

UNIVERSITÄT
BAYREUTH

Kernel-based Approximation Methods with Applications in Uncertainty Quantification

Von der Universität Bayreuth
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
genehmigte Abhandlung

von
Lorenz Gollwitzer
aus Neustadt an der Waldnaab

1. Gutachter: Prof. Dr. Holger Wendland
2. Gutachter: Prof. Dr. Christian Rieger

Tag der Einreichung: 25. Juni 2026
Tag des Kolloquiums: 16. Oktober 2026

Acknowledgement

First, I want to thank my supervisor, Prof. Dr. Holger Wendland. He opened the possibility for me to not just write a doctoral thesis under his supervision, but additionally doing this in a 'fast-track' program. He gave me freedom to pursue my own ideas, but also provided guidance and feedback whenever needed. I am grateful for the time and effort he devoted to bring this thesis to a successful conclusion. I also want to thank Prof. Dr. Mihai Păun for giving me the possibility to finish this thesis. Without both, this thesis would not have been possible.
Moreover, I would like to thank Prof. Dr. Christian Rieger for agreeing to serve as the second examiner of this thesis. This work was partially conducted in the DFG project "Kernbasierte Multilevelverfahren für hochdimensionale Approximationsprobleme auf dünnen Gittern - Herleitung, Analyse und Anwendung in der Uncertainty Quantification" (project number: 452806809). I gratefully acknowledge the financial support provided.

I also would like to thank (in alphabetical order) my colleagues Daniel, Robert and Rüdiger, for all the fruitful, entertaining and encouraging discussions on both mathematical and non-math related topics. In particular, I want to thank Daniel, who put much effort in the proofreading of this thesis. His encouragement finally motivated me to finish, and I am very grateful to him for this. I also want to thank Naemi, who shared her profound knowledge of the english language with me and helped me to shape this thesis in a form that is (hopefully) understandable.

Of course, there are many people I have to thank, and it would exceed the scope of this acknowledgement to mention everyone I have in mind. Nonetheless, there are some people I want to thank separately:
At first, there are my probably longest and best friends, Adrian, Felix, Moritz and Niki. We know each other for most of our lives now, and they probably know me better than even I do. I cannot express how grateful and proud I am to call them my friends. Our regular get-togethers are always one of the highlights of the year.
I also especially want to thank Carola. She was the first one to read this thesis completly and corrected many mistakes I made. That aside, she was always there for me and had an open ear for me and my problems, gave me good advice, and I gladly remember the endless hours of talking and activities we pursued together. I am deeply grateful for her encouragement, our friendship and all the time we spent together.
Last but not least, I want to thank my parents and my sister. They always believed in me, helped me in every way possible, and encouraged me throughout all this time. They provided support for me in every way possible, yet brought me back down to earth whenever I needed it. This is the kind of family I wish for everyone, and I am very happy to have them in my life.

Abstract

In the thesis “Kernel-based Approximation Methods with Applications in Uncertainty Quantification”, we motivate the presented results by a partial differential equation that depends on random inputs, which we will therefore model using a random field.
In applications, this random field is usually not known everywhere on the spatial domain and must be replaced by a proper surrogate. We pursue this idea and try to approximate this random field with a pathwise approach using kernel-based approximation techniques. We begin by introducing basic function spaces in combination with tensor products, followed by theoretical foundations concerning approximation theory using kernels and the theory behind random fields.
In order to apply the reconstruction methods on realizations of the random field, we study the so-called sample path regularity, which connects regularity assumptions of the covariance function of the random field to the smoothness of its sample paths. Here, a slight generalization of the known theory has been achieved. Afterwards, we use this theory to transfer the deterministic approximation theory to random fields. We show that this approach is well-posed, i.e., we obtain a random field, for which the mean and covariance function can be stated and are themselves good approximations to the corresponding counterparts of the underlying random field. Inside that chapter we assume different settings, depending on the regularity of the covariance function and the paths of the random field.
Numerical considerations lead to another approximation scheme, the kernel-based multilevel method. We first introduce the method itself, before we proceed to prove new properties concerning the approximation space of the method and, in particular, a nodal representation. Subsequently, we apply this scheme on the sample paths of a random field, where the nodal representation from before allows us to obtain an explicit representation of the covariance function of the emerging random field.
Next, we present two different approaches to generalize the multilevel method on tensor product spaces. We analyze both methods separately, provide convergence results, state similar properties as in the untensorized case for both methods and compare them afterwards.
Then all presented methods will be applied to the introductory problem. We first establish well-posedness, i.e., the existence of a unique solution using the surrogate is shown, before we prove that the presented error estimates in the different scenarios lead to good approximations of the true solution as well.

Zusammenfassung

In dieser Abhandlung „Kernel-based Approximation Methods with Application in Uncertainty Quantification“ (Kernbasierte Approximationsmethoden mit Anwendung in der Quantifizierung von Unsicherheiten) betrachten wir als Motivation eine elliptische partielle Differentialgleichung, welche von zufälligen Eingabeparametern abhängt und mit Hilfe von einem stochastischen Prozess modelliert wird. In den meisten Anwendungen ist dieser Prozess jedoch nicht auf dem ganzen zugrundeliegenden Gebiet bekannt, weshalb dieser durch ein passendes Surrogat ersetzt werden muss. Unser Ziel ist es daher, diesen stochastischen Prozess durch kernbasierte Verfahren pfadweise zu approximieren.
Dazu führen wir zunächst die grundlegenden Funktionsräume im Zusammenhang mit Tensorprodukten ein. Danach folgen die wesentlichen Begriffe für kernbasierte Approximationsmethoden sowie die Grundlagen für stochastische Prozesse. Um die Kernmethoden auf die Pfade des Prozesses anwenden zu können, beschäftigen wir uns danach mit der Frage, inwiefern die Regularität der zum Prozess gehörigen Kovarianzfunktion auf die Pfade übertragen werden kann. Hier wurde eine kleine Verallgemeinerung der bisher bekannten Ergebnisse erzielt. Im nächsten Abschnitt wird dann die Wohldefiniertheit des pfadweisen Approximationsansatzes gezeigt. Insbesondere wird dort bewiesen, dass durch diesen Ansatz ebenfalls ein stochastischer Prozess entsteht, dessen Erwartungswert- und Kovarianzfunktion explizit angegeben werden kann und gute Approximationen an die jeweiligen Gegenstücke des ursprünglichen Prozesses sind. Danach übertragen wir die Fehlerabschätzungen des deterministischen Falles auf die vorliegende Situation, wobei wir nach verschiedenen Szenarien, abhängig von der Glätte der Pfade und der Kovarianzfunktion, unterscheiden.
Aus Gründen numerischer Stabilität wenden wir uns danach der kernbasierten Multilevelmethode zu, um die Pfade des Prozesses zu rekonstruieren. Zunächst stellen wir das Verfahren vor, bevor wir neue Erkenntnisse bezüglich des Approximationsraumes der Methode und insbesondere einer nodalen Basisdarstellung herleiten. Diese nodale Basisdarstellung wird insbesondere bei der Anwendung der Multilevelmethode auf stochastische Prozesse relevant, da sie uns erlaubt, eine explizite Darstellung für die entstehende Kovarianzfunktion des neuen Prozesses anzugeben.
Danach verallgemeinern wir die Multilevelmethode mit dem Tensorprodukt auf zwei verschiedene Arten. Zunächst werden beide Methoden separat analysiert, indem wir für beide Methoden Konvergenzresultate beweisen und ähnliche Eigenschaften wie im untensorierten Verfahren herleiten. Anschließend werden beide Methoden verglichen.
Schließlich werden die vorgestellten Methoden auf das einführende Problem angewendet, wobei wir zunächst die Wohldefiniertheit, sprich die Existenz einer Lösung der Differentialgleichung mit dem Surrogat sicherstellen, bevor wir mit Hilfe der vorgestellten Fehlertheorie folgern, dass dadurch auch die Lösung des Problems hinreichend gut angenähert werden kann.

Contents

1 Introduction **1**

2 Preliminaries **4**
2.1 Terminology and notation 4
2.2 Function spaces 5
2.2.1 Sobolev extension and embedding theorem 8
2.3 Bochner integral and Bochner spaces 9
2.3.1 Bochner integrability 10
2.3.2 Bochner spaces 12
2.4 Tensor products 15
2.4.1 Algebraic tensor product space 15
2.4.2 Algebraic tensor product of normed spaces 18
2.4.3 Examples of tensor product spaces 20
2.4.4 Tensor product of Hilbert spaces 23
2.4.5 Mixed regularity Sobolev spaces 23
2.4.6 Linear operators on tensor product spaces 25
2.4.7 Tensor product extension and embedding 26
2.4.8 Bochner and tensor product spaces 28

3 Kernel-based approximation **29**
3.1 Reproducing kernel Hilbert space 29
3.1.1 Mercer's theorem 35
3.2 Function approximation using kernels 36
3.2.1 Kernel-based interpolation 37
3.2.2 Penalized least squares approach 39
3.3 Error analysis 40
3.3.1 Convergence of kernel methods 45
3.4 Stability of the approximation process 47
3.5 Tensorizing approximations 49
3.5.1 Computational cost 51
3.5.2 Error estimates using tensorized kernels 53

4 Random Fields **57**
4.1 Properties of random fields 57
4.2 Karhunen-Loève expansion 61
4.2.1 Eigenvalue decay 67
4.3 Sample path regularity 68

5 Approximation of random fields **72**
5.1 Pathwise approximation 72

5.2 Kernel-based pathwise approximation 74
5.2.1 Error estimates 76
5.3 Karhunen-Loève expansion of the interpolated process 83
5.4 Covariance estimation 89

6 Kernel-based multilevel method 92
6.1 Motivation 92
6.2 Multilevel method – setup 95
6.3 Convergence of the multilevel method 97
6.4 Properties of the multilevel scheme 101
6.4.1 Representations in the general case 101
6.4.2 Representation of the multilevel scheme in the nested case 106
6.5 Application to random fields 116

7 Multilevel methods on mixed regularity spaces 119
7.1 Multilevel method with tensorized kernels 119
7.1.1 Properties of the method 130
7.1.2 Approximation of positive definite functions 134
7.2 Tensor product of multilevel methods 136
7.3 Comparison of the methods 138
7.3.1 Computational and evaluation cost 140

8 Application to PDEs with randomness 144
8.1 Formalizing the problem 144
8.2 Exchanging the diffusion coefficient 149
8.2.1 Parameterization of the differential equation 151
8.3 Approximating the solution 155

9 Outlook 160

Literature 162

1 Introduction

A very important task in various sciences, for example in physics or geostatics, is to simulate real-world behavior using mathematical models. Often, such models lead to partial differential equations, whose solution is intended to describe the underlying real-world process in a suitable manner. The aim is to integrate as many external influences as possible to obtain a model representing as much information as possible and therefore yielding a very accurate prediction. In reality, it might not be feasible to model all possible influences deterministically as such highly complex models are challenging to solve explicitly. Thus, a common approach is to make use of random inputs, which encode all possible unknown influences in the surrounding area.
The aim of *uncertainty quantification* is to make this randomness 'tractable', i.e., to provide algorithms to first simulate the given problem and then deduce statistical information and evidence from the simulation while maintaining numerical efficiency. It is, as [51, Section 1.1] puts it, the "coming together of probability theory and statistical practice with the real-world".
To make this concrete, we introduce the following typical example, which comes from geostatics and models the flow of water through a porous medium. It is practically infeasible to determine the porous medium explicitly, thus one models it using a *diffusion coefficient* a which depends on the underlying spatial domain as well as unknown parameters $\omega \in \Omega$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. Such a function is also called a *random field*. The partial differential equation can then be obtained by *Darcy's law* coupled with the mass conservation law, which leads to (cf. [35, Chapter 7] and [51, Example 1.1]):

$$\begin{aligned} q(\omega, x) + a(\omega, x)\nabla_x u(\omega, x) &= F(x), \quad x \in \mathcal{D}, \\ \operatorname{div}_x(q(\omega, x)) &= 0, \quad x \in \mathcal{D}. \end{aligned} \tag{1.1}$$

Here, $\mathcal{D} \subseteq \mathbb{R}^d$ with $d = 2$ or $d = 3$ denotes a bounded domain (the *spatial domain*), q denotes the *velocity* of the water, and u the *residual pressure*. As we assumed randomness of a, this induces randomness in q and u; thus, both are random fields as well. In order to formulate a well-posed differential equation, we have to impose boundary conditions on $\partial\mathcal{D}$, which will here be Dirichlet boundary conditions for simplicity.
We now reformulate (1.1): By applying the negative of the divergence operator to the first equation, we can use the mass conservation law of the velocity q to completely eliminate this parameter and obtain the differential equation

$$\begin{aligned} -\operatorname{div}_x\left(a(\omega, x)\nabla_x u(\omega, x)\right) &= f(x), \qquad x \in \mathcal{D}, \\ u(\omega, x) &= 0, \qquad x \in \partial\mathcal{D}, \end{aligned} \tag{1.2}$$

where $f(x) := -\operatorname{div} F(x)$ for all $x \in \mathcal{D}$. Our goal is now to find a function u that solves (1.2). Now (1.2) is an elliptic partial differential equation of second order with random inputs. Under minor assumptions on the random field a, (1.2) permits (weak) solvability

using well-known theory (cf. Chapter 8). That existence result on the other hand is non-constructive, which forces us to approximate the solution using, e.g., a Galerkin type approach using finite elements. Proceeding that way, many evaluations of the diffusion coefficient are needed. These may be hard to obtain, thus one seeks to construct a proper surrogate for the random field a. A common approach for this is to assume that a permits a series expansion $a(\omega, x) = m(x) + \sum_{j=1}^{\infty} \sqrt{\lambda_j} \xi_j(\omega) e_j(x)$, where $\omega \in \Omega$ and $x \in \mathcal{D}$ (usually the Karhunen-Loève expansion, see Corollary 4.2.5). This expansion then gets truncated after $n \in \mathbb{N}$ terms to obtain a random field that only depends on n real-valued random variables. In the literature, such a random field is also called *finite noise*. The next common step is to transform this random field into a parametric, but deterministic function using the push-forward measure of $(\xi_1, \dots, \xi_n)$ and to then solve the resulting parametrized partial differential equation with this random field as a surrogate. From this point on, one can deduce different statistical properties, such as a *quantity of interest* of the solution.
However, in reality, such expansions are hardly ever known explicitly and thus have to be approximated as well. While this can be done using deterministic methods in the spatial domain, the randomness seems far more delicate to handle. When considering the aforementioned Karhunen-Loève expansion, the distribution of the random variables ξ_j is usually unknown (except for Gaussian random fields, see Corollary 4.2.6), which forces us to additionally assume the distribution of these random variables (see for example [18]). Also, other spectral approaches such as the *(generalized) polynomial chaos approximation* (see [51, Section 11.3]) use knowledge of the underlying probability measure to construct orthogonal polynomials.
These issues motivate us to consider a different approach for choosing a surrogate for the diffusion coefficient a. In this thesis, specifically in Chapter 5 and Section 6.5, we will use *kernel-based approximation methods* to construct a surrogate for the random field a, which only uses information of the random field at sampling points $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$, with $N \in \mathbb{N}$ fixed, i.e., we assume the $\mathbb{R}^N$-valued random variable $(a(\cdot, x_1), \dots, a(\cdot, x_N))^T$ is known.
We will see that this approach inherently leads to a finite noise random field as well, yet the finitely many random variables are now given by $a(\cdot, x_1), \dots, a(\cdot, x_N)$. At this point, we can proceed as before, although it seems more feasible to assume that the common distribution of these random variables is known rather than to use the distributions emerging from the Karhunen-Loève or similar expansions of a. Additionally, we can use the Karhunen-Loève expansion (on top of the approximation process) to further reduce the dimensionality of the parametric problem. In our case, this expansion can now be calculated explicitly, as will be shown in Section 5.3. Another main advantage of this approach is that we also gain an explicit representation of the mean and covariance function of the surrogate random field, which are good approximations of the original mean and covariance function of a, respectively.

From an abstract point of view, the above can be established by transferring deterministic kernel-based approximation methods to random fields using a pathwise approach. Following this idea, the thesis is structured as follows: First, we introduce notation and fundamental results used in Chapter 2 by giving a short overview of basic function spaces, an introduction to Bochner integrability and Bochner spaces and a review of results for

tensor products and tensor product spaces.
In Chapter 3, we introduce reproducing kernel Hilbert spaces and present error estimates for specific kernel-based approximation methods, namely kernel-based interpolation and penalized least squares approach, both under certain regularity assumptions of the target function.
Following up, Chapter 4 provides a short overview of random fields in Section 4.1 and the already mentioned Karhunen-Loève expansion in Section 4.2. As we want to apply the deterministic error estimates on random fields, we conclude this chapter by stating results concerning the so-called *sample path regularity*, where we will prove a slight generalization of currently known results on fractional Sobolev spaces. Having done this, in Chapter 5 we connect Chapter 3 and Chapter 4 and provide error estimates for random fields and their resulting covariance functions in different scenarios.
Motivated by numerical considerations, we then go on in Chapter 6 with a kernel-based multiscale scheme, which is also known as the multilevel method. We provide novel representations (such as a Lagrange type representation) leading to new insights for this approximation scheme. Section 6.5 then finishes the chapter, where we proceed as in before and apply the deterministic multilevel method on random fields.
In Chapter 7, we consider two different approaches to generalize the multilevel method on mixed regularity Sobolev spaces. We present convergence results for both methods in Section 7.1 and Section 7.2, together with various properties of the methods, inspired by the ideas developed in Chapter 6, before we compare both methods in Section 7.3.
In Chapter 8, we head back to the introductory problem and apply the stated methods on the partial differential equation in a more rigorous manner. We present two common approaches that lead to weak formulations and state the underlying theory to prove that in both cases a unique solution exists. Having done this, we prove that under the right assumptions these solutions coincide.
Then we proceed as proposed before and prove that the differential equation with surrogate leads to a well-posed problem, provide error bounds for the solutions and give an exemplary strategy for tackling the emerging uncertainty.

