i'}(\mathbf{s},\mathbf{t}) := k\big(\boldsymbol{\kappa}_i(\mathbf{s}), \boldsymbol{\kappa}_{i'}(\mathbf{t})\big).$$

To this end, without loss of generality, we assume that the geometry D is scaled such that $\operatorname{diam}(D) \leq 1$ and thus

$$\|\kappa_i\|_{L^\infty(\square)} \leq 1, \quad \|\boldsymbol{\kappa}_{i'}\|_{L^\infty(\square)} \leq 1. \tag{3.4}$$

Then, the partial derivatives of $\tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})$ can be expressed in terms of the Faà di Bruno formula (see [CS96]) in accordance with

$$\begin{aligned} \partial_{\mathbf{s}}^{\boldsymbol{\alpha}} \partial_{\mathbf{t}}^{\boldsymbol{\beta}} \tilde{k}_{i,i'}(\mathbf{s},\mathbf{t}) = \sum_{\substack{1 \leqslant |\boldsymbol{\alpha}'| \leqslant |\boldsymbol{\alpha}| \\ 1 \leqslant |\boldsymbol{\beta}'| \leqslant |\boldsymbol{\beta}|}} \partial_{\mathbf{x}}^{\boldsymbol{\alpha}'} \partial_{\mathbf{y}}^{\boldsymbol{\beta}'} \tilde{k}_{i,i'}(\mathbf{s},\mathbf{t}) \Bigg(\sum_{P(\boldsymbol{\alpha},\boldsymbol{\alpha}')} \boldsymbol{\alpha}! \prod_{a=1}^{|\boldsymbol{\alpha}|} \frac{(\partial_{\mathbf{s}}^{\boldsymbol{\nu}_a} \boldsymbol{\kappa}_i(\mathbf{s}))^{\boldsymbol{\mu}_a}}{(\boldsymbol{\mu}_a!)(\boldsymbol{\nu}_a!)^{|\boldsymbol{\mu}_a|}} \Bigg) \\ \cdot \Bigg(\sum_{P(\boldsymbol{\beta},\boldsymbol{\beta}')} \boldsymbol{\beta}! \prod_{b=1}^{|\boldsymbol{\beta}|} \frac{(\partial_{\mathbf{t}}^{\boldsymbol{\nu}_b} \boldsymbol{\kappa}_{i'}(\mathbf{t}))^{\boldsymbol{\mu}_b}}{(\boldsymbol{\mu}_b!)(\boldsymbol{\nu}_b!)^{|\boldsymbol{\mu}_b|}} \Bigg) \end{aligned} \tag{3.5}$$

with multi-indices $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^2$ and $\boldsymbol{\alpha}', \boldsymbol{\beta}' \in \mathbb{N}^3$. Here and in the sequel, the set $P(\boldsymbol{\alpha}, \boldsymbol{\alpha}')$ is given by (1.3). Nevertheless, we state the definition here again, in order to clarify the

purpose of each occurring multiindex.

$$P(\boldsymbol{\alpha},\boldsymbol{\alpha}') := \Big\{(\boldsymbol{\mu}_a,\boldsymbol{\nu}_a)_{a=1}^{|\boldsymbol{\alpha}|} \in (\mathbb{N}^3,\mathbb{N}^2)^{|\boldsymbol{\alpha}|} \text{ with } \sum_a \boldsymbol{\mu}_a = \boldsymbol{\alpha}' \text{ and } \sum_a |\boldsymbol{\mu}_a|\boldsymbol{\nu}_a = \boldsymbol{\alpha}$$
$$\text{and } \exists 1 \leqslant s \leqslant |\boldsymbol{\alpha}| : |\boldsymbol{\mu}_a| = |\boldsymbol{\nu}_a| = 0 \text{ for all } 1 \leqslant a \leqslant s,$$
$$|\boldsymbol{\mu}_a| \neq 0 \text{ for all } s+1 \leq a \leqslant |\boldsymbol{\alpha}| \text{ and } \mathbf{0} \prec \boldsymbol{\nu}_{s+1} \prec \cdots \prec \boldsymbol{\nu}_{|\boldsymbol{\alpha}|}\Big\}.$$

Moreover, in our two-dimensional setting, the relation $\boldsymbol{\nu} \prec \boldsymbol{\nu}'$ for $|\boldsymbol{\nu}| = |\boldsymbol{\nu}'|$ becomes $\nu_1 < \nu_1'$.

Since the parameterizations are analytic and in view of our scaling (3.4), the Cauchy integral formula, cf. [AE06], implies

$$[\partial_{\mathbf{s}}^{\boldsymbol{\alpha}} \boldsymbol{\kappa}_i(\mathbf{s})]_j \leqslant \frac{\boldsymbol{\alpha}!}{\rho_i^{|\boldsymbol{\alpha}|}}, \quad [\partial_{\mathbf{t}}^{\boldsymbol{\beta}} \boldsymbol{\kappa}_i(\mathbf{t})]_j \leqslant \frac{\boldsymbol{\beta}!}{\rho_{i'}^{|\boldsymbol{\beta}|}}, \quad j = 1,2,3 \tag{3.6}$$

for some $\rho_i, \rho_{i'} \in (0,1]$. Inserting these estimates into (3.5) yields