2 Preliminaries

This chapter will be divided into three main parts: Section 2.1 and Section 2.2 introduce basic notations, followed by definitions and properties of several well-known function spaces. In Section 2.3, we provide a short overview of Bochner integrability and Bochner spaces, while Section 2.4 deals with the tensor product and tensor product spaces.

2.1 Terminology and notation

In this thesis, we denote with $\mathbb{N}$ the natural numbers (not including zero) and write explicitly $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. In the following, $d, n \in \mathbb{N}$ always denote dimensions and $\emptyset \neq \mathcal{D} \subseteq \mathbb{R}^d$ and $\emptyset \neq \mathcal{T} \subseteq \mathbb{R}^n$ are measurable, pathwise connected domains. Additional assumptions, such as boundedness or boundary conditions, will always be stated separately. Except in Section 2.3, we always assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. Generic constants $C > 0$ may differ from line to line. For norms on $\mathbb{R}^d$, we use the following convention:

Definition 2.1.1.
For $1 \leq p \leq \infty$, the ℓ^p norms on $\mathbb{R}^d$ are defined as

$$\|x\|_p = \begin{cases} \left(\sum_{j=1}^d |x_j|^p\right)^{\frac{1}{p}}, & 1 \leq p < \infty, \\ \max_{1 \leq j \leq d} |x_j|, & p = \infty. \end{cases}$$

For $p = 2$, this norm is also called the *Euclidean norm* on $\mathbb{R}^d$. When we consider a multi-index $\alpha \in \mathbb{N}_0^d$, we also write

$$|\alpha| := \|\alpha\|_1.$$

For Banach or Hilbert spaces, we denote the operator norm as follows:

Definition 2.1.2.
Let U and V be Banach spaces. We denote the space of all *linear and bounded* operators $A \colon U \to V$ by

$$\mathcal{L}(U, V) := \{A \colon U \to V \mid A \text{ linear and bounded}\}.$$

For an element $A \in \mathcal{L}(U, V)$, the *operator norm* of A is given by

$$\|A\|_{U \to V} := \sup_{\substack{u \in U \\ \|u\|_U \leq 1}} \|Au\|_V = \sup_{\substack{u \in U \\ \|u\|_U = 1}} \|Au\|_V.$$

If $V = \mathbb{R}$, we denote by $U^* := \mathcal{L}(U, \mathbb{R})$ the *dual space* of U. The elements of U^* are also called *functionals* on U.

2.2 Function spaces

An important space for us is the space of (classical) differentiable functions.

Definition 2.2.1.
The space of all *continuous* functions $f\colon \mathcal{D} \to \mathbb{R}$ is denoted by

$$C(\mathcal{D}) := \{f\colon \mathcal{D} \to \mathbb{R} \mid f \text{ continuous on } \mathcal{D}\}.$$

The subspace

$$C_b(\mathcal{D}) := \{f \in C(\mathcal{D}) \mid \|f\|_{C_b(\mathcal{D})} < \infty\},$$

of all bounded continuous functions on $\mathcal{D}$ equipped with the norm

$$\|f\|_{C_b(\mathcal{D})} := \sup_{x\in\mathcal{D}} |f(x)|$$

is a separable Banach space.
Let $k \in \mathbb{N}_0$ be arbitrary but fixed. A function $f\colon \mathcal{D} \to \mathbb{R}$ is said to be k-times *continuously differentiable* if all classical derivatives $D^\alpha f$ for $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq k$ exist and are additionally continuous. The corresponding space is defined accordingly as

$$C^k(\mathcal{D}) := \{f \in C(\mathcal{D}) \mid D^\alpha f \in C(\mathcal{D}), \alpha \in \mathbb{N}_0^d, |\alpha| \leq k\}.$$

Note that $C^0(\mathcal{D}) = C(\mathcal{D})$. The subspace

$$C_b^k(\mathcal{D}) := \{f \in C^k(\mathcal{D}) \mid D^\alpha f \in C_b(\mathcal{D}),\ \alpha \in \mathbb{N}_0^d,\ |\alpha| \leq k\},$$

of $C^k(\mathcal{D})$ equipped with the norm

$$\|f\|_{C_b^k(\mathcal{D})} := \max_{|\alpha|\leq k} \|D^\alpha f\|_{C_b(\mathcal{D})},$$

again yields a separable Banach space $\left(C_b^k(\mathcal{D}), \|\cdot\|_{C_b^k(\mathcal{D})}\right)$.
The space of the *infinitely often differentiable functions* is defined as

$$C^\infty(\mathcal{D}) := \bigcap_{k=1}^{\infty} C^k(\mathcal{D}),$$

and $C_c^\infty(\mathcal{D}) \subseteq C^\infty(\mathcal{D})$ consists of all functions $f \in C^\infty(\mathcal{D})$ whose *support*

$$\operatorname{supp}(f) := \overline{\{x \in \mathcal{D} \mid f(x) \neq 0\}}$$

is compact.

Next, we define intermediate spaces between $C_b^k(\mathcal{D})$ and $C_b^{k+1}(\mathcal{D})$, the so-called *Hölder spaces*.

Definition 2.2.2 (Hölder space)**.**
Let $k \in \mathbb{N}_0$ and $0 < \gamma < 1$. For $f\colon \mathcal{D} \to \mathbb{R}$, define

$$[f]_\gamma := \sup_{\substack{x,y\in\mathcal{D}\\ x\neq y}} \frac{|f(x) - f(y)|}{\|x - y\|_2^\gamma} \in [0, \infty].$$

Then $C^{(k,\gamma)}(\mathcal{D}) \subseteq C_b^k(\mathcal{D})$ consists of all functions $f \in C_b^k(\mathcal{D})$ such that the following norm is finite:

$$\|f\|_{C^{(k,\gamma)}(\mathcal{D})} := \sum_{\substack{\alpha \in \mathbb{N}_0^d \\ |\alpha| \leq k}} \|D^\alpha f\|_{C_b(\mathcal{D})} + \sum_{\substack{\alpha \in \mathbb{N}_0^d \\ |\alpha| = k}} [D^\alpha f]_\gamma \,.$$

We call the space $C^{(k,\gamma)}(\mathcal{D})$ *Hölder space* and the functions $f \in C^{(k,\gamma)}(\mathcal{D})$ *Hölder continuous.*

Remark.
For $\gamma = 0$, we define the Hölder spaces $C^{k,0}(\mathcal{D}) := C^k(\mathcal{D})$ as the usual spaces of differentiable functions; thus, we set

$$C^s(\mathcal{D}) := \begin{cases} C_b^k(\mathcal{D}), & s = k \in \mathbb{N}_0, \\ C^{(k,\gamma)}(\mathcal{D}), & s = k + \gamma,\ k \in \mathbb{N}_0,\ 0 < \gamma < 1. \end{cases}$$

For every $s \geq 0$, the inclusion $C^s(\mathcal{D}) \subseteq C_b(\mathcal{D})$ holds. Whenever there is no confusion about the finiteness of this norm, we suppress the boundedness in the following and just write $\|\cdot\|_{C(\mathcal{D})} := \|\cdot\|_{C_b(\mathcal{D})}$.

Whereas we previously classified the function spaces based on their classical differentiability, we also consider spaces classified by *weak differentiability* (see for example [15, Chapter 5.2]). This leads to *Sobolev spaces.*

Definition 2.2.3 (L^p and Sobolev-space)**.**
Let $1 \leq p \leq \infty$. We denote the L^p (semi-)norms on $\mathcal{D}$ with

$$\|f\|_{L^p(\mathcal{D})} := \begin{cases} \left(\int_{\mathcal{D}} |f(x)|^p \, dx\right)^{\frac{1}{p}}, & 1 \leq p < \infty, \\ \sup\left\{C \geq 0 \mid |f(x)| \leq C \text{ almost everywhere on } \mathcal{D}\right\}, & p = \infty. \end{cases}$$

We then call

$$L^p(\mathcal{D}) := \{f \colon \mathcal{D} \to \mathbb{R} \mid \|f\|_{L^p(\mathcal{D})} < \infty\}$$

the *Lebesgue spaces* on $\mathcal{D}$.
Now let $k \in \mathbb{N}$ be arbitrary. The spaces

$$W^{k,p}(\mathcal{D}) := \left\{ f \in L^p(\mathcal{D}) \mid D^\alpha f \in L^p(\mathcal{D}), \alpha \in \mathbb{N}_0^d, |\alpha| \leq k \right\},$$

are called the (k,p) *Sobolev spaces.* We define the $W^{k,p}$ semi-norm as

$$|f|_{W^{k,p}(\mathcal{D})} := \left(\sum_{|\alpha| = k} \|D^\alpha f\|_{L^p(\mathcal{D})}^p \right)^{\frac{1}{p}},$$

which induces the norm

$$\|f\|_{W^{k,p}(\mathcal{D})} := \sum_{\ell=1}^{k} \left(|f|_{W^{\ell,p}(\mathcal{D})}^p \right)^{\frac{1}{p}} = \left(\sum_{|\alpha| \leq k} \|D^\alpha f\|_{L^p(\mathcal{D})}^p \right)^{\frac{1}{p}},$$

where in both cases $D^\alpha f$ denotes the α-th weak derivative of f. Identifying all functions f which are equal almost everywhere on $\mathcal{D}$, $\|\cdot\|_{L^p(\mathcal{D})}$ and $\|\cdot\|_{W^{k,p}(\mathcal{D})}$ are indeed norms and the corresponding spaces are Banach spaces. For $1 \leq p < \infty$, these are separable, and, for $p = 2$, they are additionally Hilbert spaces. In this case, we also write $H^k(\mathcal{D}) := W^{k,2}(\mathcal{D})$.

Usually, one first defines an equivalence relation on the functions $f\colon \mathcal{D} \to \mathbb{R}$ before introducing L^p spaces. For the sake of brevity, we assume the formally correct procedure is known.
The following subspace of $H^1(\mathcal{D})$ will be crucial in Chapter 8.

Definition 2.2.4.
The space $H_0^1(\mathcal{D})$ is defined as

$$H_0^1(\mathcal{D}) := \overline{C_c^\infty(\mathcal{D})}^{\|\cdot\|_{H^1(\mathcal{D})}} \subseteq H^1(\mathcal{D}).$$

One can interpret $H_0^1(\mathcal{D})$ as the functions $f \in H^1(\mathcal{D})$ that *vanish* on the boundary of $\mathcal{D}$. We can define an inner product on $H_0^1(\mathcal{D})$ by

$$\langle u, v\rangle_{H_0^1(\mathcal{D})} := \int_{\mathcal{D}} \nabla u(x) \cdot \nabla v(x)\, dx, \quad u, v \in H_0^1(\mathcal{D}).$$

The induced norm $\|\cdot\|_{H_0^1(\mathcal{D})}$ is also called the *energy norm*. By the Poincarè inequality (see [15, §5.6.1. Theorem 3]) it defines an equivalent norm to the inherited norm $\|\cdot\|_{H^1(\mathcal{D})}$ on $H_0^1(\mathcal{D})$.

With the *Sobolev-Slobodeckij spaces*, we introduce, similar to Hölder spaces, Sobolev spaces for non-integer order. We follow the definition of [56, Chapter 11.6].

Definition and Remark 2.2.5 (Sobolev-Slobodeckij)**.**
Let $1 \le p < \infty$, $k \in \mathbb{N}_0$, $0 < s < 1$ and $\sigma := k + s$. Then the *Sobolev-Slobodeckij space* $W^{\sigma,p}(\mathcal{D})$ is defined as

$$W^{\sigma,p}(\mathcal{D}) := \left\{ u \in W^{k,p}(\mathcal{D}) \mid \|u\|_{W^{\sigma,p}(\mathcal{D})} < \infty \right\},$$

where

$$|u|^p_{W^{\sigma,p}(\mathcal{D})} := \sum_{|\alpha|=k} \int_{\mathcal{D}} \int_{\mathcal{D}} \frac{|D^\alpha u(x) - D^\alpha u(y)|^p}{\|x-y\|^{d+ps}}\, dy\, dx$$

and

$$\|u\|_{W^{\sigma,p}(\mathcal{D})} := \left(\|u\|^p_{W^{k,p}(\mathcal{D})} + |u|^p_{W^{\sigma,p}(\mathcal{D})} \right)^{\frac{1}{p}}.$$

For $s = 0$, we use the usual Sobolev space $W^{k,p}(\mathcal{D})$ as defined before. These are separable Banach spaces for all $1 < p < \infty$, and, for $p = 2$, also Hilbert spaces, which will be denoted as before with $H^\sigma(\mathcal{D}) := W^{\sigma,2}(\mathcal{D})$.

For $p = 2$ and $\mathcal{D} = \mathbb{R}^d$, we can use the Fourier transform to define an equivalent norm on this space. Here, we use the same convention as [56, Definition 5.15].

Definition and Remark 2.2.6 (Fourier transform)**.**

1. Let $f \in L^1(\mathbb{R}^d)$. The *Fourier transform* of f is defined as

$$\widehat{f}(\omega) := (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} e^{-ix^T\omega} f(x)\, dx, \quad \omega \in \mathbb{R}^d.$$

2. Using density of $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d) \subseteq L^2(\mathbb{R}^d)$, the Fourier transform can be extended to a linear and bounded operator $\mathfrak{F}\colon L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$. Moreover, for all functions $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ the equality $\mathfrak{F}(f) = \widehat{f}$ holds. Thus, we again write $\mathfrak{F}(f) := \widehat{f}$ for all $f \in L^2(\mathbb{R}^d)$ but note that in general the formula above does not hold for $f \in L^2(\mathbb{R}^d) \setminus L^1(\mathbb{R}^d)$.

3. For all $f \in L^2(\mathbb{R}^d)$, the equation $\|f\|_{L^2(\mathbb{R}^d)} = \|\widehat{f}\|_{L^2(\mathbb{R}^d)}$ holds, which is also known as the *Plancherel theorem*.

Theorem 2.2.7 (Equivalent description of $H^\sigma(\mathbb{R}^d)$)**.**
Let $\sigma \geq 0$. Then the norm

$$\|u\|^2_{H^\sigma(\mathbb{R}^d)} := \int_{\mathbb{R}^d} \left|\widehat{f}(\omega)\right|^2 \left(1 + \|\omega\|_2^2\right)^\sigma \, d\omega$$

is equivalent to the norm on $H^\sigma(\mathbb{R}^d)$ previously defined in Definition and Remark 2.2.5.

Proof. The statement can be found in [2, Chapter 14] or [53, Chapter 2.3.3]. □

2.2.1 Sobolev extension and embedding theorem

Since we have a particularly nice expression for the Sobolev spaces of fractional order on $\mathbb{R}^d$, we seek a way to *extend* functions on a particular domain $\mathcal{D} \subseteq \mathbb{R}^d$ to this space. Unfortunately, in order to state such a result, we have to impose conditions on the boundary of $\mathcal{D}$ (see [9, Definition 1.4.4]).

Definition 2.2.8 (Lipschitz boundary)**.**
Let $\mathcal{D} \subseteq \mathbb{R}^d$ with $d \geq 2$. We call the boundary $\partial\mathcal{D}$ a *Lipschitz boundary* if there exists an arbitrary index set I, a family of open subsets $(U_i)_{i\in I} \subseteq \mathbb{R}^d$, constants $\varepsilon, M > 0$ and an integer $N \in \mathbb{N}$ such that the following statements hold:

- For all $x \in \partial\mathcal{D}$, there exists an $i \in I$ such that $B_\varepsilon(x) \subseteq U_i$;
- For every $i \in I$, there are at most N sets $U_{i_1}, \ldots, U_{i_N}$ such that, for all $1 \leq j \leq N$, the intersection $U_i \cap U_{i_j} \neq \emptyset$;
- For every $i \in I$, there is a domain $\mathcal{D}_i \subseteq \mathbb{R}^d$ such that $U_i \cap \mathcal{D} = U_i \cap \mathcal{D}_i$, and the boundary of $\mathcal{D}_i$ is a graph of a Lipschitz function φ_i, i.e.,

$$\mathcal{D}_i = \{(x, y) \in \mathbb{R}^d \mid x \in \mathbb{R}^{d-1}, y < \varphi_i(x)\}$$

  and the *Lipschitz norm* of φ_i fulfills

$$\|\varphi_i\|_{C_b(\mathcal{D})} + \sup_{\substack{x,y\in\mathcal{D}\\ x\neq y}} \frac{|\varphi_i(x) - \varphi_i(y)|}{\|x - y\|_2} \leq M.$$

We can now state the *Sobolev extension theorem* on such domains.

Theorem 2.2.9 (Extension operator)**.**
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a domain with Lipschitz boundary. Then there exists a linear operator $E\colon L^2(\mathcal{D}) \to L^2(\mathbb{R}^d)$ that, for all $\sigma \geq 0$, fulfills the following properties:

(i) $E_{|H^\sigma(\mathcal{D})}\colon H^\sigma(\mathcal{D}) \to H^\sigma(\mathbb{R}^d)$ is linear and bounded;

(ii) For all $f \in H^\sigma(\mathcal{D})$, it holds that $(E_{|H^\sigma(\mathcal{D})}(f))_{|\mathcal{D}} = f$, i.e., the extension of f restricted to $\mathcal{D}$ is again f.