$$\begin{aligned}|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\partial_{\mathbf{t}}^{\boldsymbol{\beta}} \tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})| &\leqslant \sum_{\substack{1\leqslant|\boldsymbol{\alpha}'|\leqslant|\boldsymbol{\alpha}|\\ 1\leqslant|\boldsymbol{\beta}'|\leqslant|\boldsymbol{\beta}|}} |\partial_{\mathbf{x}}^{\boldsymbol{\alpha}'}\partial_{\mathbf{y}}^{\boldsymbol{\beta}'} \tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})| \Bigg(\sum_{p(\boldsymbol{\alpha},\boldsymbol{\alpha}')} \boldsymbol{\alpha}! \prod_{a=1}^{|\boldsymbol{\alpha}|} \frac{\rho_i^{-|\boldsymbol{\nu}_a||\boldsymbol{\mu}_a|}}{\boldsymbol{\mu}_a!}\Bigg) \\ &\qquad \cdot \Bigg(\sum_{p(\boldsymbol{\beta},\boldsymbol{\beta}')} \boldsymbol{\beta}! \prod_{b=1}^{|\boldsymbol{\beta}|} \frac{\rho_{i'}^{-|\boldsymbol{\nu}_b||\boldsymbol{\mu}_b|}}{\boldsymbol{\mu}_b!}\Bigg) \\ &= \frac{\boldsymbol{\alpha}!\boldsymbol{\beta}!}{\rho_i^{|\boldsymbol{\alpha}|}\rho_{i'}^{|\boldsymbol{\beta}|}} \sum_{\substack{1\leqslant|\boldsymbol{\alpha}'|\leqslant|\boldsymbol{\alpha}|\\ 1\leqslant|\boldsymbol{\beta}'|\leqslant|\boldsymbol{\beta}|}} |\partial_{\mathbf{x}}^{\boldsymbol{\alpha}'}\partial_{\mathbf{y}}^{\boldsymbol{\beta}'} \tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})| \sum_{p(\boldsymbol{\alpha},\boldsymbol{\alpha}')} \prod_{a=1}^{|\boldsymbol{\alpha}|} \frac{1}{\boldsymbol{\mu}_a!} \sum_{p(\boldsymbol{\beta},\boldsymbol{\beta}')} \prod_{b=1}^{|\boldsymbol{\beta}|} \frac{1}{\boldsymbol{\mu}_b!}.\end{aligned}$$

We shall next determine upper bounds of the two last terms of this expression. To that end, we employ the identity

$$\sum_{p(\boldsymbol{\alpha},\boldsymbol{\alpha}')} \prod_{b=1}^{|\boldsymbol{\alpha}|} \frac{1}{\boldsymbol{\mu}_b!} = \frac{|s^+(\boldsymbol{\alpha},\boldsymbol{\alpha}')|}{\boldsymbol{\alpha}'!},$$

provided by [CS96], where

$$s^+(\boldsymbol{\alpha},\boldsymbol{\alpha}') := \Big\{(\boldsymbol{\eta}_1,\ldots,\boldsymbol{\eta}_{|\boldsymbol{\alpha}'|}) \colon |\boldsymbol{\eta}_a| \neq 0 \text{ and } \sum_a \boldsymbol{\eta}_a = \boldsymbol{\alpha}\Big\}.$$

To bound the cardinality of the set $s^+(\boldsymbol{\alpha},\boldsymbol{\alpha}')$ we use the estimates for the number of weak integer compositions, cf. [HM09], in each of the two components of $\boldsymbol{\alpha} = (\alpha_1,\alpha_2)$. This yields the trivial combinatorial estimate

$$|s^+(\boldsymbol{\alpha},\boldsymbol{\alpha}')| \leqslant \binom{\alpha_1+|\boldsymbol{\alpha}'|-1}{|\boldsymbol{\alpha}'|-1}\binom{\alpha_2+|\boldsymbol{\alpha}'|-1}{|\boldsymbol{\alpha}'|-1} \leqslant 2^{|\boldsymbol{\alpha}|+2(|\boldsymbol{\alpha}'|-1)}.$$

Thus, we arrive at

$$|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\partial_{\mathbf{t}}^{\boldsymbol{\beta}} \tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})| \leqslant \frac{2^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}\boldsymbol{\alpha}!\boldsymbol{\beta}!}{16\rho_i^{|\boldsymbol{\alpha}|}\rho_{i'}^{|\boldsymbol{\beta}|}} \sum_{\substack{1\leqslant|\boldsymbol{\alpha}'|\leqslant|\boldsymbol{\alpha}|\\ 1\leqslant|\boldsymbol{\beta}'|\leqslant|\boldsymbol{\beta}|}} \frac{4^{|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|}}{\boldsymbol{\alpha}'!\boldsymbol{\beta}'!} |\partial_{\mathbf{x}}^{\boldsymbol{\alpha}'}\partial_{\mathbf{y}}^{\boldsymbol{\beta}'} \tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})|.$$

By inserting the statement (3.1), we deduce

$$\begin{aligned}\left|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\partial_{\mathbf{t}}^{\boldsymbol{\beta}}\tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})\right| &\leqslant \frac{c_k}{16}\frac{2^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}{\rho_i^{|\boldsymbol{\alpha}|}\rho_{i'}^{|\boldsymbol{\beta}|}}\boldsymbol{\alpha}!\boldsymbol{\beta}! \\ &\quad\cdot\sum_{\substack{1\leqslant|\boldsymbol{\alpha}'|\leqslant|\boldsymbol{\alpha}|\\1\leqslant|\boldsymbol{\beta}'|\leqslant|\boldsymbol{\beta}|}}\frac{(|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|)!}{\boldsymbol{\alpha}'!\boldsymbol{\beta}'!}\left(\frac{4}{r_k}\right)^{|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|}\frac{1}{\|\boldsymbol{\kappa}_i(\mathbf{s})-\boldsymbol{\kappa}_{i'}(\mathbf{t})\|_2^{2(1+q)+|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|}}.\end{aligned}$$