We call E the (universal) extension operator *from $H^\sigma(\mathcal{D})$ to $H^\sigma(\mathbb{R}^d)$. As the operator is the same for all $\sigma \geq 0$, we suppress the index $H^\sigma(\mathcal{D})$ and write $Ef := E_{|H^\sigma(\mathcal{D})}(f)$ for $f \in H^\sigma(\mathcal{D})$.*

Proof. For $\sigma = k \in \mathbb{N}_0$, the theorem is stated in [9, Theorem 1.4.5]. The most general form can be found in [45, Theorem 5.8], where this statement is proven for Besov spaces $B^r_{p,q}(\mathcal{D})$. However, for non-integer $\sigma > 0$, we have $H^\sigma(\mathcal{D}) = B^\sigma_{2,2}(\mathcal{D})$ (cf. [43, Chapter 2.1.2]) and thus the claim. □

Another classical result concerning Sobolev spaces is the so-called *Sobolev embedding theorem*. The theorem specifically embeds these spaces into Hölder spaces and thus allows us to conclude classical differentiability up to a specific order. For us, the case $p = 2$ will be sufficient.

Theorem 2.2.10 (Sobolev embedding)**.**
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a domain with Lipschitz boundary and $\sigma > \frac{d}{2}$. Then for all $0 \leq s < \sigma - \frac{d}{2}$ the inclusion

$$\iota\colon H^\sigma(\mathcal{D}) \to C^s(\mathcal{D}), \quad f \mapsto f$$

is linear and bounded in the sense that there exists a representative of $f \in H^\sigma(\mathcal{D})$ which is an element of $C^s(\mathcal{D})$. We also say that $H^\sigma(\mathcal{D}) \hookrightarrow C^s(\mathcal{D})$ is continuously embedded*.*

Proof. See [25, Theorem 3.32, Exercise 3.34] for the case $\mathcal{D} = \mathbb{R}^d$. The general result follows by Theorem 2.2.9. □

2.3 Bochner integral and Bochner spaces

In this section, we extend the classical theory for integration of real-valued functions to Banach space-valued functions, as this will be important later on for Banach space-valued random variables. Although we will use the following results only for a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, there is no advantage of restricting ourselves to this case. Therefore, we assume throughout this section that $(\Omega, \mathcal{F}, \mu)$ is a measure space and B a separable Banach space which is endowed with the Borel-σ-Algebra

$$\mathcal{B}(B) := \sigma\left(\{A \subseteq B \mid A \text{ open}\}\right).$$

The following results can be found in [26, Chapter 1].

2.3.1 Bochner integrability

As we will see, the basic idea of the Bochner integral is analogous to the integration of real-valued functions. We start by defining *measurability* of a function.

Definition 2.3.1 (Measurable function)**.**
A function $f\colon \Omega \to B$ is called $(\mathcal{F}, \mathcal{B}(B))$-*measurable* if

$$f^{-1}(A) := \{x \in \Omega \mid f(x) \in A\} \in \mathcal{F}$$

for all $A \in \mathcal{B}(B)$. We then denote by

$$\sigma(f) := \left\{f^{-1}(A) \mid A \in \mathcal{B}(B)\right\} \subseteq \mathcal{F}$$

the σ-algebra generated by f.
If the choices of $\mathcal{F}$ and $\mathcal{B}(B)$ are clear from the context, we simply say that f is *measurable*.

We emphasize that we always assume a *separable* Banach space B. When we drop the separability, one can show that the σ-algebra

$$\sigma(B^*) := \sigma\left(\left\{x^{-1}(M) \mid x \in B^*, M \in \mathcal{B}(\mathbb{R})\right\}\right)$$

created by all linear functionals on B is a subset of the Borel σ-algebra on B, but in general a proper one. For separable B it is shown in [26, Proposition 1.1.1] that these two σ-algebras coincide.
To define an integral for such functions, we proceed as in the real-valued case using step functions (see [26, Definition 1.1.13, Definition 1.1.14]).

Definition 2.3.2.

1. A function $f\colon \Omega \to B$ is called μ-*simple* if there exists a natural number $n \in \mathbb{N}$, coefficients $b_1, \dots, b_n \in B$, and sets $A_1, \dots, A_n \in \mathcal{F}$ with $\mu(A_j) < \infty$ such that f can be written as
$$f(t) = \sum_{j=1}^{n} 1_{A_j}(t) b_j \in B, \quad t \in \Omega,$$
where for a set $A \subseteq B$ the indicator function 1_A is defined as
$$1_A\colon B \to \mathbb{R}, \quad x \mapsto \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

2. An arbitrary function $f\colon \Omega \to B$ is called μ-*measurable* (with values in B) if there exists a sequence of μ-simple functions $(f_n)_{n\in\mathbb{N}}$ such that for all $t \in \Omega$ the sequence $(f_n(t))_{n\in\mathbb{N}} \subseteq B$ converges pointwise to $f(t)$.

For a μ-measurable function, the sequence $(f_n)_{n\in\mathbb{N}}$ can be selected such that the norms of f_n are uniformly bounded by the norm of f ([26, Corollary 1.1.21]).

Corollary 2.3.3.
Let $f\colon \Omega \to B$ be μ-measurable. Then there exists a sequence of μ-simple functions $(f_n)_{n\in\mathbb{N}}$ such that for all $n \in \mathbb{N}$ and $t \in \Omega$ the bound $\|f_n(t)\|_B \leq \|f(t)\|_B$ holds and $(f_n(t))_{n\in\mathbb{N}} \subseteq B$ converges to $f(t)$.

Now the two concepts of measurability can be connected ([26, Proposition 1.1.16]).

Proposition 2.3.4.
Let $f\colon \Omega \to B$.

1. *If f is μ-measurable, then f is equal to a measurable function μ-almost everywhere.*
2. *If μ is σ-finite and f is equal to a measurable function μ-almost everywhere, then f is also μ-measurable.*

We note that the preceding corollary explicitly uses that B is separable, as in this case the *strong measurability* (cf. [26, Definition 1.1.4]) and the measurability of Definition 2.3.1 coincide.
We now use μ-measurability to define the Bochner integral ([26, Definition 1.2.1]).

Definition 2.3.5 (Bochner integral)**.**

1. For a μ-simple function $f(t) = \sum_{j=1}^{n} 1_{A_j}(t) b_j$, we set
$$\int_\Omega f d\mu := \sum_{j=1}^{n} \mu(A_j) b_j \in B.$$
2. Let $f\colon \Omega \to B$ be μ-measurable. Then f is called *Bochner integrable (with respect to μ)* if there is a sequence of μ-simple functions $(f_n)_{n\in\mathbb{N}}$ such that
$$\lim_{n\to\infty} \int_\Omega \|f(t) - f_n(t)\|_B \, d\mu(t) = 0.$$
In this case, we define the *Bochner integral of f* as
$$\int_\Omega f(t)\, d\mu(t) := \lim_{n\to\infty} \int_\Omega f_n(t)\, d\mu(t).$$

This definition uses that $\|f - f_n\|_B$ is integrable as a real-valued function with respect to μ. In [26, Proposition 1.2.2], we can find an equivalent characterization of the Bochner integrability.

Proposition 2.3.6.
A μ-measurable function $f\colon \Omega \to B$ is Bochner integrable with respect to μ if and only if
$$\int_\Omega \|f(t)\|_B \, d\mu(t) < \infty$$
as a real-valued integral. In this case, the bound
$$\left\| \int_\Omega f(t)\, d\mu(t) \right\|_B \leq \int_\Omega \|f(t)\|_B \, d\mu(t) < \infty$$
holds.

Next, we observe that the Bochner integral can be interchanged with linear and bounded operators. This follows directly by definition, but is also a special case of [26, Theorem 1.2.4].

Corollary 2.3.7.
Let B and B' be two separable Banach spaces and $T\colon B \to B'$ a linear and bounded operator. Then, for every Bochner integrable function $f\colon \Omega \to B$, the concatenation of f with T is also Bochner integrable with respect to μ, and it holds that

$$T\left(\int_\Omega f(t)\,d\mu(t)\right) = \int_\Omega (T\circ f)(t)\,d\mu(t).$$

For more details on rules using the Bochner integral, we again refer to [26, Chapter 1].

2.3.2 Bochner spaces

Proposition 2.3.6 motivates us to define the so-called *Bochner spaces* analogously to the L^p spaces in the real-valued case (see [26, Definition 1.2.15]).

Definition 2.3.8 (Bochner space)**.**
Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, B a separable Banach space, and $1 \le p < \infty$.

1. We call two μ-measurable functions $f, g\colon \Omega \to B$ *equivalent* if there exists a set $N \in \mathcal{F}$ of measure $\mu(N) = 0$ such that $f = g$ on $\Omega \setminus N$. This defines an equivalence relation on the μ-measurable functions. As usual, we identify the equivalence classes with their representative and note that the Bochner integrals of two representatives of the same equivalence class are equal.

2. We define the *Bochner space* $L^p(\Omega, \mu; B)$ as the linear space of (the equivalence classes of) all μ-measurable functions $f\colon \Omega \to B$ such that

$$\int_\Omega \|f(t)\|_B^p\,d\mu(t) < \infty.$$

   This space endowed with the norm

$$\|f\|_{L^p(\Omega,\mu;B)} := \left(\int_\Omega \|f(t)\|_B^p\,d\mu(t)\right)^{\frac{1}{p}}, \quad f \in L^p(\Omega,\mu;B),$$

   is again a Banach space. When there is no confusion, we suppress the measure μ in our notation and write $L^2(\Omega; B) = L^2(\Omega, \mu; B)$. Additionally, for $B = \mathbb{R}$, we write $L^p(\Omega, \mu) := L^p(\Omega, \mu; \mathbb{R})$. If $\mu = \lambda$ is the Lebesgue measure on $\mathbb{R}$, we simply obtain the L^p spaces defined in Definition 2.2.3 and identify $L^p(\Omega, \lambda, \mathbb{R}) = L^p(\Omega)$.

Of course, one can also define the space $L^\infty(\Omega; B)$, but we will work mostly with the spaces $L^1(\Omega; B)$ and $L^2(\Omega; B)$ with either $B = C_b(\mathcal{D})$ or $B = H^\sigma(\mathcal{D})$ for a bounded domain $\mathcal{D} \subseteq \mathbb{R}^d$.
In this case, we also obtain a Hölder type estimate.

Lemma 2.3.9.
Let $(B_i, \|\cdot\|_{B_i})$ for $i = 1, 2, 3$ be separable Banach spaces and $\varphi\colon B_1 \times B_2 \to B_3$ a bounded bilinear form (cf. Definition 2.4.1) such that

$$\|\varphi(f,g)\|_{B_3} \leq \|\varphi\|\|f\|_{B_1}\|g\|_{B_2}$$

holds, where

$$\|\varphi\| := \sup_{\|f\|_{B_1}, \|g\|_{B_2} \leq 1} \|\varphi(f,g)\|_{B_3}.$$

Let $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$ and functions $f \in L^p(\Omega; B_1)$ and $g \in L^q(\Omega; B_2)$ be given. Then $\varphi(f,g)$ is an element of $L^1(\Omega; B_3)$, and the estimate

$$\|\varphi(f,g)\|_{L^1(\Omega;B_3)} \leq \|\varphi\|\|f\|_{L^p(\Omega;B_1)}\|g\|_{L^q(\Omega;B_2)}$$

holds.

Proof. First, let $f = \sum_{j=1}^n 1_{A_j} f_j \in L^p(\Omega; B_1)$ and $g = \sum_{k=1}^m 1_{C_k} g_k \in L^q(\Omega; B_2)$ be two simple functions. Then

$$\varphi(f(t), g(t)) = \sum_{j=1}^n \sum_{k=1}^m 1_{A_j}(t) 1_{C_k}(t) \varphi(f_j, g_k)$$

is again a simple function, hence integrable. If now $f \in L^p(\Omega; B_1)$ and $g \in L^q(\Omega; B_2)$ are both μ-measurable, then there are two sequences of simple functions $(f_j)_{j\in\mathbb{N}} \subseteq L^p(\Omega; B_1)$ and $(g_j)_{j\in\mathbb{N}} \subseteq L^q(\Omega; B_2)$ that converge pointwise to f and g, respectively. Additionally, we can apply Corollary 2.3.3 and choose the sequence $(f_j)_{j\in\mathbb{N}}$ in such a manner that for all $t \in \Omega$ the bound $\|f_j(t)\|_{B_1} \leq \|f\|_{B_1}$ holds. Then

$$\begin{aligned}
&\|\varphi(f_j(t), g_j(t)) - \varphi(f(t), g(t))\|_{B_3} \\
&\leq \|\varphi(f_j(t), g_j(t) - g(t))\|_{B_3} + \|\varphi(f_j(t) - f(t), g(t))\|_{B_3} \\
&\leq \|\varphi\| \left(\|f_j(t)\|_{B_1}\|g_j(t) - g(t)\|_{B_2} + \|f_j(t) - f(t)\|_{B_1}\|g(t)\|_{B_2}\right) \\
&\leq \|\varphi\| \left(\|f(t)\|_{B_1}\|g_j(t) - g(t)\|_{B_2} + \|f_j(t) - f(t)\|_{B_1}\|g(t)\|_{B_2}\right)
\end{aligned}$$

and the right-hand side tends to 0 for $j \to \infty$, hence $\varphi(f,g)$ is again μ-measurable.
As φ is a bounded bilinear form, we see that

$$\begin{aligned}
\int_\Omega \|\varphi(f(t), g(t))\|_{B_3}\, d\mu(t) &\leq \|\varphi\| \int_\Omega \|f(t)\|_{B_1}\|g(t)\|_{B_2}\, d\mu(t) \\
&\leq \|\varphi\| \left(\int_\Omega \|f(t)\|_{B_1}^p\, d\mu(t)\right)^{\frac{1}{p}} \left(\int_\Omega \|g(t)\|_{B_2}^q\, d\mu(t)\right)^{\frac{1}{q}} \\
&= \|\varphi\|\|f\|_{L^p(\Omega;B_1)}\|g\|_{L^q(\Omega;B_2)},
\end{aligned}$$

using Hölder inequality in $L^1(\Omega, \mu)$.
This implies that the mapping $\Omega \to B_3, t \mapsto \varphi(f(t), g(t))$ is integrable and also yields the stated estimate. □

We now apply this lemma in several settings.

Corollary 2.3.10.

1. *For all $f, g \in L^2(\Omega; L^2(\mathcal{D}))$ we have*

$$\|fg\|_{L^1(\Omega;L^1(\mathcal{D}))} \leq \|f\|_{L^2(\Omega;L^2(\mathcal{D}))}\|g\|_{L^2(\Omega;L^2(\mathcal{D}))}.$$

2. *For all $f, g \in L^2(\Omega; C_b(\mathcal{D}))$*

$$\|fg\|_{L^1(\Omega;C_b(\mathcal{D}))} \leq \|f\|_{L^2(\Omega;C_b(\mathcal{D}))}\|g\|_{L^2(\Omega;C_b(\mathcal{D}))}$$

   holds.

This implies that in both cases the Bochner integral

$$\int_\Omega f(t)g(t)\,d\mu(t)$$

in the corresponding space exists.

Proof. First, we know that both $L^2(\mathcal{D})$ and $C_b(\mathcal{D})$ are separable Banach spaces. We choose the bilinear forms

$$\varphi\colon L^2(\mathcal{D}) \times L^2(\mathcal{D}) \to L^1(\mathcal{D}), \quad (f, g) \mapsto fg,$$

and

$$\varphi\colon C_b(\mathcal{D}) \times C_b(\mathcal{D}) \to C_b(\mathcal{D}), \quad (f, g) \mapsto fg,$$

where, in both cases, the usual multiplication of functions is used. One can easily check that $\|\varphi\| = 1$ holds in both cases, hence Lemma 2.3.9 provides the claim. □

Corollary 2.3.11.
Assume that μ is a finite measure on Ω (i.e., $\mu(\Omega) < \infty$), B is a separable Banach space and $1 \leq p < q < \infty$. Then, $L^q(\Omega; B) \subseteq L^p(\Omega; B)$, and we obtain the bound

$$\|f\|_{L^p(\Omega;B)} \leq \mu(\Omega)^{\frac{1}{p}-\frac{1}{q}}\|f\|_{L^q(\Omega;B)}.$$

Proof. The proof is a simple application of the real-valued case with the function $\|f(\cdot)\|_B$. □

2.4 Tensor products

We now give a short overview of the most important concepts concerning tensor products used in this thesis. Most of the results presented here can be found in [21], [34], [44] and [24], which also provide a more in depth discussion of tensor products. In the following, we always use $\mathbb{R}$ as the underlying field of the occurring linear spaces.

2.4.1 Algebraic tensor product space

We start with the definition of bilinearity.

Definition 2.4.1 (Bilinear form)**.**
Let V_1, V_2 and T be linear spaces. A mapping $\varphi\colon V_1 \times V_2 \to T$ is called a *bilinear form* if it is linear in both arguments, i.e., for all $u_1, u_2 \in V_1$, $v_1, v_2 \in V_2$ and $\alpha, \beta \in \mathbb{R}$

$$\begin{aligned}\varphi(u_1 + \alpha u_2, v_1) &= \varphi(u_1, v_1) + \alpha\varphi(u_2, v_1),\\ \varphi(u_1, v_1 + \beta v_2) &= \varphi(u_1, v_1) + \beta\varphi(u_1, v_2)\end{aligned}$$

holds. The subspace

$$\operatorname{Im}(\varphi) := \operatorname{span}\{\varphi(v_1, v_2) \mid v_1 \in V_1, v_2 \in V_2\} \subseteq T$$

is called the *image* of φ.