Since $\operatorname{diam}(D)\leq 1$, we find

$$\frac{1}{\|\boldsymbol{\kappa}_i(\mathbf{s})-\boldsymbol{\kappa}_{i'}(\mathbf{t})\|_2^{2(1+q)+|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|}}\leq\frac{1}{\|\boldsymbol{\kappa}_i(\mathbf{s})-\boldsymbol{\kappa}_{i'}(\mathbf{t})\|_2^{2(1+q)+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}} \tag{3.7}$$

for all $|\boldsymbol{\alpha}'|\leq|\boldsymbol{\alpha}|$, $|\boldsymbol{\beta}'|\leq|\boldsymbol{\beta}|$ and $\mathbf{s},\mathbf{t}\in\square$. Thus, it holds

$$\begin{aligned}\left|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\partial_{\mathbf{t}}^{\boldsymbol{\beta}}\tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})\right| &\leqslant \frac{c_k}{16}\boldsymbol{\alpha}!\boldsymbol{\beta}!\frac{2^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}\rho_i^{-|\boldsymbol{\alpha}|}\rho_{i'}^{-|\boldsymbol{\beta}|}}{\|\boldsymbol{\gamma}_i(\mathbf{s})-\boldsymbol{\gamma}_{i'}(\mathbf{t})\|_2^{2(1+q)+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}} \\ &\quad\cdot\underbrace{\sum_{\substack{1\leqslant|\boldsymbol{\alpha}'|\leqslant|\boldsymbol{\alpha}|\\1\leqslant|\boldsymbol{\beta}'|\leqslant|\boldsymbol{\beta}|}}\frac{(|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|)!}{\boldsymbol{\alpha}'!\boldsymbol{\beta}'!}\left(\frac{4}{r_k}\right)^{|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|}}_{=:R(\boldsymbol{\alpha},\boldsymbol{\beta})}.\end{aligned}$$

Next, we shall derive an estimation for $R(\boldsymbol{\alpha},\boldsymbol{\beta})$. Obviously, it is

$$(|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|)!=\binom{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}{|\boldsymbol{\beta}|}|\boldsymbol{\alpha}|!|\boldsymbol{\beta}|!=\binom{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}{|\boldsymbol{\alpha}|}|\boldsymbol{\alpha}||\boldsymbol{\beta}|!. \tag{3.8}$$

Additionally, we have by means of multinomial coefficients

$$(|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|)!=\binom{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}{|\boldsymbol{\beta}|}\binom{|\boldsymbol{\alpha}|}{\boldsymbol{\alpha}}\binom{|\boldsymbol{\beta}|}{\boldsymbol{\beta}}\boldsymbol{\alpha}!\boldsymbol{\beta}!.$$

Further, we know by the multinomial theorem that

$$(x_1+\cdots+x_m)^{|\boldsymbol{\alpha}|}=\sum_{|\boldsymbol{\alpha}'|=|\boldsymbol{\alpha}|}\binom{|\boldsymbol{\alpha}|}{\boldsymbol{\alpha}'}x_1^{\alpha_1'}\cdots x_m^{\alpha_m'}$$

for multi-indices $\boldsymbol{\alpha},\boldsymbol{\alpha}'\in\mathbb{N}^m$ and $x_1,\ldots,x_m\in\mathbb{R}$. Assuming $r_k\leqslant 1$, we arrive at

$$\begin{aligned}R(\boldsymbol{\alpha},\boldsymbol{\beta}) &= \sum_{\substack{1\leqslant|\boldsymbol{\alpha}'|\leqslant|\boldsymbol{\alpha}|\\1\leqslant|\boldsymbol{\beta}'|\leqslant|\boldsymbol{\beta}|}}\frac{(|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|)!}{\boldsymbol{\alpha}'!\boldsymbol{\beta}'!}\left(\frac{4}{r_k}\right)^{|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|}=\sum_{a=1}^{|\boldsymbol{\alpha}|}\sum_{b=1}^{|\boldsymbol{\beta}|}\sum_{\substack{|\boldsymbol{\alpha}'|=a\\|\boldsymbol{\beta}'|=b}}\frac{(a+b)!}{\boldsymbol{\alpha}'!\boldsymbol{\beta}'!}\left(\frac{4}{r_k}\right)^{a+b} \\ &= \sum_{a=1}^{|\boldsymbol{\alpha}|}\sum_{b=1}^{|\boldsymbol{\beta}|}\binom{a+b}{b}\sum_{\substack{|\boldsymbol{\alpha}'|=a\\|\boldsymbol{\beta}'|=b}}\binom{a}{\boldsymbol{\alpha}'}\binom{b}{\boldsymbol{\beta}'}\left(\frac{4}{r_k}\right)^{a+b} \\ &\leqslant \sum_{a=1}^{|\boldsymbol{\alpha}|}\sum_{b=1}^{|\boldsymbol{\beta}|}2^{a+b}\left(\frac{12}{r_k}\right)^{a+b}=\sum_{a=1}^{|\boldsymbol{\alpha}|}\left(\frac{24}{r_k}\right)^{a}\sum_{b=1}^{|\boldsymbol{\beta}|}\left(\frac{24}{r_k}\right)^{b} \\ &\leqslant \left(\frac{24}{24-r_k}\right)^2\left(\frac{24}{r_k}\right)^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}\leqslant 2\left(\frac{24}{r_k}\right)^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}.\end{aligned}$$

Altogether, this yields

$$|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\partial_{\mathbf{t}}^{\boldsymbol{\beta}}\tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})| \leqslant \frac{c_k}{8}\boldsymbol{\alpha}!\boldsymbol{\beta}!\left(\frac{48}{r_k}\right)^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}\frac{\rho_i^{-|\boldsymbol{\alpha}|}\rho_{i'}^{-|\boldsymbol{\beta}|}}{\|\boldsymbol{\kappa}_i(\mathbf{s})-\boldsymbol{\kappa}_{i'}(\mathbf{t})\|_2^{2(1+q)+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}. \tag{3.9}$$

Now, we can easily estimate the decay of the transported kernel function (III.3.4) by the Leibniz formula. It holds

$$\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\partial_{\mathbf{t}}^{\boldsymbol{\beta}}k_{i,i'}(\mathbf{s},\mathbf{t}) = \sum_{\substack{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}\\ \boldsymbol{\beta}'\leqslant\boldsymbol{\beta}}}\binom{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'}\binom{\boldsymbol{\beta}}{\boldsymbol{\beta}'}\partial_{\mathbf{s}}^{\boldsymbol{\alpha}'}\partial_{\mathbf{t}}^{\boldsymbol{\beta}'}\tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})\partial_{\mathbf{s}}^{\boldsymbol{\alpha}-\boldsymbol{\alpha}'}\kappa_i(\mathbf{s})\partial_{\mathbf{t}}^{\boldsymbol{\beta}-\boldsymbol{\beta}'}\kappa_{i'}(\mathbf{t}). \tag{3.10}$$

Since the surface measures are analytic, we can estimate them also by the Cauchy integral formula. It is

$$|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}|\boldsymbol{\kappa}_i|(\mathbf{s})| \leqslant \frac{\boldsymbol{\alpha}!}{\tilde{\rho}_i^{|\boldsymbol{\alpha}|}}|||\boldsymbol{\kappa}_i|||_{L^\infty(\square)}, \quad |\partial_{\mathbf{t}}^{\boldsymbol{\beta}}|\kappa_{i'}|(\mathbf{t})| \leqslant \frac{\boldsymbol{\beta}!}{\tilde{\rho}_{i'}^{|\boldsymbol{\beta}|}}|||\boldsymbol{\kappa}_{i'}|||_{L^\infty(\square)}$$

for some $\tilde{\rho}_i, \tilde{\rho}_{i'} > 0$. We insert these estimates into (3.10) and arrive at

$$\begin{aligned}&|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\partial_{\mathbf{t}}^{\boldsymbol{\beta}}k_{i,i'}(\mathbf{s},\mathbf{t})|\\ &\quad\leqslant \boldsymbol{\alpha}!\boldsymbol{\beta}!|||\boldsymbol{\kappa}_i|||_{L^\infty(\square)}|||\boldsymbol{\kappa}_{i'}|||_{L^\infty(\square)}\sum_{\substack{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}\\ \boldsymbol{\beta}'\leqslant\boldsymbol{\beta}}}\frac{|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}'}\partial_{\mathbf{t}}^{\boldsymbol{\beta}'}\tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})|}{\boldsymbol{\alpha}'!\boldsymbol{\beta}'!}\tilde{\rho}_i^{-|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|}\tilde{\rho}_{i'}^{-|\boldsymbol{\beta}-\boldsymbol{\beta}'|}.\end{aligned}$$

Next, we use (3.9) to replace the derivatives $\partial_{\mathbf{s}}^{\boldsymbol{\alpha}'}\partial_{\mathbf{t}}^{\boldsymbol{\beta}'}\tilde{k}_{i,i'}(\mathbf{s},\mathbf{t})$ which, in view of (3.7) and with $R_{i,i'} := \min\{\rho_i,\rho_{i'},\tilde{\rho}_i,\tilde{\rho}_{i'},1\}$, leads to

$$\begin{aligned}|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\partial_{\mathbf{t}}^{\boldsymbol{\beta}}k_{i,i'}(\mathbf{s},\mathbf{t})| &\leqslant \frac{c_k}{8}\boldsymbol{\alpha}!\boldsymbol{\beta}!|||\boldsymbol{\kappa}_i|||_{L^\infty(\square)}|||\boldsymbol{\kappa}_{i'}|||_{L^\infty(\square)}\\ &\qquad\cdot\sum_{\substack{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}\\ \boldsymbol{\beta}'\leqslant\boldsymbol{\beta}}}\frac{\boldsymbol{\alpha}'!\boldsymbol{\beta}'!}{\boldsymbol{\alpha}'!\boldsymbol{\beta}'!}\left(\frac{48}{r_k}\right)^{|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|}\frac{\rho_i^{-|\boldsymbol{\alpha}'|}\tilde{\rho}_i^{-|\boldsymbol{\alpha}-\boldsymbol{\alpha}'|}\rho_{i'}^{-|\boldsymbol{\beta}'|}\tilde{\rho}_{i'}^{-|\boldsymbol{\beta}-\boldsymbol{\beta}'|}}{\|\boldsymbol{\kappa}_i(\mathbf{s})-\boldsymbol{\kappa}_{i'}(\mathbf{t})\|_2^{2(1+q)+|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|}}\\ &\leqslant \frac{c_k}{8}\boldsymbol{\alpha}!\boldsymbol{\beta}!\frac{|||\boldsymbol{\kappa}_i|||_{L^\infty(\square)}|||\boldsymbol{\kappa}_{i'}|||_{L^\infty(\square)}}{\|\boldsymbol{\kappa}_i(\mathbf{s})-\boldsymbol{\kappa}_{i'}(\mathbf{t})\|_2^{2(1+q)+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}\sum_{\substack{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}\\ \boldsymbol{\beta}'\leqslant\boldsymbol{\beta}}}\left(\frac{48}{R_{i,i'}r_k}\right)^{|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|}.\end{aligned}$$

The remaining sum can be estimated via

$$\sum_{\substack{\boldsymbol{\alpha}'\leqslant\boldsymbol{\alpha}\\ \boldsymbol{\beta}'\leqslant\boldsymbol{\beta}}}q^{|\boldsymbol{\alpha}'|+|\boldsymbol{\beta}'|} \leqslant \frac{q^{\alpha_1+1}-1}{q-1}\cdot\frac{q^{\alpha_2+1}-1}{q-1}\cdot\frac{q^{\beta_1+1}-1}{q-1}\cdot\frac{q^{\beta_2+1}-1}{q-1} \leqslant \left(\frac{q}{q-1}\right)^4 q^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}.$$