Remark 2.4.2.
Note that $\varphi(V_1, V_2)$ is not a linear space in general, which forces us to consider the linear span. Also, per definition, every element $v \in \operatorname{Im}(\varphi)$ is given as a linear combination

$$v = \sum_{j=1}^{N} \alpha_j \varphi(v_1^{(j)}, v_2^{(j)}),$$

where $\alpha_j \in \mathbb{R}, v_1^{(j)} \in V_1$, and $v_2^{(j)} \in V_2$ for $1 \leq j \leq N$. Using the linearity in the first argument, we set $\widetilde{v}_1^{(j)} := \alpha_j v_1^{(j)}$ and rewrite v as

$$v = \sum_{j=1}^{N} \varphi(\widetilde{v}_1^{(j)}, v_2^{(j)}).$$

Thus, without loss of generality, we can assume that $\alpha_j = 1$ holds for all $1 \leq j \leq N$. In particular, we see that the representation of v is not unique.

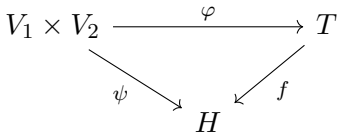
We now want to use bilinear forms to define the *algebraic tensor product space* of V_1 and V_2 as the image of such mappings. However, it is obvious that, in general, there are many bilinear forms with domain of definition given by $V_1 \times V_2$ that possess different images. As it turns out, the so-called *universal property* (see [21, Chapter 1.4]) is the right requirement to guarantee the well-posedness of this idea.

Definition 2.4.3 (Universal property)**.**
Let V_1, V_2 and T be linear spaces and $\varphi\colon V_1 \times V_2 \to T$ bilinear. Then φ possesses the *universal property* if

1. φ generates T, i.e., $\mathrm{Im}(\varphi) = T$,
2. for every bilinear form $\psi\colon V_1 \times V_2 \to H$ with $\mathrm{Im}(\psi) = H$ for some linear space H, there is a linear mapping $f\colon T \to H$, such that

   $$\psi = f \circ \varphi,$$

   holds, i.e., the following diagram commutes:



When we consider two bilinear mappings $\varphi_1\colon V_1 \times V_2 \to T_1$ and $\varphi_2\colon V_1 \times V_2 \to T_2$, where both possess the universal property, we can identify the images spaces T_1 and T_2 (see [21, Chapter 1.6]).

Theorem 2.4.4.
Let V_1, V_2, T_1 and T_2 be linear spaces and $\varphi_j\colon V_1 \times V_2 \to T_j$, for $j = 1, 2$, two bilinear forms with the universal property. Then there is a linear and bijective mapping $f\colon T_1 \to T_2$, i.e., T_1 and T_2 are equal up to isomorphism.

This allows us to define the *algebraic tensor product space* of V_1 and V_2:

Definition 2.4.5 (Algebraic tensor product space)**.**
Let V_1, V_2 and T be linear spaces and $\varphi\colon V_1 \times V_2 \to T$ a bilinear mapping with the universal property. Then T is called the *algebraic tensor product space* of V_1 and V_2. We denote this space by

$$V_1 \otimes V_2 := T.$$

Additionally, we call

$$v_1 \otimes v_2 := \varphi(v_1, v_2)$$

for $v_1 \in V_1$ and $v_2 \in V_2$, an *elementary tensor* in $V_1 \otimes V_2$. Hence, every element $v \in V_1 \otimes V_2$ can be written as a linear combination of elementary tensors,

$$v = \sum_{j=1}^{N} v_1^{(j)} \otimes v_2^{(j)},$$

for $N \in \mathbb{N}$, $v_1^{(j)} \in V_1$ and $v_2^{(j)} \in V_2$ for $1 \leq j \leq N$.

Next, we collect some basic results for this space, which are for example given in [21, Chapter 1.5 – 1.7].

Corollary 2.4.6.
Let V_1 and V_2 be two linear spaces.

1. *The algebraic tensor product space $V_1 \otimes V_2$ always exists;*
2. *Let $\mathcal{B}_j$ be a basis of V_j for $j = 1, 2$. Then $\mathcal{B}_1 \otimes \mathcal{B}_2 := \{b_1 \otimes b_2 \mid b_1 \in \mathcal{B}_1, b_2 \in \mathcal{B}_2\}$ is a basis of $V_1 \otimes V_2$;*
3. *The dimension of $V_1 \otimes V_2$ is given by $\dim(V_1 \otimes V_2) = \dim(V_1)\dim(V_2)$;*
4. *For every $u \in V_1 \otimes V_2 \setminus \{0\}$, there is a uniquely determined minimal number $r \in \mathbb{N}$ and linearly independent vectors $v_1^{(j)} \in V_1, v_2^{(j)} \in V_2$ for $1 \leq j \leq r$, such that*

$$u = \sum_{j=1}^{r} v_1^{(j)} \otimes v_2^{(j)}.$$

This number r is called the rank of u.

Example.
Consider the two Euclidean spaces $\mathbb{R}^d$ and $\mathbb{R}^n$. Then the *dyadic product*

$$\varphi_1 \colon \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^{d \times n}, \quad (u, v) \mapsto \begin{pmatrix} u_1 v_1 & \dots & u_1 v_n \\ u_2 v_1 & \dots & u_2 v_n \\ \vdots & \vdots & \vdots \\ u_d v_1 & \dots & u_d v_n \end{pmatrix} = uv^T$$

is a bilinear form that possesses the universal property, with $\mathrm{Im}(\varphi_1) = \mathbb{R}^{d \times n}$. Hence, the algebraic tensor product space of $\mathbb{R}^d \otimes \mathbb{R}^n$ is $\mathbb{R}^{d \times n}$. On the other hand, the algebraic tensor product space of $\mathbb{R}^d$ and $\mathbb{R}^n$ can also be identified by using the *Kronecker product* as

$$\varphi_2 \colon \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^{dn}, \quad (u, v) \mapsto (u_1 v_1, \dots, u_1 v_n, u_2 v_1, \dots, u_d v_n)^T.$$

Obviously the two linear spaces $\mathbb{R}^{d \times n}$ and $\mathbb{R}^{dn}$ are isomorphic. If not stated otherwise, we will always use the *dyadic product* and write $u \otimes v := \varphi_1(u, v)$ for $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^n$.

Another application is the tensor product of two function spaces.

Theorem 2.4.7.
Let $V_1 \subseteq \{v_1 \colon \mathcal{D} \to \mathbb{R}\}$ and $V_2 \subseteq \{v_2 \colon \mathcal{T} \to \mathbb{R}\}$ be two linear spaces of real-valued functions defined on $\mathcal{D}$ and $\mathcal{T}$, respectively. Then $V_1 \otimes V_2 \subseteq \{v \colon \mathcal{D} \times \mathcal{T} \to \mathbb{R}\}$, and an elementary tensor is given by

$$(v_1 \otimes v_2)(x, y) := v_1(x) v_2(y), \quad v_1 \in V_1, v_2 \in V_2, x \in \mathcal{D}, y \in \mathcal{T}.$$

Proof. This statement can be implicitly found in [34, Theorem 1.13] or [24, Chapter 1.1.3]. □

2.4.2 Algebraic tensor product of normed spaces

The previous results are all given for linear spaces with no additional structure. If V_1 and V_2 are both normed spaces, it would be desirable for the algebraic tensor product space $V_1 \otimes V_2$ to be a normed space as well. In particular, we have to define norms on the algebraic tensor product space. This can be achieved using different constructions. We start with the so-called *projective norm* ([24, Definition 4.50]).

Definition 2.4.8 (Projective norm)**.**
Let $(V_1, \|\cdot\|_1)$ and $(V_2, \|\cdot\|_2)$ be two normed spaces with algebraic tensor product space $V_1 \otimes V_2$. The *projective norm* on $V_1 \otimes V_2$ is given by

$$\|v\|_\wedge := \inf\left\{ \sum_{j=1}^{N} \|v_1^{(j)}\|_1 \|v_2^{(j)}\|_2 \;\middle|\; v = \sum_{j=1}^{N} v_1^{(j)} \otimes v_2^{(j)} \right\}, \quad v \in V_1 \otimes V_2.$$

We omit the proof that the projective norm is indeed a norm on $V_1 \otimes V_2$, but refer to [24, Lemma 4.51]. Additionally, this norm fulfills the *cross norm* property (see [24, Definition 4.37]).

Definition 2.4.9 (Cross norm)**.**
Let $(V_1, \|\cdot\|_1)$ and $(V_2, \|\cdot\|_2)$ be two normed spaces and $\|\cdot\|\colon V_1 \otimes V_2 \to \mathbb{R}$ a norm on the algebraic tensor product space $V_1 \otimes V_2$. This norm is called a *cross norm* if for all elementary tensors $v = v_1 \otimes v_2$ the property

$$\|v\| = \|v_1\|_1 \|v_2\|_2$$

holds.

The cross norm property of the projective norm is proven in [24, Lemma 4.51] as well.
A second way of defining a norm on $V_1 \otimes V_2$ leads to the so-called *injective norm* ([24, Definition 4.58]).

Definition and Remark 2.4.10 (Injective norm)**.**
Let $(V_1, \|\cdot\|_1)$ and $(V_2, \|\cdot\|_2)$ be two normed spaces with corresponding dual spaces V_1^* and V_2^*.

1. Let $\lambda_1 \in V_1^*, \lambda_2 \in V_2^*$. For $v = \sum_{j=1}^{N} v_1^{(j)} \otimes v_2^{(j)} \in V_1 \otimes V_2$, we define the linear mapping $\lambda_1 \otimes \lambda_2\colon V_1 \otimes V_2 \to \mathbb{R}$ via

   $$(\lambda_1 \otimes \lambda_2)(v) := \sum_{j=1}^{N} \lambda_1\left(v_1^{(j)}\right) \lambda_2\left(v_2^{(j)}\right).$$

   This is well-defined, i.e., independent of the representation of v.

2. Let $v \in V_1 \otimes V_2$ be arbitrary. The *injective norm* on $V_1 \otimes V_2$ is defined as

   $$\|v\|_\vee := \sup\left\{ |(\lambda_1 \otimes \lambda_2)(v)| \;\middle|\; \lambda_j \in V_j^*, \|\lambda_j\|_{V_j^*} \leq 1,\, j = 1, 2 \right\}.$$

   This indeed defines a norm on $V_1 \otimes V_2$. Furthermore, it is a cross norm (see [24, Lemma 4.59] for both statements).

Another important property for norms on an algebraic tensor product space is the following (cf. [24, Definition 4.72]):

Definition and Remark 2.4.11 (Reasonable norm)**.**
Let V_1 and V_2 be normed spaces. A norm $\|\cdot\|_V$ on $V := V_1 \otimes V_2$ is called *reasonable*, if for all linear functionals $\lambda_1 \in V_1^*$ and $\lambda_2 \in V_2^*$, the equality

$$\|\lambda_1 \otimes \lambda_2\|_{V^*} = \sup_{\|v\|_V=1} |(\lambda_1 \otimes \lambda_2)(v)| = \|\lambda_1\|_{V_1^*}\|\lambda_2\|_{V_2^*}$$

holds. It is shown in [24, Corollary 4.64, Lemma 4.59] that both the injective and projective norm are reasonable on V.

As it turns out, the injective and projective norm stand out among all other reasonable cross norms on the algebraic tensor product space.

Proposition 2.4.12.
Let U_1 and U_2 be linear spaces with algebraic tensor product space $U = U_1 \otimes U_2$.

1. *Assume $\|\cdot\|_U$ is a reasonable norm on U. Then, for all $u \in U$, it holds that*

$$\|u\|_\vee \leq \|u\|_U;$$

2. *Assume $\|\cdot\|_U$ is a cross norm on U. Then, for all $u \in U$, it holds that*

$$\|u\|_U \leq \|u\|_\wedge;$$

3. *Assume $\|\cdot\|_U$ is a reasonable cross norm on U. Then, for all $u \in U$, it holds that*

$$\|u\|_\vee \leq \|u\|_U \leq \|u\|_\wedge.$$

Proof. Obviously, the third statement is an immediate consequence of the first two. While the first one can be found in [24, Proposition 4.52], the second one is proven in [24, Lemma 4.56]. The constant $C > 0$ occurring in the proofs can be chosen as $C = 1$. □

We have seen that we can use norms on the underlying spaces V_1 and V_2 to equip the algebraic tensor product space $V_1 \otimes V_2$ with a norm. However, when considering complete normed spaces, i.e., Banach spaces, the algebraic tensor product space does not generally have to be complete. Therefore, we consider the *tensor product space*.

Definition 2.4.13 (Tensor product space)**.**
Let V_1, V_2 be two normed spaces and $V_1 \otimes V_2$ equipped with the norm $\|\cdot\|$.
Then the space $V_1 \overline{\otimes} V_2$ is defined as the completion of $V_1 \otimes V_2$ under the norm $\|\cdot\|$, i.e.,

$$V_1 \overline{\otimes} V_2 := \overline{V_1 \otimes V_2}^{\|\cdot\|},$$

equipped with the norm $\|\cdot\|$. We call $V_1 \overline{\otimes} V_2$ the *tensor product space* of V_1 and V_2 with norm $\|\cdot\|$.

Using this concept, we now seek to define two tensor product spaces, which will be used throughout this thesis.

2.4.3 Examples of tensor product spaces

In the following, we define the *mixed regularity spaces*, i.e., tensor product spaces of differentiable functions. To this end, we introduce the following shorthand notation:

Notation.
For $\alpha_1 \in \mathbb{N}_0^d$ and $\alpha_2 \in \mathbb{N}_0^n$ and $v\colon \mathcal{D} \times \mathcal{T} \to \mathbb{R}$ sufficiently regular, we set

$$D^{\alpha_1,\alpha_2} v(x,y) = D_x^{\alpha_1} D_y^{\alpha_2} v(x,y),$$

where $D_y^{\alpha_2}$ only acts on the variables in the second argument $y \in \mathcal{T}$ and $D_x^{\alpha_1}$ only on the ones in the first argument $x \in \mathcal{D}$.

Theorem 2.4.14.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ and $\mathcal{T} \subseteq \mathbb{R}^n$ and $k_1, k_2 \in \mathbb{N}_0$. The mixed regularity space $C_{\text{mix}}^{k_1,k_2}(\mathcal{D} \times \mathcal{T})$ is defined as

$$C_{\text{mix}}^{k_1,k_2}(\mathcal{D} \times \mathcal{T}) := C_b^{k_1}(\mathcal{D}) \overline{\otimes} C_b^{k_2}(\mathcal{T}),$$

where the closure with respect to the injective norm is given by

$$\|v\|_{C_{\text{mix}}^{k_1,k_2}(\mathcal{D}\times\mathcal{T})} := \|v\|_\vee = \max_{|\alpha_1| \le k_1} \max_{|\alpha_2| \le k_2} \|D^{\alpha_1,\alpha_2} v\|_{C(\mathcal{D}\times\mathcal{T})}.$$

Moreover, the mixed regularity space is separable and permits the representation

$$C_{\text{mix}}^{k_1,k_2}(\mathcal{D} \times \mathcal{T}) = \{v \in C_b(\mathcal{D} \times \mathcal{T}) \mid D^{\alpha_1,\alpha_2} v \in C_b(\mathcal{D} \times \mathcal{T}), \\ \alpha_1 \in \mathbb{N}_0^d, \alpha_2 \in \mathbb{N}_0^n, |\alpha_j| \le k_j, j = 1,2\Big\}.$$

If $k_1 = k_2 = k$, we write $C_{\text{mix}}^k(\mathcal{D} \times \mathcal{T}) := C_{\text{mix}}^{k,k}(\mathcal{D} \times \mathcal{T})$.

Proof. See [24, Remark 4.69] for the general idea. The mixed norm with $k_1 = k_2 = 1$ is then introduced in [24, Example 4.80], which can be trivially generalized to arbitrary $k_1, k_2 \in \mathbb{N}_0$. □

We want to provide a similar norm-representation for the tensor product of Hölder spaces. To this end, we have to equip these spaces with an equivalent norm.

Lemma 2.4.15.
Let $s = k + \gamma > 0$ with $k \in \mathbb{N}_0$ and $0 \le \gamma < 1$ be given. Then the norm

$$\|v\| := \begin{cases} \max_{|\alpha|\le k} \max_{|\beta|=k} \left(\|D^\alpha v\|_{C_b(\mathcal{D})}, \sup_{\substack{x,y\in\mathcal{D} \\ x\neq y}} \frac{|D^\beta v(x) - D^\beta v(y)|}{\|x-y\|^\gamma} \right), & \gamma > 0, \\ \max_{|\alpha|\le k} \|D^\alpha v\|_{C_b(\mathcal{D})}, & \gamma = 0, \end{cases}$$

is equivalent to the standard norm $\|\cdot\|_{C^s(\mathcal{D})}$ on $C^s(\mathcal{D})$ given in Definition 2.2.2.

Proof. First, we see that the above defines a norm on $C^s(\mathcal{D})$. For $\gamma = 0$, the statement obviously holds, and for $\gamma > 0$, the norm equivalence follows by

$$\begin{aligned}
\|v\| &= \max_{|\alpha|\leq k} \max_{|\beta|=k} \left(\|D^\alpha v\|_{C_b(\mathcal{D})}, \sup_{\substack{x,y\in\mathcal{D}\\ x\neq y}} \frac{|D^\beta v(x) - D^\beta v(y)|}{\|x-y\|^\gamma} \right) \\
&\leq \sum_{|\alpha|\leq k} \|D^\alpha v\|_{C_b(\mathcal{D})} + \sum_{|\beta|=k} \sup_{\substack{x,y\in\mathcal{D}\\ x\neq y}} \frac{|D^\beta v(x) - D^\beta v(y)|}{\|x-y\|^\gamma} \\
&= \|v\|_{C^{k,\gamma}(\mathcal{D})} \\
&\leq \left(\binom{k+d}{d} + \binom{d+k-1}{k} \right) \|v\|,
\end{aligned}$$

where we use that the number of multi-indices $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq k$ is given by $\binom{k+d}{d}$ and for $|\beta| = k$ by $\binom{d+k-1}{k}$. □

In the following, we denote the norm defined in the previous lemma by $\|\cdot\|_{C^s(\mathcal{D})}$. Next, it is our goal to extend the result given in Theorem 2.4.14 to Hölder spaces.