For $q = 48/(R_{i,i'}r_k)$, we have

$$\left(\frac{48}{48-R_{i,i'}r_k}\right)^4 \leqslant 2.$$

With the help of (3.8), it finally follows

$$|\partial_{\mathbf{s}}^{\boldsymbol{\alpha}}\partial_{\mathbf{t}}^{\boldsymbol{\beta}}k_{i,i'}(\mathbf{s},\mathbf{t})| \leqslant \frac{c_k}{4}(|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|)!\left(\frac{48}{R_{i,i'}r_k}\right)^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}\frac{|||\boldsymbol{\kappa}_i|||_{L^\infty(\square)}|||\boldsymbol{\kappa}_{i'}|||_{L^\infty(\square)}}{\|\boldsymbol{\kappa}_i(\mathbf{s})-\boldsymbol{\kappa}_{i'}(\mathbf{t})\|_2^{2(1+q)+|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}.$$

This is the desired estimate (3.3). □

BIBLIOGRAPHY

[AE06] H. Amann and J. Escher. *Analysis II*. Birkhäuser, Basel, 2006.

[AE08] H. Amann and J. Escher. *Analysis III*. Birkhäuser, Basel, 2008.

[Alt07] H. W. Alt. *Lineare Funktionalanalysis*. Springer, London, 4th edition, 2007.

[AS64] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. Applied mathematics series. Dover Publications, N. Chemsford, MA, 1964.

[Beb00] M. Bebendorf. Approximation of boundary element matrices. *Numerische Mathematik*, 86(4):565–589, 2000.

[BG04] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004.

[BG05] S. Börm and L. Grasedyck. Hybrid cross approximation of integral operators. *Numerische Mathematik*, 101(2):221–249, 2005.

[BG06] M. Bebendorf and R. Grzhibovskis. Accelerating Galerkin BEM for linear elasticity using adaptive cross approximation. *Mathematical Methods in the Applied Sciences*, 29(14):1721–1747, 2006.

[BK94] J. H. Bramble and J. T. King. A robust finite element method for nonhomogeneous Dirichlet problems in domains with curved boundaries. *Mathematics of Computation*, 63(207):1–17, 1994.

[BL76] J. Bergh and J. Löfström. *Interpolation Spaces. An Introduction*. Springer, Berlin, 1976.

[BL77] N. H. F. Beebe and J. Linderberg. Simplifications in the generation and transformation of two-electron integrals in molecular calculations. *International Journal of Quantum Chemistry*, 12(4):683–705, 1977.

[BNT07] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 45(3):1005–1034, 2007.

[Boc33] S. Bochner. Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind. *Fundamenta Mathematicae*, 20(1):262–176, 1933.

[BPX90] J. H. Bramble, J. E. Pasciak, and J. Xu. Parallel multilevel preconditioners. *Mathematics of Computation*, 55(191):1–22, 1990.

[BR03] M. Bebendorf and S. Rjasanow. Adaptive low-rank approximation of collocation matrices. *Computing*, 70(1):1–24, 2003.

[Bra07] D. Braess. *Finite Elemente: Theorie, Schnelle Löser und Anwendungen in der Elastizitätstheorie*. Springer, London, 2007.

[BS08] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer, Berlin, 3rd edition, 2008.

[BTNT12] J. Beck, R. Tempone, F. Nobile, and L. Tamellini. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Mathematical Models and Methods in Applied Sciences*, 22(9):1250023, 2012.

[BTZ04] I. Babuška, R. Tempone, and G. Zouraris. Galerkin finite element approximations of stochastic elliptic partial differential equations. *SIAM Journal on Numerical Analysis*, 42(2):800–825, 2004.

[CDS10] A. Cohen, R. DeVore, and C. Schwab. Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs. *Foundations of Computational Mathematics*, 10, 2010.

[CK07] C. Canuto and T. Kozubek. A fictitious domain approach to the numerical solution of PDEs in stochastic domains. *Numerische Mathematik*, 107(2):257–293, 2007.

[CNT13] J. E. Castrillon-Candas, F. Nobile, and R. F. Tempone. Analytic regularity and collocation approximation for PDEs with random domain deformations. *ArXiv e-prints 1312.7845*, 2013.

[CS96] G. M. Constantine and T. H. Savits. A multivariate Faà di Bruno formula with applications. *Transactions of the American Mathematical Society*, 248:503–520, 1996.

[CS13] A. Chernov and C. Schwab. First order k-th moment finite element analysis of nonlinear operator equations with stochastic data. *Mathematics of Computation*, 82(284):1859–1888, 2013.

[CZ98] Z. Chen and J. Zou. Finite element methods and their convergence for elliptic and parabolic interface problems. *Numerische Mathematik*, 79(2):175–202, 1998.

[DeV98] R. A. DeVore. Nonlinear approximation. *Acta numerica*, 7:51–150, 1998.

[DHS06] W. Dahmen, H. Harbrecht, and R. Schneider. Compression techniques for boundary integral equations. Asymptotically optimal complexity estimates. *SIAM Journal on Numerical Analysis*, 43(6):2251–2271, 2006.

[DM96] E. G. D'yakonov and S. F. McCormick. *Optimization in Solving Elliptic Problems*. CRC Press, Boca Raton, 1996.

[DU77] J. Diestel and J. J. Uhl. *Vector Measures*. Mathematical surveys and monographs. American Mathematical Society, Providence, 1977.