Theorem 2.4.16.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ and $\mathcal{T} \subseteq \mathbb{R}^n$ be two bounded domains, $s = k+\gamma$ and $t = \ell+\eta$, where $k, \ell \in \mathbb{N}_0$ and $0 < \gamma, \eta < 1$. Then, for all functions $v \in C^{s,t}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T}) := C^s(\mathcal{D})\overline{\otimes}C^t(\mathcal{T})$, the injective norm $\|v\|_{C^{s,t}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} := \|v\|_\vee$ is finite. Additionally, this norm is given by

$$\begin{aligned}
\|v\|_\vee = \max_{|\alpha_1|\leq k} \max_{|\alpha_2|=k} \max_{|\beta_1|\leq \ell} \max_{|\beta_2|=\ell} \Bigg\{ & \|D^{\alpha_1,\beta_1} v\|_{C(\mathcal{D}\times\mathcal{T})}, \\
& \sup_{x_1\in\mathcal{D}} \sup_{\substack{x_2,y_2\in\mathcal{T}\\ x_2\neq y_2}} \frac{|D^{\alpha_1,\beta_2} v(x_1,x_2) - D^{\alpha_1,\beta_2} v(x_1,y_2)|}{\|x_2-y_2\|_2^\eta}, \\
& \sup_{\substack{x_1,y_1\in\mathcal{D}\\ x_1\neq y_1}} \sup_{x_2\in\mathcal{T}} \frac{|D^{\alpha_2,\beta_1} v(x_1,x_2) - D^{\alpha_2,\beta_1} v(y_1,x_2)|}{\|x_1-y_1\|_2^\gamma}, \\
& \sup_{\substack{x_1,y_1\in\mathcal{D}\\ x_1\neq y_1}} \sup_{\substack{x_2,y_2\in\mathcal{T}\\ x_2\neq y_2}} \frac{|D^{\alpha_2,\beta_2} v(x_1,x_2) - D^{\alpha_2,\beta_2} v(x_1,y_2) - D^{\alpha_2,\beta_2} v(y_1,x_2) + D^{\alpha_2,\beta_2} v(y_1,y_2)|}{\|x_1-y_1\|_2^\gamma \|x_2-y_2\|_2^\eta} \Bigg\}.
\end{aligned} \tag{2.1}$$

Proof. We only have to prove that the injective norm is indeed given by the stated expression. The proof we present next closely follows the ideas from Theorem 2.4.14 and is similar to [24, Remark 4.69, Example 4.80]. For brevity, we denote the right hand-side by $\|v\|_*$ and set $V_1 := C^s(\mathcal{D})$ and $V_2 := C^t(\mathcal{T})$.
First, we consider $v \in V_1 \otimes V_2$ as an element of the algebraic tensor product space and write

$$v(x,y) = \sum_{j=1}^N v_1^{(j)}(x) v_2^{(j)}(y).$$

Our goal is to show that $\|v\|_\vee \geq \|v\|_*$ and $\|v\|_\vee \leq \|v\|_*$.
For the first inequality, we can see that for every $x_1 \neq y_1 \in \mathcal{D}$ and $\alpha_1, \beta_1 \in \mathbb{N}_0^d$ with $|\alpha_1| \leq k, |\beta_1| = k$, the functionals

$$\delta_{\alpha_1,x_1}(v_1) := D^{\alpha_1} v_1(x_1), \quad \delta_{\beta_1,x_1,y_1}(v_1) := \frac{D^{\beta_1} v_1(x_1) - D^{\beta_1} v_1(y_1)}{\|x_1 - y_1\|^\gamma}$$

are elements of V_1^*. Analogously, we define linear functionals $\kappa_{\alpha_2,x_2}, \kappa_{\beta_2,x_2,y_2} \in V_2^*$ with suitable $\alpha_2, \beta_2 \in \mathbb{N}_0^n$ and denote by $M_1 \subseteq V_1^*$ and $M_2 \subseteq V_2^*$ the set of these functionals. Obviously, the operator norms of all elements in M_1 and M_2 are less or equal than one, which allows us to bound the injective norm from below by

$$\|v\|_\vee = \sup_{\substack{\lambda_1 \in V_1^*, \lambda_2 \in V_2^* \\ \|\lambda_1\|_{V_1^*}, \|\lambda_2\|_{V_2^*} \leq 1}} |(\lambda_1 \otimes \lambda_2) v| \geq \sup_{(\lambda_1,\lambda_2) \in M_1 \times M_2} |(\lambda_1 \otimes \lambda_2) v| = \|v\|_*,$$

where the last equality follows by plugging in the definition of M.
For the other inequality, consider $\lambda_1 \in V_1^*$, $\lambda_2 \in V_2^*$ and $v = \sum_{j=1}^N v_1^{(j)} \otimes v_2^{(j)} \in V_1 \otimes V_2$. Then

$$(\lambda_1 \otimes \lambda_2)(v) = \sum_{j=1}^N \lambda_1\left(v_1^{(j)}\right) \lambda_2\left(v_2^{(j)}\right) = \lambda_2^y \lambda_1^x v(x,y),$$

where λ_1^x indicates that we apply λ_1 to the mapping $x \mapsto v(x,\cdot)$. Next, we define the function $v_2(y) := \lambda_1^x v(x,y) \in V_2$, thus we can apply the boundedness of λ_2 to obtain

$$\begin{aligned} |\lambda_2^y \lambda_1^x v(x,y)| &= |\lambda_2^y v_2(y)| \leq \|\lambda_2\|_{V_2^*} \|v_2\|_{V_2} \\ &= \|\lambda_2\|_{V_2^*} \max_{|\alpha_2| \leq \ell} \max_{|\beta_2| = \ell} \left\{ \|D^{\alpha_2} v_2\|_{C(\mathcal{T})}, \sup_{\substack{x_2, y_2 \in \mathcal{T} \\ x_2 \neq y_2}} \frac{|D^{\beta_2} v_2(x_2) - D^{\beta_2} v_2(y_2)|}{\|x_2 - y_2\|_2^\eta} \right\}. \end{aligned} \tag{2.2}$$

To estimate this further, we seek to bound for example the last expression in the maximum. For this, choose $\beta_2 \in \mathbb{N}_0^n$ with $|\beta_2| = \ell$ arbitrary, but fixed. Then, for $x_2, y_2 \in \mathcal{T}$ with $x_2 \neq y_2$, we plug in the definition of v_2 to see that

$$\begin{aligned} \left| D^{\beta_2} v_2(x_2) - D^{\beta_2} v_2(y_2) \right| &= \left| \sum_{j=1}^N \lambda_1\left(v_1^{(j)}\right) D^{\beta_2} v_2^{(j)}(x_2) - \sum_{j=1}^N \lambda_1\left(v_1^{(j)}\right) D^{\beta_2} v_2^{(j)}(y_2) \right| \\ &= \left| \lambda_1 \left(\sum_{j=1}^N v_1^{(j)} \left(D^{\beta_2} v_2^{(j)}(x_2) - D^{\beta_2} v_2^{(j)}(y_2) \right) \right) \right| \\ &\leq \|\lambda_1\|_{V_1^*} \left\| \sum_{j=1}^N v_1^{(j)} \left(D^{\beta_2} v_2^{(j)}(x_2) - D^{\beta_2} v_2^{(j)}(y_2) \right) \right\|_{V_1}. \end{aligned}$$

Therefore, the last expression in the maximum of (2.2) can be bounded by

$$\frac{|D^{\beta_2} v_2(x_2) - D^{\beta_2} v_2(y_2)|}{\|x_2 - y_2\|_2^\eta} \leq \|\lambda_1\|_{V_1^*} \frac{\left\| \sum_{j=1}^N v_1^{(j)} \left(D^{\beta_2} v_2^{(j)}(x_2) - D^{\beta_2} v_2^{(j)}(y_2) \right) \right\|_{V_1}}{\|x_2 - y_2\|_2^\eta}. \tag{2.3}$$

If we now consider the latter expression of (2.3), we can insert the definition of $\|\cdot\|_{V_1}$. Here, we focus again on the second term in the occurring maximum. For $\beta_1 \in \mathbb{N}_0^d$ with $|\beta_1| = k$ fixed, this can be rewritten as

$$\begin{aligned}
&\sup_{\substack{x_2,y_2\in\mathcal{T}\\ x_2\neq y_2}} \sup_{\substack{x_1,y_1\in\mathcal{D}\\ x_1\neq y_1}} \frac{1}{\|x_2-y_2\|_2^{\eta}\|x_1-y_1\|_2^{\gamma}} \Bigg| \sum_{j=1}^{N} D^{\beta_1} v_1^{(j)}(x_1) \left(D^{\beta_2} v_2^{(j)}(x_2) - D^{\beta_2} v_2^{(j)}(y_2)\right) \\
&\qquad\qquad - \sum_{j=1}^{N} D^{\beta_1} v_1^{(j)}(y_1) \left(D^{\beta_2} v_2^{(j)}(x_2) - D^{\beta_2} v_2^{(j)}(y_2)\right) \Bigg| \\
&= \sup_{\substack{x_2,y_2\in\mathcal{T}\\ x_2\neq y_2}} \sup_{\substack{x_1,y_1\in\mathcal{D}\\ x_1\neq y_1}} \frac{\left|D^{\beta_1,\beta_2} v(x_1,x_2) - D^{\beta_1,\beta_2} v(x_1,y_2) - D^{\beta_1,\beta_2} v(y_1,x_2) + D^{\beta_1,\beta_2} v(y_1,y_2)\right|}{\|x_2-y_2\|_2^{\eta}\|x_1-y_1\|_2^{\gamma}}.
\end{aligned}$$

Considering the maximum of all possible derivatives $\beta_1 \in \mathbb{N}_0^d$ and $\beta_2 \in \mathbb{N}_0^n$ yields the last term in the stated norm representation (2.1). The other ones follow with similar arguments in the other cases. Thus, we obtain the equality of the expressions for all elements $v \in V_1 \otimes V_2$. This stays true for all elements in $V_1 \overline{\otimes} V_2$ after completion. □

2.4.4 Tensor product of Hilbert spaces

Next, we assume that $V_1 = H_1$ and $V_2 = H_2$ are given by Hilbert spaces $(H_1, \langle\cdot,\cdot\rangle_{H_1})$ and $(H_2, \langle\cdot,\cdot\rangle_{H_2})$, respectively. In this case, we can define an inner product on $H_1 \otimes H_2$, which therefore leads to a pre-Hilbert space.

Theorem 2.4.17.
Let $(H_1, \langle\cdot,\cdot\rangle_1)$ and $(H_2, \langle\cdot,\cdot\rangle_2)$ be two Hilbert spaces. Define, for $u = \sum_{j=1}^{N} u_1^{(j)} \otimes u_2^{(j)}$ and $v = \sum_{k=1}^{M} v_1^{(k)} \otimes v_2^{(k)}$ in $H_1 \otimes H_2$, the bilinear form

$$\langle u, v\rangle := \sum_{j=1}^{N}\sum_{k=1}^{M} \langle u_1^{(j)}, v_1^{(k)}\rangle_1 \langle u_2^{(j)}, v_2^{(k)}\rangle_2.$$

Then $\langle\cdot,\cdot\rangle$ is independent of the representations of u and v and yields an inner product on $H_1 \otimes H_2$, the induced inner product *of H_1 and H_2. The induced norm $\|f\| := \sqrt{\langle f, f\rangle}$ for $f \in H_1 \otimes H_2$ is a reasonable cross norm and the tensor product space $H_1 \overline{\otimes} H_2$ with respect to the induced norm is again a Hilbert space.*

Proof. It is shown in [24, Lemma 4.147] that the stated definition is well-defined and yields an inner product on $H_1 \otimes H_2$. Then [24, Lemma 4.150] proves that the norm is a reasonable cross norm. Therefore, $H_1 \otimes H_2$ is a pre-Hilbert space. This allows us to complete this space with respect to the induced norm $\|\cdot\|$. As the completion preserves the inner product, the result is a Hilbert space. □

2.4.5 Mixed regularity Sobolev spaces

In Theorem 2.4.14 and Theorem 2.4.16, we defined the tensor product space of differentiable and Hölder continuous functions, respectively. Of course, there is an analog for Sobolev spaces using weak derivatives.

In the following, we focus on the result for Hilbert spaces, i.e., the case $p = 2$ in Definition and Remark 2.2.5.

Theorem 2.4.18 (Mixed regularity Sobolev spaces)**.**
Let $\sigma \geq 0$ and $\tau \geq 0$. The mixed regularity Sobolev spaces *are defined as*

$$H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T}) := H^{\sigma}(\mathcal{D}) \overline{\otimes} H^{\tau}(\mathcal{T}),$$

endowed with the induced inner product and norm as defined in Theorem 2.4.17. Moreover, if $\sigma = k \in \mathbb{N}_0$ and $\tau = \ell \in \mathbb{N}_0$, the mixed regularity Sobolev space is given by

$$H^{k,\ell}_{\text{mix}}(\mathcal{D} \times \mathcal{T}) = \left\{ v \in L^2(\mathcal{D} \times \mathcal{T}) \mid D^{\alpha,\beta} v(x,y) \in L^2(\mathcal{D} \times \mathcal{T}),\ |\alpha| \leq k,\ |\beta| \leq \ell \right\}.$$

Particularly, in the case $k = \ell = 0$, we obtain $L^2(\mathcal{D}) \overline{\otimes} L^2(\mathcal{T}) = L^2(\mathcal{D} \times \mathcal{T})$.

We again omit the proof, but note that the idea is very similar to the one of Theorem 2.4.14. For mixed regularity Hilbert spaces on $\mathbb{R}^d \times \mathbb{R}^n$, it is, in light of Theorem 2.2.7, not surprising that we can again characterize this space by using Fourier transformation.

Corollary 2.4.19 (Fourier characterization of $H^{\sigma,\tau}_{\text{mix}}(\mathbb{R}^d \times \mathbb{R}^n)$)**.**
Let $\sigma, \tau \geq 0$. Then the mixed regularity Sobolev space $H^{\sigma,\tau}_{\text{mix}}(\mathbb{R}^d \times \mathbb{R}^n)$ is isomorphic to

$$H^{\sigma,\tau}_{\text{mix}}(\mathbb{R}^d \times \mathbb{R}^n) \cong \left\{ v \in L^2(\mathbb{R}^d \times \mathbb{R}^n) \mid \|f\|_{\sigma,\tau} < \infty \right\},$$

where $\|\cdot\|_{\sigma,\tau}$ is defined as

$$\|v\|^2_{\sigma,\tau} = \int_{\mathbb{R}^d} \int_{\mathbb{R}^n} |\widehat{v}(\omega_1, \omega_2)|^2 \left(1 + \|\omega_1\|_2^2\right)^{\sigma} \left(1 + \|\omega_2\|_2^2\right)^{\tau} d\omega_2 \, d\omega_1.$$

Here, the Fourier transform of v is with respect to both variables $(x, y) \in \mathbb{R}^d \times \mathbb{R}^n \cong \mathbb{R}^{d+n}$ in Definition and Remark 2.2.6.

Proof. As seen in Theorem 2.2.7, we can define an equivalent norm on $H^{\sigma}(\mathbb{R}^d)$ by

$$\|v_1\|^2_{\sigma} = \frac{1}{\sqrt{2\pi}^d} \int_{\mathbb{R}^d} |\widehat{v_1}(\omega_1)|^2 \left(1 + \|\omega_1\|_2^2\right)^{\sigma} d\omega_1, \quad v_1 \in H^{\sigma}(\mathbb{R}^d),$$

and the same holds for $\|\cdot\|_{\tau}$ on $H^{\tau}(\mathbb{R}^n)$.
If we first consider an element $v = \sum_{j=1}^{N} v_1^{(j)} \otimes v_2^{(j)} \in H^{\sigma}(\mathbb{R}^d) \otimes H^{\tau}(\mathbb{R}^n)$, we can explicitly calculate the induced norm by

$$\begin{aligned}
\|v\|^2 &= \sum_{j=1}^{N} \sum_{k=1}^{N} \langle v_1^{(j)}, v_1^{(k)} \rangle_{\sigma} \langle v_2^{(j)}, v_2^{(k)} \rangle_{\tau} \\
&= \sum_{j=1}^{N} \sum_{k=1}^{N} \int_{\mathbb{R}^d} \widehat{v_1^{(j)}}(\omega_1) \overline{\widehat{v_1^{(k)}}(\omega_1)} \left(1 + \|\omega_1\|_2^2\right)^{\sigma} d\omega_1 \\
&\qquad \int_{\mathbb{R}^n} \widehat{v_2^{(j)}}(\omega_2) \overline{\widehat{v_2^{(k)}}(\omega_2)} \left(1 + \|\omega_2\|_2^2\right)^{\tau} d\omega_2 \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}^n} |\widehat{v}(\omega_1, \omega_2)|^2 \left(1 + \|\omega_1\|_2^2\right)^{\sigma} \left(1 + \|\omega_2\|_2^2\right)^{\tau} d\omega_2 \, d\omega_1 \\
&= \|v\|^2_{\sigma,\tau}.
\end{aligned}$$

Since the algebraic tensor product space $H^\sigma(\mathbb{R}^d) \otimes H^\tau(\mathbb{R}^n)$ forms a dense subset of $H^{\sigma,\tau}_{\text{mix}}(\mathbb{R}^d \times \mathbb{R}^n)$ and the norms coincide, the statement remains valid after completion and the representation of the space follows. □

2.4.6 Linear operators on tensor product spaces

In the last sections, we provided the foundation to define tensor product spaces. Next, we want to study linear operators between such spaces. Similar to the linear functionals introduced in Definition and Remark 2.4.10, we can define the tensor product of operators as follows:

Theorem 2.4.20 (Tensor product operator)**.**
Let V_1, V_2, U_1 and U_2 be linear spaces and $A_j \colon U_j \to V_j$ for $j = 1, 2$ linear. Then there is a unique linear mapping

$$A_1 \otimes A_2 \colon U_1 \otimes U_2 \to V_1 \otimes V_2$$

such that for every tensor $u = \sum_{j=1}^N u_1^{(j)} \otimes u_2^{(j)} \in U_1 \otimes U_2$ it holds that

$$A_1 \otimes A_2(u) = \sum_{j=1}^N (A_1 u_1^{(j)}) \otimes (A_2 u_2^{(j)}),$$

independently of the representation of u. We call $A_1 \otimes A_2$ the tensor product operator *of A_1 and A_2. Moreover,* $\operatorname{Im}(A_1 \otimes A_2) = \operatorname{Im}(A_1) \otimes \operatorname{Im}(A_2)$.