[DZ11] C. Delfour and J.P. Zolésio. *Shapes and Geometries: Metrics, Analysis, Differential Calculus, and Optimization, Second Edition*. Advances in Design and Control. Society for Industrial and Applied Mathematics, Philadelphia, 2011.

[EEU07] M. Eiermann, O. G. Ernst, and E. Ullmann. Computational aspects of the stochastic finite element method. *Computing and Visualization in Science*, 10(1):3–15, 2007.

[Epp00a] K. Eppler. Boundary integral representations of second derivatives in shape optimization. *Discussiones Mathematicae (Differential Inclusion Control and Optimization)*, 20:63–78, 2000.

[Epp00b] K. Eppler. Optimal shape design for elliptic equations via BIE-methods. *Journal of Applied Mathematics and Computer Science*, 10:487–516, 2000.

[Epp07] K. Eppler. *Efficient Shape Optimization Algorithms for Elliptic Boundary Value Problems*. Habilitation Thesis. Technische Universität Berlin, Germany, 2007.

[FGP83] G. J. Fix, M. D. Gunzburger, and J. S. Peterson. On finite element approximations of problems having inhomogeneous essential boundary conditions. *Computers & Mathematics with Applications*, 9(5):687–700, 1983.

[FST05] P. Frauenfelder, C. Schwab, and R. Todor. Finite elements for elliptic problems with stochastic coefficients. *Computer Methods in Applied Mechanics and Engineering*, 194(2-5):205–228, 2005.

[FWA+09] L. Foster, A. Waagen, N. Aijaz, M. Hurley, A. Luis, J. Rinsky, C. Satyavolu, M. J. Way, P. Gazis, and A. Srivastava. Stable and efficient Gaussian process calculations. *The Journal of Machine Learning Research*, 10:857–882, 2009.

[GH14] M. Griebel and H. Harbrecht. Approximation of bi-variate functions: singular value decomposition versus sparse grids. *IMA Journal of Numerical Analysis*, 34(1):28–54, 2014.

[Gie01] K. Giebermann. Multilevel approximation of boundary integral operators. *Computing*, 67(3):183–207, 2001.

[GKN+13] I. Graham, F. Kuo, J. Nichols, R. Scheichl, C. Schwab, and I. Sloan. Quasi-Monte Carlo finite element methods for elliptic PDEs with log-normal random coefficient. *SAM-Report*, 2013-14, 2013. To appear in Numerische Mathematik.

[Gro62] O. A. Gross. Preferential arrangements. *American Mathematical Monthly*, pages 4–8, 1962.

[GS91] R. Ghanem and P. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer, New York, 1991.

[GSZ92] M. Griebel, M. Schneider, and C. Zenger. A combination technique for the solution of sparse grid problems. In P. de Groen and R. Beauwens, editors, *Iterative Methods in Linear Algebra*, pages 263–281. IMACS, Elsevier, North Holland, 1992.

[Hac85] W. Hackbusch. *Multi-grid Methods and Applications*. Springer, Berlin, 1985.

[Hac99] W. Hackbusch. A sparse matrix arithmetic based on $\mathcal{H}$-matrices part I: Introduction to $\mathcal{H}$-matrices. *Computing*, 62(2):89–108, 1999.

[Hac09] W. Hackbusch. *Hierarchische Matrizen: Algorithmen und Analysis*. Springer, Heidelberg, 2009.

[Hal60] J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1):84–90, 1960.

[Hal76] P. R. Halmos. *Measure Theory*. Graduate Texts in Mathematics. Springer, New York, 1976.

[Har10a] H. Harbrecht. Finite element based second moment analysis for elliptic problems in stochastic domains. In G. Kreiss, P. Lötstedt, A. Målqvist, and M. Neytcheva, editors, *Numerical Mathematics and Advanced Applications 2009*, pages 433–441. Springer Berlin Heidelberg, 2010.

[Har10b] H. Harbrecht. A finite element method for elliptic problems with stochastic input data. *Applied Numerical Mathematics*, 60(3):227–244, 2010.

[HB02] W. Hackbusch and S. Börm. $\mathcal{H}^2$-matrix approximation of integral operators by interpolation. *Applied Numerical Mathematics*, 43(1–2):129 – 143, 2002.

[HCB05] T. J. R. Hughes, J. A. Cottrell, and Y. Bazilevs. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Computer Methods in Applied Mechanics and Engineering*, 194(39):4135–4195, 2005.

[Hig90] N. J. Higham. Analysis of the Cholesky decomposition of a semi-definite matrix. In *Reliable Numerical Computation*, pages 161–185, Oxford, 1990. Oxford University Press.

[Hig02] N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, 2nd edition, 2002.

[HJ94] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, 1994.

[HL89] J. Hoschek and D. Lasser. *Grundlagen der Geometrischen Datenverarbeitung*. Teubner, Stuttgart, 1989.

[HL13] H. Harbrecht and J. Li. First order second moment analysis for stochastic interface problems based on low-rank approximation. *ESAIM: Mathematical Modelling and Numerical Analysis*, 47:1533–1552, 2013.

[HM09] S. Heubach and T. Mansour. *Combinatorics of Compositions and Words*. CRC Press, Boca Raton, 2009.

[Hol73] R. B. Holmes. Smoothness of certain metric projections on Hilbert space. *Transactions of the American Mathematical Society*, 184:87–100, 1973.

[HP57] E. Hille and R. S. Phillips. *Functional Analysis and Semi-Groups*, volume 31. American Mathematical Society, Providence, 1957.

[HP13] H. Harbrecht and M. Peters. Comparison of fast boundary element methods on parametric surfaces. *Computer Methods in Applied Mechanics and Engineering*, 261–262:39–55, 2013.

[HPS12] H. Harbrecht, M. Peters, and R. Schneider. On the low-rank approximation by the pivoted Cholesky decomposition. *Applied Numerical Mathematics*, 62:28–440, 2012.

[HPS13a] H. Harbrecht, M. Peters, and M. Siebenmorgen. Combination technique based k-th moment analysis of elliptic problems with random diffusion. *Journal of Computational Physics*, 252:128–141, 2013.

[HPS13b] H. Harbrecht, M. Peters, and M. Siebenmorgen. Tractability of the quasi-monte carlo quadrature with Halton points for elliptic PDEs with random diffusion. *Preprint 2013-28, Mathematisches Institut, Universität Basel*, 2013.

[HPS14a] H. Harbrecht, M. Peters, and M. Siebenmorgen. Efficient approximation of random fields for numerical applications. *Preprint 2014-01, Mathematisches Institut, Universität Basel*, 2014.

[HPS14b] H. Harbrecht, M. Peters, and M. Siebenmorgen. Numerical solution of elliptic diffusion problems on random domains. *Preprint 2014-08, Mathematisches Institut, Universität Basel*, 2014.

[HR09] H. Harbrecht and M. Randrianarivony. Wavelet BEM on molecular surfaces: parametrization and implementation. *Computing*, 86(1):1–22, 2009.

[HR10] H. Harbrecht and M. Randrianarivony. From computer aided design to wavelet BEM. *Computing and Visualization in Science*, 13(2):69–82, 2010.

[HR11] H. Harbrecht and M. Randrianarivony. Wavelet BEM on molecular surfaces: solvent excluded surfaces. *Computing*, 92(4):335–364, 2011.

[HS06] H. Harbrecht and R. Schneider. Wavelet Galerkin schemes for boundary integral equations. Implementation and quadrature. *SIAM Journal on Scientific Computing*, 27(4):1347–1370, 2006.

[HSS08a] H. Harbrecht, R. Schneider, and C. Schwab. Multilevel frames for sparse tensor product spaces. *Numerische Mathematik*, 110(2):199–220, 2008.

[HSS08b] H. Harbrecht, R. Schneider, and C. Schwab. Sparse second moment analysis for elliptic problems in stochastic domains. *Numerische Mathematik*, 109(3):385–414, 2008.

[Kir93] A. Kirsch. The domain derivative and two applications in inverse scattering theory. *Inverse Problems*, 9(1):81–96, 1993.

[Kny97] A. Knyazev. New estimates for Ritz vectors. *Mathematics of Computation*, 66(219):985–995, 1997.

[KR86] R. V. Kadison and J. R. Ringrose. *Fundamentals of the Theory of Operator Algebras. V1: Elementary Theory.* Academic Press, New York, 1986.

[KR88] B. W. Kernighan and D. M. Ritchie. *The C Programming Language.* Prentice-Hall, Upper Saddle River, NJ, 1988.

[LC85] W. A. Light and E. W. Cheney. *Approximation Theory in Tensor Product Spaces.* Lecture Notes in Mathematics Volume 1169. Springer, Berlin-Heidelberg, 1985.

[Len86] M. Lenoir. Optimal isoparametric finite elements and error estimates for domains involving curved boundaries. *SIAM Journal on Numerical Analysis*, 23(3):562–580, 1986.

[LMWZ10] J. Li, J. M. Melenk, B. Wohlmuth, and J. Zou. Optimal a priori estimates for higher order finite elements for elliptic interface problems. *Applied Numerical Mathematics*, 60(1):19–37, 2010.

[Loè77] M. Loève. *Probability Theory. I+II.* Number 45 in Graduate Texts in Mathematics. Springer, New York, 4th edition, 1977.

[LS96] R. B. Lehoucq and D. C. Sorensen. Deflation techniques for an implicitly restarted Arnoldi iteration. *SIAM Journal on Matrix Analysis and Applications*, 17(4):789–821, 1996.

[LSY98] R. B. Lehoucq, D. C. Sorensen, and C. Yang. *Arpack User's Guide: Solution of Large-Scale Eigenvalue Problems with Implicityly Restorted Arnoldi Methods (Software, Environments, Tools).* Society for Industrial and Applied Mathematics, Philadelphia, 1998.

[Luc04] C. Lucas. LAPACK-style codes for level 2 and 3 pivoted Cholesky factorizations. *LAPACK Working*, 2004.

[Mat86] B. Matérn. *Spatial Variation.* Springer Lecture Notes in Statistics. Springer, New York, 1986.

[Mer09] J. Mercer. Functions of positive and negative type, and their connection with the theory of integral equations. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, pages 415–446, 1909.

[MK05] H. Matthies and A. Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 194(12-16):1295–1331, 2005.

[MNK11] P. S. Mohan, P. B. Nair, and A. J. Keane. Stochastic projection schemes for deterministic linear elliptic partial differential equations on random domains. *International Journal for Numerical Methods in Engineering*, 85(7):874–895, 2011.

[Mül98] C. Müller. *Analysis of Spherical Symmetries in Euclidean Spaces*. Number 129 in Applied Mathematical Sciences Series. Springer, Berlin-Heidelberg, 1998.

[Nie92] H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1992.

[Ovt06] E. Ovtchinnikov. Cluster robust error estimates for the Rayleigh–Ritz approximation I: Estimates for invariant subspaces. *Linear Algebra and its Applications*, 415(1):167–187, 2006.

[Pir82] O. Pironneau. Optimal shape design for elliptic systems. In R.F. Drenick and F. Kozin, editors, *System Modeling and Optimization*, volume 38 of *Lecture Notes in Control and Information Sciences*, pages 42–66. Springer, Berlin-Heidelberg, 1982.