Proof. See [34, Definition 1.29] and the observation afterwards. The equality for the images follows directly and its proof is also omitted here. □

If we now additionally assume that the spaces U_j and V_j, for $j = 1, 2$, are normed, the question arises whether the tensor product operator $A_1 \otimes A_2$ consisting of two linear and bounded operators $A_j \colon U_j \to V_j, j = 1, 2$ is again bounded. Of course, the answer to this question depends on the choice of norms used on the tensor spaces $U_1 \otimes U_2$ and $V_1 \otimes V_2$.

Definition 2.4.21 (Compatible norms)**.**
Let U_1, U_2, V_1 and V_2 be normed spaces. Assume that $U := U_1 \otimes U_2$ is equipped with a norm $\|\cdot\|_U$ and $V := V_1 \otimes V_2$ with $\|\cdot\|_V$. Then we call the norm $\|\cdot\|_U$ *compatible* with $\|\cdot\|_V$ if for all linear and bounded operators $A_j \colon U_j \to V_j$ with $j = 1, 2$ the operator norm fulfills the bound

$$\|A_1 \otimes A_2\|_{U \to V} = \sup_{\|u\|_U \leq 1} \|(A_1 \otimes A_2)(u)\|_V \leq \|A_1\|_{U_1 \to V_1} \|A_2\|_{U_2 \to V_2}.$$

Remark 2.4.22.

1. The definition above is often stated in the case $U_1 = V_1$ and $U_2 = V_2$, where the norms are called *uniform cross norms* (cf. [24, Definition 4.84] or [34, Definition 1.11]).

2. If $\|\cdot\|_U$ is compatible with $\|\cdot\|_V$, then the tensor product operator $A_1 \otimes A_2$ is again a linear and bounded operator.

3. This allows us to *extend* the operator from the algebraic tensor product space $U_1 \otimes U_2$ to the tensor product space $U_1 \overline{\otimes} U_2$ with the well-known extension theorem [59, Satz II.1.5] for Banach spaces, while maintaining the same operator norm.

   Therefore, we can define the tensor product operator of A_1 and A_2 on the tensor product spaces $U_1 \overline{\otimes} U_2$ and $V_1 \overline{\otimes} V_2$ as

   $$A_1 \overline{\otimes} A_2 \colon U_1 \overline{\otimes} U_2 \to V_1 \overline{\otimes} V_2.$$

4. Additionally, assume that $W := W_1 \otimes W_2$ is the algebraic tensor product space of two normed spaces W_1 and W_2, where W is equipped with a norm $\|\cdot\|_W$ such that $\|\cdot\|_V$ is compatible with $\|\cdot\|_W$. Then for two linear and bounded operators $B_j \colon V_j \to W_j$ with $j = 1, 2$, we obtain

   $$(B_1 \overline{\otimes} B_2)(A_1 \overline{\otimes} A_2) = (B_1 A_1) \overline{\otimes} (B_2 A_2) \colon U_1 \overline{\otimes} U_2 \to W_1 \overline{\otimes} W_2.$$

Next, we present several examples of norms that are compatible.

Proposition 2.4.23.
Let U_1, U_2, V_1 and V_2 be linear spaces with algebraic tensor product spaces $U = U_1 \otimes U_2$ and $V = V_1 \otimes V_2$.

1. *The projective norm $\|\cdot\|_{\wedge,U}$ on U is compatible with every cross norm $\|\cdot\|_V$ on V;*
2. *Every reasonable norm $\|\cdot\|_U$ on U is compatible with the injective norm $\|\cdot\|_{\vee,V}$ on V;*
3. *If U_1, U_2, V_1 and V_2 are Hilbert spaces and U, V are endowed with the induced inner product norms $\|\cdot\|_U$ and $\|\cdot\|_V$, then $\|\cdot\|_U$ is compatible with $\|\cdot\|_V$.*

Proof. The main idea for the first two statements is given in [34, Lemma 1.12], while the third statement follows the ideas of [34, Lemma 1.35 and Lemma 1.36]. These can directly be generalized to our setting. □

2.4.7 Tensor product extension and embedding

We now want to use the tools proven previously to generalize both the Sobolev extension Theorem 2.2.9 and the embedding Theorem 2.2.10 to the mixed regularity spaces defined in Theorem 2.4.18.

Theorem 2.4.24 (Tensor product extension)**.**
Let $\mathcal{D} \subseteq \mathbb{R}^d$ and $\mathcal{T} \subseteq \mathbb{R}^n$ be two domains with Lipschitz boundary. Then there exists a linear operator $E \colon L^2(\mathcal{D} \times \mathcal{T}) \to L^2(\mathbb{R}^d \times \mathbb{R}^n)$ that, for all $\sigma, \tau \geq 0$, fulfills the following properties:

(i) $E_{|H^{\sigma,\tau}_{\text{mix}}(\mathcal{D}\times\mathcal{T})} \colon H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T}) \to H^{\sigma,\tau}_{\text{mix}}(\mathbb{R}^d \times \mathbb{R}^n)$ *is linear and bounded;*

(ii) $(E_{|H^{\sigma,\tau}_{\text{mix}}(\mathcal{D}\times\mathcal{T})}(f))_{|\mathcal{D}\times\mathcal{T}} = f$, *i.e., the extension of $f \in H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$ restricted to $\mathcal{D} \times \mathcal{T}$ is again f.*

Proof. Denote by $E_{\mathcal{D}}\colon L^2(\mathcal{D}) \to L^2(\mathbb{R}^d)$ and $E_{\mathcal{T}}\colon L^2(\mathcal{T}) \to L^2(\mathbb{R}^n)$ the extension operators for $\mathcal{D}$ and $\mathcal{T}$ given in Theorem 2.2.9, and let $\sigma, \tau \geq 0$. Since $H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$ and $H^{\sigma,\tau}_{\text{mix}}(\mathbb{R}^d \times \mathbb{R}^n)$ are Hilbert spaces, we know by Proposition 2.4.23 that the induced inner product norms are compatible, and hence, by Remark 2.4.22, we infer that

$$E_{\mathcal{D}} \overline{\otimes} E_{\mathcal{T}}\colon H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T}) \to H^{\sigma,\tau}_{\text{mix}}(\mathbb{R}^d \times \mathbb{R}^n)$$

is linear and bounded. We now claim that $E_{\mathcal{D}} \overline{\otimes} E_{\mathcal{T}}$ fulfills the properties. To prove this statement, we find, for every $f \in H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$, a sequence $(f_j)_{j\in\mathbb{N}} \subseteq H^{\sigma}(\mathcal{D}) \otimes H^{\tau}(\mathcal{T})$ such that $f_j \to f$ for $j \to \infty$ in $H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$. After choosing a subsequence, we can assume without restriction that the sequence also converges pointwise almost everywhere to f. As E is linear and bounded, we obtain

$$E(f)_{|\mathcal{D}\times\mathcal{T}}(x, y) = \lim_{j\to\infty} E(f_j)_{|\mathcal{D}\times\mathcal{T}}(x, y) = \lim_{j\to\infty} f_j(x, y) = f(x, y)$$

for almost all $(x, y) \in \mathcal{D} \times \mathcal{T}$. Thus, E fulfills the required properties and the claim follows. □

With a similar argument, we can state a tensorized version of the Sobolev embedding Theorem 2.2.10 as well.

Theorem 2.4.25 (Tensor product embedding)**.**
Let $\mathcal{D} \subseteq \mathbb{R}^d$ and $\mathcal{T} \subseteq \mathbb{R}^n$ be two domains with Lipschitz boundary.
Then, for all $s, t, \sigma, \tau \geq 0$ such that $\sigma > \frac{d}{2} + s$ and $\tau > \frac{n}{2} + t$, the embedding

$$\iota_{\mathcal{D}\times\mathcal{T}}\colon H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T}) \to C^{s,t}_{\text{mix}}(\mathcal{D} \times \mathcal{T}), \quad f \mapsto f,$$

is linear and bounded.

Proof. From Theorem 2.2.10 we know that the inclusions

$$\iota_{\mathcal{D}}\colon H^{\sigma}(\mathcal{D}) \to C^{s}(\mathcal{D}),$$
$$\iota_{\mathcal{T}}\colon H^{\tau}(\mathcal{T}) \to C^{t}(\mathcal{T})$$

are linear and bounded. The tensor product space $H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$ is endowed with the induced norm, which is a reasonable cross norm by Theorem 2.4.17, and the tensor product space $C^{s,t}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$ is endowed with the injective norm. Hence, we can apply Proposition 2.4.23 to see that $\|\cdot\|_{H^{\sigma,\tau}_{\text{mix}}(\mathcal{D}\times\mathcal{T})}$ is compatible with $\|\cdot\|_{C^{s,t}_{\text{mix}}(\mathcal{D}\times\mathcal{T})}$ and infer that

$$\iota_{\mathcal{D}} \overline{\otimes} \iota_{\mathcal{T}}\colon H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T}) \to C^{s,t}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$$

is linear and bounded. To prove that $\iota_{\mathcal{D}} \overline{\otimes} \iota_{\mathcal{T}} = \iota_{\mathcal{D}\times\mathcal{T}}$, we first observe that the statement obviously holds for elements of $H^{\sigma}(\mathcal{D}) \otimes H^{\tau}(\mathcal{T})$. For $f \in H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$ arbitrary, there exists a sequence $(f_j)_{j\in\mathbb{N}} \subseteq H^{\sigma}(\mathcal{D}) \otimes H^{\tau}(\mathcal{T})$ that converges to f in $H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$, after choosing a subsequence additionally pointwise almost everywhere. The sequence $(f_j)_{j\in\mathbb{N}} \subseteq C^{s,t}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$ also forms a Cauchy sequence and therefore converges to an element $h \in C^{s,t}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$, in particular pointwise. Thus, $h = f$ almost everywhere and the identity $\iota_{\mathcal{D}} \overline{\otimes} \iota_{\mathcal{T}}(f) = f$ holds. □

2.4.8 Bochner and tensor product spaces

We conclude this introductory chapter by connecting the Bochner spaces introduced in Section 2.3 with tensor product spaces. As it turns out, in some cases we can equip the tensor product space with a suitable norm and prove that it coincides with the corresponding Bochner space. To this end, it is shown in [26, Lemma 1.2.19] that the algebraic tensor product is a dense subset of the Bochner space.

Lemma 2.4.26.
Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and B a separable Banach space. Then the algebraic tensor product space $L^p(\Omega, \mu) \otimes B$ is a dense subset of $L^p(\Omega; B)$ for all $1 \leq p < \infty$.

Proof. Let $f \in L^p(\Omega; B)$ be arbitrary. As f is μ-measurable, we can use Corollary 2.3.3 to find a sequence $(f_n)_{n\in\mathbb{N}}$ of μ-simple functions that converges pointwise to f such that, for all $t \in \Omega$ and $n \in \mathbb{N}$, the bound $\|f_n(t)\|_B \leq \|f(t)\|_B$ holds. As the functions f_n are μ-simple for all $n \in \mathbb{N}$, we easily see that $f_n \in L^p(\Omega) \otimes B$. The convergence of $(f_n)_{n\in\mathbb{N}}$ to f for $n \to \infty$ in $L^p(\Omega; B)$ follows with the use of the dominated convergence theorem. □

We can apply this lemma to prove the following:

Proposition 2.4.27.
Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.

1. *Let B be a separable Banach space. Then $L^1(\Omega; B) = L^1(\Omega)\overline{\otimes}_\wedge B$, where the completion is with respect to the projective norm $\|\cdot\|_\wedge$;*
2. *Let H be a Hilbert space. Then $L^2(\Omega; H) = L^2(\Omega)\overline{\otimes} H$, where the completion is with respect to the induced Hilbert space norm of the tensor product.*

Proof. As shown, for example, in [26, Proposition 1.4.4], the two norms $\|\cdot\|_{L^1(\Omega;B)}$ and $\|\cdot\|_\wedge$ coincide on the algebraic tensor product space $L^1(\Omega)\otimes B$. As this is a dense subspace of $L^2(\Omega; B)$ and $L^2(\Omega)\overline{\otimes}B$, the spaces are the equal.
If H is a Hilbert space and $f = \sum_{j=1}^n f_1^{(j)} \otimes f_2^{(j)} \in L^2(\Omega) \otimes H$, the induced inner product norm can be written as

$$\begin{aligned}
\|f\|^2_{L^2(\Omega)\otimes H} &= \sum_{j=1}^n \sum_{k=1}^n \langle f_1^{(j)}, f_1^{(k)}\rangle_{L^2(\mathcal{D})} \langle f_2^{(j)}, f_2^{(k)}\rangle_H \\
&= \int_\Omega \sum_{j=1}^n \sum_{k=1}^n f_1^{(j)}(t) f_1^{(k)}(t) \langle f_2^{(j)}, f_2^{(k)}\rangle_H \, d\mu(t) \\
&= \int_\Omega \|f(t)\|_H^2 \, d\mu(t),
\end{aligned}$$

thus the two norms again coincide on $L^2(\Omega) \otimes H$. As before, this subspace is dense in both spaces $L^2(\Omega; H)$ and $L^2(\Omega)\overline{\otimes}H$, hence they are equal. □

Remark.
One can generalize the preceding proposition to $L^p(\Omega; B)$ for $1 \leq p < \infty$. For this, we have to equip $L^p(\Omega)\otimes B$ with the norm α_p defined in [34, Definition 1.45]. Afterwards, one can apply the results given in [34, Lemma 1.47, Lemma 1.49] to see that the completion with respect to this norm fulfills $L^p(\Omega, \mu)\overline{\otimes}_{\alpha_p} B = L^p(\Omega; B)$.

3 Kernel-based approximation

A central aspect of this thesis are *reproducing kernel Hilbert spaces (RKHS)*. An introductory overview article can be found in [6], but we will use [56] and [49] as our main references. Throughout this chapter, we assume that $\mathcal{D} \subseteq \mathbb{R}^d$ is a pathwise connected domain.

3.1 Reproducing kernel Hilbert space

Definition 3.1.1 (Reproducing kernel Hilbert space)**.**
Let $H \subseteq \{ f \colon \mathcal{D} \to \mathbb{R}\}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. We call a function $K \colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ *reproducing kernel* of H if

- for all $x \in \mathcal{D}$, the function $K(\cdot, x)$ is an element of H,
- for all $x \in \mathcal{D}$ and $f \in H$, the *reproducing property* $\langle f, K(\cdot, x) \rangle = f(x)$ holds.

In this case, we call H *reproducing kernel Hilbert space* or *native space of* K and denote it by $H_K(\mathcal{D}) := H$. For the inner product, we write $\langle \cdot, \cdot \rangle_{H_K(\mathcal{D})}$ or simply $\langle \cdot, \cdot \rangle_K$ (and analogously $\| \cdot \|_{H_K(\mathcal{D})}$ or $\| \cdot \|_K$) when the underlying domain is clear from the context.

As the last notation already implies, one can show that a reproducing kernel is unique if it exists (cf. [56, Definition 10.1]). Furthermore, additional properties can be derived ([56, Theorem 10.3]).

Theorem 3.1.2.
Let $H = H_K(\mathcal{D})$ be a reproducing kernel Hilbert space with kernel K. Then the following statements hold:

1. *K is symmetric, i.e., for all $x, y \in \mathcal{D}$, the kernel fulfills $K(x, y) = K(y, x)$;*
2. *Every sequence $(f_n)_{n \in \mathbb{N}} \subseteq H$ which converges to f in H also converges pointwise to f on $\mathcal{D}$.*

To prove existence of a reproducing kernel given a Hilbert space, we can use the following equivalent description ([56, Theorem 10.2]).

Theorem 3.1.3.
Let H be a Hilbert space of functions $f \colon \mathcal{D} \to \mathbb{R}$. Then the following statements are equivalent:

1. *H possesses a reproducing kernel K;*
2. *For every $x \in \mathcal{D}$, the point evaluation functional $\delta_x \colon H \to \mathbb{R}, \delta_x(f) := f(x)$ is linear and bounded, i.e., $\delta_x \in H^*$.*

While this condition allows us to determine whether a given Hilbert space possesses a reproducing kernel, we can also give a condition for a function $K\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ that guarantees the existence of a corresponding reproducing kernel Hilbert space H such that K is the reproducing kernel of this space (cf [56, Definition 6.1, Theorem 6.3]).

Definition 3.1.4 (Positive semi-definiteness)**.**
Let $K\colon \mathcal{D}\times\mathcal{D} \to \mathbb{R}$. We call K *positive semi-definite* if K is symmetric and, for all $N \in \mathbb{N}$, $x_1, \dots, x_N \in \mathcal{D}$ and $\alpha_1, \dots, \alpha_N \in \mathbb{R}$, the quadratic form satisfies

$$\sum_{k=1}^{N}\sum_{\ell=1}^{N} \alpha_k \alpha_\ell K(x_k, x_\ell) \geq 0.$$

If this expression is strictly positive for all mutually distinct points $x_1, \dots, x_N \in \mathcal{D}$ and all coefficients $\alpha_1, \dots, \alpha_N \in \mathbb{R}$, where $\alpha_j \neq 0$ for at least one $j \in \{1, \dots, N\}$, K is called *positive definite*.

Remark 3.1.5.
There is also a close connection to the positive (semi-)definiteness of matrices: For $N \in \mathbb{N}$, let $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ be a finite point set and $\alpha = (\alpha_1, \dots, \alpha_N)^T \in \mathbb{R}^N$. We define the *kernel matrix* of $K\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ in X as $K(X, X) := (K(x_k, x_\ell))_{1\leq k,\ell\leq N} \in \mathbb{R}^{N\times N}$. Then we can rewrite the preceding condition as

$$0 \leq \sum_{k=1}^{N}\sum_{l=1}^{N} \alpha_k \alpha_l K(x_k, x_l) = \alpha^T K(X, X)\alpha,$$

i.e., the positive (semi-)definiteness of K ensures that for all finite point sets $X \subseteq \mathcal{D}$ the corresponding kernel matrix is positive (semi-)definite.

With this definition, we can state the so-called *Moore-Aronszajin* theorem ([49, Theorem 4.21]):

Theorem 3.1.6 (Moore-Aronszajin)**.**
Let $K\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$. Then there is a Hilbert space $H = H_K(\mathcal{D})$ of functions $f\colon \mathcal{D} \to \mathbb{R}$ such that K is the reproducing kernel of H if and only if K is positive semi-definite. In this case,

$$H_{pre} := \operatorname{span}\{K(\cdot, x) \mid x \in \mathcal{D}\}$$

is a dense subset of $H_K(\mathcal{D})$.

This theorem allows us to choose a symmetric and positive definite kernel $K\colon \mathcal{D}\times\mathcal{D} \to \mathbb{R}$ and guarantees the existence of a Hilbert space $H_K(\mathcal{D})$ with an inner product $\langle\cdot,\cdot\rangle_K$. Unfortunately, this Hilbert space $H_K(\mathcal{D})$ is a rather abstract space, which is why seek a way to characterize this space in a more explicit way. While there are various methods given in [56, Chapter 10], we focus on a special class of reproducing kernels on $\mathbb{R}^d$, namely the *translation invariant* functions. In this setting, we consider kernel functions $K\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ which are given using an even and real-valued function $\Phi\colon \mathbb{R}^d \to \mathbb{R}$ and fulfill $K(x, y) = \Phi(x - y)$ for $x, y \in \mathbb{R}^d$. With a slight abuse of notation, we call Φ a positive definite function whenever Φ is even and the kernel $K(x, y) = \Phi(x - y)$ for $x, y \in \mathbb{R}^d$ is positive definite. The reproducing kernel Hilbert space of Φ will be denoted by $H_\Phi(\mathbb{R}^d) := H_K(\mathbb{R}^d)$. For kernels of this structure, we can use the Fourier transform to characterize their native spaces.

Theorem 3.1.7.
Let $\Phi \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ be a positive definite function. Then the Fourier transform of Φ is real-valued, $\widehat{\Phi} \geq 0$ and $\widehat{\Phi}$ is not vanishing on all of $\mathbb{R}^d$. Moreover, the native space of Φ is given by

$$H_\Phi(\mathbb{R}^d) = \left\{ f \in L^2(\mathbb{R}^d) \cap C(\mathbb{R}^d) \mid \frac{\widehat{f}}{\sqrt{\widehat{\Phi}}} \in L^2(\mathbb{R}^d) \right\},$$

and the inner product can be represented as

$$\langle f, g \rangle_\Phi = (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \frac{\widehat{f}(\omega)\overline{\widehat{g}}(\omega)}{\widehat{\Phi}(\omega)} \, d\omega, \quad f, g \in H_\Phi(\mathbb{R}^d).$$

Proof. It is shown in [56, Theorem 6.11] that the Fourier transform of Φ is positive and non-vanishing on all of $\mathbb{R}^d$, while the representation of the native space and the inner product is proven in [56, Theorem 10.12]. □

In particular, this theorem guarantees that the functions in the reproducing kernel Hilbert space $H_\Phi(\mathbb{R}^d)$ are again continuous functions. If we now compare this result with Theorem 2.2.7, we get the following corollary:

Corollary 3.1.8.
Let $\Phi \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ be a positive definite function and $\sigma > \frac{d}{2}$. Assume further that $\widehat{\Phi}$ is equivalent to $(1 + \|\cdot\|_2^2)^{-\sigma}$, i.e., that there are two constants $c_1, c_2 > 0$ such that, for all $\omega \in \mathbb{R}^d$,

$$c_1 \left(1 + \|\omega\|_2^2\right)^{-\sigma} \leq \widehat{\Phi}(\omega) \leq c_2 \left(1 + \|\omega\|_2^2\right)^{-\sigma} \tag{3.1}$$

holds. Then the native space $H_\Phi(\mathbb{R}^d)$ coincides algebraically with $H^\sigma(\mathbb{R}^d)$, and the norms are equivalent with constants

$$\sqrt{c_1}(2\pi)^{\frac{d}{4}} \|f\|_{H_\Phi(\mathbb{R}^d)} \leq \|f\|_{H^\sigma(\mathbb{R}^d)} \leq \sqrt{c_2}(2\pi)^{\frac{d}{4}} \|f\|_{H_\Phi(\mathbb{R}^d)}$$

for all $f \in H^\sigma(\mathbb{R}^d)$. If $\widehat{\Phi}$ is equivalent to $(1 + \|\cdot\|_2^2)^{-\sigma}$, we also write $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$ to indicate that the spaces are algebraically equal and their norms are equivalent.

Proof. By (3.1), it is obvious that $\frac{\widehat{f}}{\sqrt{\widehat{\Phi}}} \in L^2(\mathbb{R}^d)$ if and only if $f \in H^\sigma(\mathbb{R}^d)$. The norm equivalence also follows immediately. □

With the previous corollary, it is now possible to give examples for positive definite functions where the induced native space $H_\Phi(\mathbb{R}^d)$ is equivalent to a Sobolev space:

Example.

1. The first class we consider are the *Matérn* kernels (see [42, Chapter 4.2]). For $\nu > 0$, these are defined as

$$\Phi_\nu(x) := \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\|x\|_2\right)^\nu \mathcal{K}_\nu \left(\sqrt{2\nu}\|x\|_2\right), \quad x \in \mathbb{R}^d,$$

where $\mathcal{K}_\nu$ denotes the Bessel function of the second kind. The Fourier transform of Φ_ν is given by

$$\widehat{\Phi_\nu}(\omega) = \frac{2^d \pi^{d/2} \, \Gamma\left(\nu + \frac{d}{2}\right) (2\nu)^\nu}{\Gamma(\nu)} \left(2\nu + 4\pi^2 \|\omega\|_2^2\right)^{-\left(\nu + \frac{d}{2}\right)}, \quad \omega \in \mathbb{R}^d,$$

i.e., $H_{\Phi_\nu}(\mathbb{R}^d) \cong H^{\nu + \frac{d}{2}}(\mathbb{R}^d)$ for all $\nu > 0$.

2. Another important class of positive definite functions on $\mathbb{R}^d$ are the *Wendland kernels* ([58, Definition 8]). They are built as follows: For $\ell \in \mathbb{N}$ and $r \in \mathbb{R}$ we define $\phi_\ell(r) := (1-r)_+^\ell$. We then set

$$\Phi_{d,k}(x) := \frac{1}{2^{k-1}(k-1)!} \int_{\|x\|_2}^1 \phi_{\lfloor d/2 \rfloor + k + 1}(t) t (t^2 - \|x\|_2^2)^{k-1} \, dt \; 1_{[0,1]}(\|x\|_2)$$

for all $k \in \mathbb{N}_0$ and $d \in \mathbb{N}$. From this definition, one can see that the functions $\Phi_{d,k}$ are radial, i.e., their value only depends on $\|x\|_2$. Moreover, for odd dimension $d \geq 3$, these functions coincide with the ones in $d-1$, i.e., $\Phi_{d,k} = \Phi_{d-1,k}$, with a slight abuse of notation. Additionally, for all $d \in \mathbb{N}$ and $k \in \mathbb{N}_0$, the Wendland kernels coincide with a univariate polynomial $p_{d,k}$ of degree $\lfloor d/2 \rfloor + 3k + 1$ evaluated in $\|x\|_2$ when $\|x\|_2 \leq 1$, i.e., $\Phi_{d,k}$ can be written as

$$\Phi_{d,k}(x) = \begin{cases} p_{d,k}(\|x\|_2), & 0 \leq \|x\|_2 \leq 1, \\ 0, & \|x\|_2 > 1. \end{cases}$$

The Fourier transform of $\Phi_{d,k}$ satisfies (see [58, Theorem 6] and [56, Theorem 10.35])

$$c_1 \left(1 + \|\omega\|_2^2\right)^{\frac{d}{2}+k+\frac{1}{2}} \leq \widehat{\Phi_{d,k}}(\omega) \leq c_2 \left(1 + \|\omega\|_2^2\right)^{\frac{d}{2}+k+\frac{1}{2}}, \quad \omega \in \mathbb{R}^d,$$

for all $k, d \in \mathbb{N}$. Moreover, $k = 0$ is permitted, if $d \geq 3$. In these cases, we obtain $H_{\Phi_{d,k}}(\mathbb{R}^d) \cong H^{k + \frac{d+1}{2}}(\mathbb{R}^d)$. In Table 3.1, we list explicit representations of some Wendland kernels up to constants, which is indicated by $\doteq$. As an abbreviation, we write $r := \|x\|_2$.

Dimension	Function	RKHS
$d = 1$	$\Phi_{1,0}(x) = (1-r)_+$	–
	$\Phi_{1,1}(x) \doteq (1-r)_+^3(3r+1)$	$H^2(\mathbb{R})$
	$\Phi_{1,2}(x) \doteq (1-r)_+^5(8r^2+5r+1)$	$H^3(\mathbb{R})$
$d \leq 3$	$\Phi_{3,0}(x) = (1-r)_+^2$	$H^2(\mathbb{R}^3)$
	$\Phi_{3,1}(x) \doteq (1-r)_+^4(4r+1)$	$H^3(\mathbb{R}^3)$, $H^{5/2}(\mathbb{R}^2)$
	$\Phi_{3,2}(x) \doteq (1-r)_+^6(35r^2+18r+3)$	$H^4(\mathbb{R}^3)$, $H^{7/2}(\mathbb{R}^2)$
	$\Phi_{3,3}(x) \doteq (1-r)_+^8(32r^3+25r^2+8r+1)$	$H^5(\mathbb{R}^3)$, $H^{9/2}(\mathbb{R}^2)$
$d \leq 5$	$\Phi_{5,0}(x) = (1-r)_+^3$	$H^3(\mathbb{R}^5)$, $H^{5/2}(\mathbb{R}^4)$
	$\Phi_{5,1}(x) \doteq (1-r)_+^5(5r+1)$	$H^4(\mathbb{R}^5)$, $H^{7/2}(\mathbb{R}^4)$
	$\Phi_{5,2}(x) \doteq (1-r)_+^7(16r^2+7r+1)$	$H^5(\mathbb{R}^5)$, $H^{9/2}(\mathbb{R}^4)$

Table 3.1: Wendland functions up to dimension 5.

In the previous statements, we only considered reproducing kernels and their Hilbert spaces on $\mathbb{R}^d$. In order to conclude results for arbitrary domains, we need to *extend* and *restrict* functions to different domains.

Theorem 3.1.9.
Let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ be a positive definite function and $\mathcal{D}_1 \subseteq \mathcal{D}_2 \subseteq \mathbb{R}^d$.
Then Φ is also positive definite on $\mathcal{D}_1$ and $\mathcal{D}_2$, and the following statements hold:

1. *Every $f \in H_\Phi(\mathcal{D}_1)$ has a natural extension $E_\Phi f$ to $\mathcal{D}_2$ in $H_\Phi(\mathcal{D}_2)$ with norm $\|f\|_{H_\Phi(\mathcal{D}_1)} = \|E_\Phi f\|_{H_\Phi(\mathcal{D}_2)}$,*
2. *For every $f \in H_\Phi(\mathcal{D}_2)$, the restriction $f_{|\mathcal{D}_1}$ of f to $\mathcal{D}_1$ is an element of $H_\Phi(\mathcal{D}_1)$ and $\|f_{|\mathcal{D}_1}\|_{H_\Phi(\mathcal{D}_1)} = \|f\|_{H_\Phi(\mathcal{D}_1)} \leq \|f\|_{H_\Phi(\mathcal{D}_2)}$ holds. In short: $H_\Phi(\mathcal{D}_1) = H_\Phi(\mathcal{D}_2)_{|\mathcal{D}_1}$.*

Proof. The first statement is given in [56, Theorem 10.46], the second one in [56, Theorem 10.47]. □

Remark.
Here, the restriction of a positive definite kernel $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ to $\mathcal{D} \subseteq \mathbb{R}^d$ is understood in the sense that the kernel function $K(x, y) := \Phi(x - y)$ for $x, y \in \mathbb{R}^d$ is restricted to $\mathcal{D} \times \mathcal{D}$, which then results in the restricted kernel $\Phi \colon \mathcal{D} - \mathcal{D} \to \mathbb{R}$, where

$$\mathcal{D} - \mathcal{D} := \{x - y \in \mathbb{R}^d \mid x, y \in \mathcal{D}\}.$$

We now can apply these results to kernels Φ with $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$ to get equivalent native spaces on $\mathcal{D} \subseteq \mathbb{R}^d$.

Corollary 3.1.10.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ a positive definite function with $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$. Then $H_\Phi(\mathcal{D}) \cong H^\sigma(\mathcal{D})$ holds as well.

Proof. For $\sigma = k \in \mathbb{N}$, a proof can be found in [56, Corollary 10.48]. However, the arguments there can easily be extended to arbitrary smoothness parameter $\sigma > \frac{d}{2}$ using the general Sobolev extension theorem as stated in Theorem 2.2.9. □

Coming back to the tensor product space of Hilbert spaces, the question arises whether the tensor product of two reproducing kernel Hilbert spaces again possesses a reproducing kernel. The next theorem answers this question.

Theorem 3.1.11.
Let $K_1 \colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ be the reproducing kernel of $H_{K_1}(\mathcal{D})$ and $K_2 \colon \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ the reproducing kernel of $H_{K_2}(\mathcal{T})$. Then $H_{K_1}(\mathcal{D}) \overline{\otimes} H_{K_2}(\mathcal{T})$ is again a reproducing kernel Hilbert space with kernel

$$K \colon (\mathcal{D} \times \mathcal{T}) \times (\mathcal{D} \times \mathcal{T}) \to \mathbb{R}, \quad K((x_1, y_1), (x_2, y_2)) = K_1(x_1, x_2) K_2(y_1, y_2).$$

Proof. The proof can for example be found in [37, Satz V 6]. □

Next, we consider the special case of translation invariant positive definite functions:

Corollary 3.1.12.
Let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ and $\Psi \in L^1(\mathbb{R}^n) \cap C(\mathbb{R}^n)$ be positive definite functions with $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$ for $\sigma > \frac{d}{2}$ and $H_\Psi(\mathbb{R}^n) \cong H^\tau(\mathbb{R}^n)$ for $\tau > \frac{n}{2}$, respectively. We set $K_1(x_1, x_2) := \Phi(x_1 - x_2)$ and $K_2(y_1, y_2) := \Psi(y_1 - y_2)$.

1. *The kernel*

$$K((x_1, y_1), (x_2, y_2)) := K_1(x_1, x_2)K_2(y_1, y_2), \quad x_1, x_2 \in \mathbb{R}^d, \quad y_1, y_2 \in \mathbb{R}^n,$$

   is reproducing kernel of $H_{K_1}(\mathbb{R}^d)\overline{\otimes}H_{K_2}(\mathbb{R}^n) = H_\Phi(\mathbb{R}^d)\overline{\otimes}H_\Psi(\mathbb{R}^n)$;

2. *K can be interpreted as a translation invariant function on $\mathbb{R}^{d+n}$; in particular,*

$$K((x_1, y_1), (x_2, y_2)) = \Phi(x_1 - x_2)\Psi(y_1 - y_2) = \Phi \otimes \Psi \left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} - \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right),$$

   where $\Phi \otimes \Psi \colon \mathbb{R}^{d+n} \to \mathbb{R}$;

3. *The Fourier transform of $\Phi \otimes \Psi$ in $\mathbb{R}^{d\times n}$ is given as the product of the Fourier transforms of Φ in $\mathbb{R}^d$ and Ψ in $\mathbb{R}^n$;*

4. *Consequently, the inner product of $H_K(\mathbb{R}^d \times \mathbb{R}^n)$ is given by*

$$\langle f, g\rangle_K := (2\pi)^{-\frac{d+n}{2}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^n} \frac{\widehat{f}(\omega_1, \omega_2)\overline{\widehat{g}(\omega_1, \omega_2)}}{\widehat{\Phi}(\omega_1)\widehat{\Psi}(\omega_2)} \, d\omega_2 \, d\omega_1,$$

   and thus $H_K(\mathbb{R}^d \times \mathbb{R}^n) \cong H^{\sigma,\tau}_{\text{mix}}(\mathbb{R}^d \times \mathbb{R}^n)$.