[Pot94a] R. Potthast. Frechét differentiability of boundary integral operators in inverse acoustic scattering. *Inverse Problems*, 10(2):431–447, 1994.

[Pot94b] R. Potthast. *Frechét-Differenzierbarkeit von Randintegraloperatoren und Randwertproblemen zur Helmholtzgleichung und zu den zeitharmonischen Maxwellgleichungen*. PHD Thesis, Georg-August-Universität Göttingen, 1994.

[PZ99] C. Pflaum and A. Zhou. Error analysis of the combination technique. *Numerische Mathematik*, 84(2):327–350, 1999.

[Rei13] C. Reisinger. Analysis of linear difference schemes in the sparse grid combination technique. *IMA Journal of Numerical Analysis*, 33(2):544–581, 2013.

[RS80] M. Reed and B. Simon. *Methods of Modern Mathematical Physics I: Functional Analysis*, volume 1. Academic Press, San Diego, 1980.

[RW05] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, Cambridge, 2005.

[Saa92] Y. Saad. *Numerical Methods for Large Eigenvalue Problems*. Manchester University Press, Manchester, 1992.

[Sch50] R. Schatten. *A Theory of Cross-Spaces*. Princeton University Press, Princeton, 1950.

[Sch60] R. Schatten. *Norm Ideals of Completely Continuous Operators*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, Berlin-Göttingen-Heidelberg, 1960.

[Sch98] R. Schneider. *Multiskalen- und Wavelet-Matrixkompression: Analysisbasierte Methoden zur Lösung Großer Vollbesetzter Gleichungssysteme*. Teubner, Stuttgart, 1998.

[SG11] C. Schwab and C. J. Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numerica*, 20:291–467, 2011.

[Sim80] J. Simon. Differentiation with respect to the domain in boundary value problems. *Numerical Functional Analysis and Optimization*, 2(7-8):649–687, 1980.

[SLS03] G. Schmidlin, C. Lage, and C. Schwab. Rapid solution of first kind boundary integral equations in $\mathbb{R}^3$. *Engineering Analysis with Boundary Elements*, 27(5):469–490, 2003.

[Sor92] D. C. Sorensen. Implicit application of polynomial filters in a k-step Arnoldi method. *SIAM Journal on Matrix Analysis and Applications*, 13(1):357–385, 1992.

[SS97] S. A. Sauter and C. Schwab. Quadrature for hp-Galerkin BEM in $\mathbb{R}^3$. *Numerische Mathematik*, 78(2):211–258, 1997.

[SS11] S. A. Sauter and C. Schwab. *Boundary Element Methods*. Springer, Berlin-Heidelberg, 2011.

[ST03a] C. Schwab and R. A. Todor. Sparse finite elements for elliptic problems with stochastic loading. *Numerische Mathematik*, 95(4):707–734, 2003.

[ST03b] C. Schwab and R. A. Todor. Sparse finite elements for stochastic elliptic problems–higher order moments. *Computing*, 71(1):43–63, 2003.

[ST06] C. Schwab and R. Todor. Karhunen-Loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, 217:100–122, 2006.

[Ste03] O. Steinbach. *Numerische Näherungsverfahren für Elliptische Randwertprobleme: Finite Elemente und Randelemente*. Advances in Numerical Mathematics. Vieweg+Teubner, Stuttgart-Leipzig-Wiesbaden, 2003.

[SX95] T. Sauer and Y. Xu. On multivariate Lagrange interpolation. *Mathematics of Computation*, 64(211):1147–1170, 1995.

[SZ92] J. Sokolowski and J. P. Zolésio. *Introduction to Shape Optimization. Shape Sensitivity Analysis*. Springer series in computational mathematics. Springer, Berlin-Heidelberg, 1992.

[Tem86] V. N. Temlyakov. Approximations of functions with bounded mixed derivative. *Trudy Matematicheskogo Instituta im. VA Steklova*, 178:3–113, 1986.

[TX06] D. M. Tartakovsky and D. Xiu. Stochastic analysis of transport in tubes with rough walls. *Journal of Computational Physics*, 217(1):248–259, 2006.

[vPS06] T. von Petersdorff and C. Schwab. Sparse finite element methods for operator equations with stochastic data. *Applications of Mathematics*, 51(2):145–180, 2006.

[Wil06] H. S. Wilf. *Generatingfunctionology*. A. K. Peters, Ltd., Natick, MA, USA, 2006.

[Wlo82] J. Wloka. *Partielle Differentialgleichungen: Sobolevräume und Randwertaufgaben*. Teubner, Stuttgart, 1982.

[XT06] D. Xiu and D. M. Tartakovsky. Numerical methods for differential equations in random domains. *SIAM Journal on Scientific Computing*, 28(3):1167–1185, 2006.

[XZ04] Y. Xu and A. Zhou. Fast boolean approximation methods for solving integral equations in high dimensions. *Journal of Integral Equations and Applications*, 16(1):83–110, 2004.

[Zei11] A. Zeiser. Fast matrix-vector multiplication in the sparse-grid Galerkin method. *Journal of Scientific Computing*, 47(3):328–346, 2011.

CURRICULUM VITAE

Personal data

Name	Michael Dieter Peters
Date of birth	20. July 1984
Place of birth	Rheinbach, Germany
Nationality	German

Education

03/2011–11/2014	PhD position in Mathematics, *University of Basel*, Basel.
10/2010–02/2011	PhD position in Mathematics, *University of Stuttgart*, Stuttgart.
10/2004–09/2010	Diploma in Mathematics, *Rheinische Friedrich-Wilhelms-Universität*, Bonn.
08/1995–06/2004	Abitur, *Städtisches Gymnasium*, Rheinbach.
09/1991–07/1995	Primary School, *Katholische Grundschule Sankt Martin*, Rheinbach.