Proof. The first statement follows directly from Theorem 3.1.11, as $H_{K_1}(\mathbb{R}^d) = H_\Phi(\mathbb{R}^d)$ and $H_{K_2}(\mathbb{R}^n) = H_\Psi(\mathbb{R}^n)$, respectively. As $\mathbb{R}^d\overline{\otimes}\mathbb{R}^n \cong \mathbb{R}^{d+n}$, we can interpret $\Phi \otimes \Psi$ as a function defined on $\mathbb{R}^{d+n}$. It holds that

$$K_1(x_1, x_2)K_2(y_1, y_2) = \Phi(x_1 - x_2)\Psi(y_1 - y_2) = \Phi \otimes \Psi \left(\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} - \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right),$$

which also proves the second statement. The Fourier transform on $\Phi \otimes \Psi$ can explicitly be calculated by

$$\begin{aligned} \widehat{\Phi \otimes \Psi}(\omega_1, \omega_2) &= (2\pi)^{-\frac{d+n}{2}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^n} e^{-ix^T\omega_1 - iy^T\omega_2} \Phi(x)\Psi(y) \, dx \, dy \\ &= (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} e^{-ix^T\omega_1} \Phi(x) \, dx \; (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} e^{-iy^T\omega_2} \Psi(y) \, dy \\ &= \widehat{\Phi}(\omega_1)\widehat{\Psi}(\omega_2), \end{aligned} \tag{3.2}$$

for $\omega_1 \in \mathbb{R}^d$ and $\omega_2 \in \mathbb{R}^n$. As K is given as a translation invariant function, we can apply Theorem 3.1.7 to obtain the norm representation in combination with (3.2). Using the equivalence for $\widehat{\Phi}$ and $\widehat{\Psi}$ thus concludes the statement by Corollary 2.4.19. □

3.1.1 Mercer's theorem

With Mercer's theorem, we now present a well known result that allows us to decompose a positive semi-definite function $K \in L^2(\mathcal{D} \times \mathcal{D})$. For this, consider the *induced integral operator of* K, which is given by

$$\mathcal{K}\colon L^2(\mathcal{D}) \to L^2(\mathcal{D}), \quad u \mapsto \int_{\mathcal{D}} K(\cdot, y)u(y)\,dy$$

and has the following properties (cf. [50]):

Remark 3.1.13.
Let $K \in L^2(\mathcal{D}\times\mathcal{D})$ be a symmetric and positive semi-definite function. Then the following holds:

1. $\mathcal{K}$ is a *positive* operator, i.e., for all $f \in L^2(\mathcal{D})$, we have $\langle \mathcal{K}(f), f\rangle_{L^2(\mathcal{D})} \geq 0$;
2. $\mathcal{K}$ is *self-adjoint*, i.e., for all $f, g \in L^2(\mathcal{D})$ it holds that $\langle \mathcal{K}(f), g\rangle_{L^2(\mathcal{D})} = \langle f, \mathcal{K}(g)\rangle_{L^2(\mathcal{D})}$;
3. $\mathcal{K}$ is a *compact* operator, i.e., for every bounded sequence $(f_n)_{n\in\mathbb{N}} \subseteq L^2(\mathcal{D})$ there is a subsequence $(f_{n_j})_{j\in\mathbb{N}}$ of $(f_n)_{n\in\mathbb{N}}$ such that $\left(\mathcal{K}(f_{n_j})\right)_{j\in\mathbb{N}}$ converges in $L^2(\mathcal{D})$.

The integral operator $\mathcal{K}$ is also called *Hilbert-Schmidt operator* and K is called *Hilbert-Schmidt kernel* of $\mathcal{K}$.

Therefore, the spectral theory for compact and self-adjoint operators ([59, Theorem VI.3.2]) implies the existence of an orthonormal basis $(e_j)_{j\in\mathbb{N}} \subseteq L^2(\mathcal{D})$ consisting of eigenfunctions with the corresponding eigenvalues $(\lambda_j)_{j\in\mathbb{N}} \subseteq [0, \infty[$, i.e.,

$$\mathcal{K}(e_j) = \int_{\mathcal{D}} K(\cdot, y)e_j(y)\,dy = \lambda_j e_j, \quad j \in \mathbb{N}.$$

We can now expand $u \in L^2(\mathcal{D})$ in its Fourier series using this orthonormal basis $(e_j)_{j\in\mathbb{N}}$. If we assume for the time being that we can interchange the integral with the limit of the Fourier series, we obtain

$$\begin{aligned}
\mathcal{K}(u) = \int_{\mathcal{D}} K(\cdot, y)u(y)\,dy &= \sum_{j=1}^{\infty} \langle e_j, u\rangle_{L^2(\mathcal{D})} \mathcal{K}(e_j) \\
&= \sum_{j=1}^{\infty} \langle e_j, u\rangle_{L^2(\mathcal{D})} \lambda_j e_j \\
&= \int_{\mathcal{D}} \left(\sum_{j=1}^{\infty} \lambda_j e_j(\cdot) e_j(y) \right) u(y)\,dy, \quad u \in L^2(\mathcal{D}),
\end{aligned}$$

hence one would assume that K can be rewritten as

$$K(x, y) = \sum_{j=1}^{\infty} \lambda_j e_j(x) e_j(y), \quad x, y \in \mathcal{D},$$

where the series converges in a certain sense on $\mathcal{D} \times \mathcal{D}$. Now Mercer's theorem states that under mild conditions on the kernel K we obtain pointwise and absolute convergence of this series. Here, we consider a special case of [50, Corollary 3.5], which reads as follows:

Theorem 3.1.14 (Mercer's theorem)**.**
Let $K\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ be a reproducing kernel such that $H_K(\mathcal{D})$ is compactly embedded in $L^2(\mathcal{D})$. Let $I \subseteq \mathbb{N}$ be a finite or countable index set and denote the positive eigenvalues and corresponding eigenfunctions of the induced integral operator $\mathcal{K}$ of the kernel K by $(\lambda_j, e_j)_{j\in I} \subseteq]0,\infty[\times L^2(\mathcal{D})$.
Assume further that one of the following two conditions is satisfied:

- *K is continuous on $\mathcal{D} \times \mathcal{D}$;*
- *For all $x \in X$, the function $K(\cdot, x)\colon \mathcal{D} \to \mathbb{R}$ is continuous and bounded.*

Then, for every fixed $x \in \mathcal{D}$, the series

$$K(x,y) = \sum_{j\in I} \lambda_j e_j(x) e_j(y)$$

converges absolutely for all $y \in \mathcal{D}$. The convergence is uniform on every subset $A \subseteq \mathcal{D}$ with $\sup_{y\in A} K(y,y) < \infty$. If K is continuous, the convergence is uniform on every set $A \times A \subseteq \mathcal{D} \times \mathcal{D}$, where A is compact.
Moreover, the sequence $(\sqrt{\lambda_j} e_j)_{j\in I}$ forms an orthonormal basis of $H_K(\mathcal{D})$.

Note that the last observation, namely that $(\sqrt{\lambda_j} e_j)_{j\in I}$ is an orthonormal basis, is not mentioned in the formulation but in the proof of [50, Corollary 3.6].
The only crucial assumption we impose is the compact embedding of $H_K(\mathcal{D})$ in $L^2(\mathcal{D})$. The following lemma provides a condition for the kernel K which ensures the compactness of this embedding (see [50, Lemma 2.3]):

Lemma 3.1.15.
Let $K \in L^2(\mathcal{D} \times \mathcal{D})$ be a reproducing kernel with Hilbert space $H_K(\mathcal{D})$. Assume further that

$$\int_{\mathcal{D}} K(x,x)\, dx < \infty.$$

Then the inclusion $\iota\colon H_K(\mathcal{D}) \to L^2(\mathcal{D}), f \mapsto f$ is compact.

3.2 Function approximation using kernels

In this section, we discuss the use of reproducing kernels in the context of *function approximation* or *interpolation*. The setting will be as follows:
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain and $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ a finite point set consisting of $N \in \mathbb{N}$ pairwise distinct points. We call X the *sampling points*. Our goal is to reconstruct an underlying but unknown function $f\colon \mathcal{D} \to \mathbb{R}$ by only accessing finitely many point evaluations of f at the points in X. Additionally, we may assume that these evaluations themselves are corrupted by some additive noise $\varepsilon_j \in \mathbb{R}$; hence, we only have access to

$$z_j = f(x_j) + \varepsilon_j, \quad 1 \le j \le N.$$

We call $Z = (z_1, \dots, z_N)^T \in \mathbb{R}^N$ the *data values* of f given X. Note that we implicitly assume the sampling points X to be ordered, which induces an ordering on the data values Z. The typical approach to obtain an approximation $\mathcal{I}_X(f)$ to f is by considering

a specific finite dimensional *approximation space*, choose a basis of this space and compute an approximation in this space to the given target function while only using its data values. To make this precise, we choose the subspace spanned by $K(\cdot, x_j)$ for $1 \leq j \leq N$, where $K\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ is a symmetric and positive definite kernel, which by Theorem 3.1.6, reproduces some Hilbert space $H_K(\mathcal{D})$. In other words, we choose our basis functions to be the kernel function K centered at the points in X. Thus, the approximation space is given by

$$V_N := \operatorname{span}\{K(\cdot, x_j) \mid 1 \leq j \leq N\}.$$

In order to fit these kernel functions in a computable way, we distinguish in the following between the *kernel interpolation* $\mathcal{I}_{X,K}$ and the *penalized least squares approximation*. The first method additionally fulfills the interpolation property $\mathcal{I}_{X,K}(f)(x_k) = z_k$ for $1 \leq k \leq N$ and is mostly used when exact data $z_j = f(x_j)$ (so $\varepsilon_j = 0$) is assumed, while the latter generally does not fulfill the interpolation property but may be more reasonable for handling inexact data. In order to formalize these methods, we introduce some notation:

Notation.
Let $\mathcal{D} \subseteq \mathbb{R}^d, \mathcal{T} \subseteq \mathbb{R}^n$, $N, M \in \mathbb{N}$, and $X = \{x_1, \ldots, x_N\} \subseteq \mathcal{D}$ and $Y = \{y_1, \ldots, y_M\} \subseteq \mathcal{T}$ be two point sets.

1. For $f\colon \mathcal{D} \to \mathbb{R}$, we define $f(X) := (f(x_1), \ldots, f(x_N))^T \in \mathbb{R}^N$;

2. Let $g\colon \mathcal{D} \times \mathcal{T} \to \mathbb{R}$. Then $g(X, Y) \in \mathbb{R}^{N \times M}$ denotes the matrix

$$g(X, Y) := \begin{pmatrix} g(x_1, y_1) & \ldots & g(x_1, y_M) \\ \vdots & \vdots & \vdots \\ g(x_N, y_1) & \ldots & g(x_N, y_M) \end{pmatrix} \in \mathbb{R}^{N \times M}.$$

   If $\mathcal{D} = \mathcal{T}$ and g is symmetric, this implies that $g(Y, X) = g(X, Y)^T$. Note that for a reproducing kernel K, the matrix $K(X, X)$ is the kernel matrix defined in Remark 3.1.5;

3. Let $g\colon \mathcal{D} \times \mathcal{T} \to \mathbb{R}$ be as above and $x \in \mathcal{D}$. Then

$$g(x, Y) := g(\{x\}, Y) = (g(x, y_1), \ldots, g(x, y_M)) \in \mathbb{R}^M.$$

Next, we introduce the previously mentioned approximation methods, starting with the kernel-based interpolation.

3.2.1 Kernel-based interpolation

The following lemma provides the existence and uniqueness of an interpolant $s_{f,X} \in V_N$, which also happens to be the best approximation to f from V_N in the native space $H_K(\mathcal{D})$ of K.

Lemma 3.2.1.
Let $K\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ be a positive definite function. Then, for all $f\colon \mathcal{D} \to \mathbb{R}$ and finite point sets $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$, there exists a uniquely determined function $s_{f,X} \in V_N$ that interpolates f in X. It is given by

$$s_{f,X}(x) = K(x, X)K(X, X)^{-1}f(X) = \sum_{j=1}^{N} \alpha_j K(x, x_j), \quad x \in \mathcal{D},$$

i.e., $\alpha = (\alpha_1, \dots, \alpha_N)^T \in \mathbb{R}^N$ is the solution of the linear equation system

$$K(X, X)\alpha = f(X).$$

If $f \in H_K(\mathcal{D})$, then $s_{f,X}$ is the best approximation *of f from V_N in $H_K(\mathcal{D})$, i.e., for all $s \in V_N$ the inequality*

$$\|s_{f,X} - f\|_{H_K(\mathcal{D})} \leq \|s - f\|_{H_K(\mathcal{D})}$$

holds.

Proof. The first part is given as a special case of [55, Theorem 1.3.1], while the best approximation property is shown in [56, Theorem 13.1].

□

Definition 3.2.2 (Kernel-based interpolant)**.**
Let $\mathcal{D} \subseteq \mathbb{R}^d$, $N \in \mathbb{N}$, $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$, and $K\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ be a positive definite function. Then the *kernel interpolant* of $f\colon \mathcal{D} \to \mathbb{R}$ in X using K is defined as

$$\mathcal{I}_{X,K}(f) := K(\cdot, X)K(X, X)^{-1}f(X) \in H_K(\mathcal{D}).$$

The representation $\mathcal{I}_{X,K}(f) = \sum_{j=1}^{N} \chi_j f(x_j)$, where

$$\chi = (\chi_1, \dots, \chi_N) = K(\cdot, X)K(X, X)^{-1},$$

is called the *Lagrange representation* of $\mathcal{I}_{X,K}(f)$ and the functions χ_j are called *Lagrange* or *cardinal* functions.

Remark 3.2.3.

1. As the kernel interpolant of $f\colon \mathcal{D} \to \mathbb{R}$ only uses the pointwise information $f(X)$, it can be calculated in our previously described setting.

2. The best approximation property implies that, by choosing $s = 0 \in V_N$ in Lemma 3.2.1, we obtain
$$\|f - \mathcal{I}_{X,K}(f)\|_{H_K(\mathcal{D})} \leq \|f\|_{H_K(\mathcal{D})}.$$

3. The operator $\mathcal{I}_{X,K}$ is linear. Moreover, as an operator $\mathcal{I}_{X,K}\colon H_K(\mathcal{D}) \to H_K(\mathcal{D})$, it is *stable*, which follows directly from the best approximation property by
$$\|\mathcal{I}_{X,K}(f)\|_{H_K(\mathcal{D})} \leq \|f - \mathcal{I}_{X,K}(f)\|_{H_K(\mathcal{D})} + \|f\|_{H_K(\mathcal{D})} \leq 2\|f\|_{H_K(\mathcal{D})}.$$
Indeed, as shown in [56, Theorem 13.2], the kernel interpolant is the function that interpolates f in X with minimal norm, i.e., one even obtains the bound
$$\|\mathcal{I}_{X,K}(f)\|_{H_K(\mathcal{D})} \leq \|f\|_{H_K(\mathcal{D})}.$$

This last observation suggests that, although we did not require our target function $f\colon \mathcal{D} \to \mathbb{R}$ in Definition 3.2.2 to be contained in $H_K(\mathcal{D})$, this property would be favorable, as the interpolation process is a stable and linear operator in this case. However, in the error estimates presented later we also state a result in which the target function f does not belong to $H_K(\mathcal{D})$.

3.2.2 Penalized least squares approach

For the second approach we assume that the given data is not exact, but contains some additive noise. In particular, we assume to only have access to data values $z_j = f(x_j) + \varepsilon_j$ for $1 \leq j \leq N$. In this scenario it is advisable to rather approximate than directly interpolate the target function, so we end up with the so-called *penalized least squares* method:

Theorem 3.2.4.
Let $K\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ be a positive definite kernel with reproducing kernel Hilbert space $H = H_K(\mathcal{D})$ and $\lambda > 0$ be a regularization parameter. Then, for all sampling points $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ and corresponding data values $Z = (z_1, \dots, z_N)^T \in \mathbb{R}^N$, the penalized least squares *problem consists of finding $s^* \in H$ such that*

$$s^* = \operatorname*{argmin}_{s \in H} \left(\sum_{j=1}^{N} (z_j - s(x_j))^2 + \lambda \|s\|_H^2 \right) \tag{3.3}$$

holds. This minimization problem has a uniquely determined solution, which is given by

$$s^*(x) = K(x, X)(K(X, X) + \lambda I_N)^{-1} Z = \sum_{j=1}^{N} \alpha_j K(x, x_j),$$

where $I_N \in \mathbb{R}^{N \times N}$ denotes the identity matrix and $\alpha = (\alpha_1, \dots, \alpha_N)^T \in \mathbb{R}^N$ is the solution of

$$(K(X, X) + \lambda I_N)\alpha = Z.$$

Proof. The proof is again an application of [55, Theorem 1.3.1].

□

Definition 3.2.5 (Penalized least squares approximant)**.**
Let $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ be a finite point set, $K\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ a symmetric and positive definite function and $\lambda > 0$. Then the *penalized least squares approximation* of $f\colon \mathcal{D} \to \mathbb{R}$ in X with data values $z_j = f(x_j) + \varepsilon_j$ for $1 \leq j \leq N$ using K is given by

$$\mathcal{I}_{X,K,\lambda}(f, \varepsilon) := K(\cdot, X) \left(K(X, X) + \lambda I_N\right)^{-1} \left(f(X) + \varepsilon\right),$$

where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^T \in \mathbb{R}^N$. For exact data values, i.e., $\varepsilon = 0$, we also write

$$\mathcal{I}_{X,K,\lambda}(f) := \mathcal{I}_{X,K,\lambda}(f, 0).$$

The representation $\mathcal{I}_{X,K,\lambda}(f) = \sum_{j=1}^{N} \chi_j (f(x_j) + \varepsilon_j)$, where

$$\chi = (\chi_1, \dots, \chi_N) = K(\cdot, X) \left(K(X, X) + \lambda I_N\right)^{-1},$$

is called the *generalized Lagrange representation* of $\mathcal{I}_{X,K,\lambda}(f, \varepsilon)$ and the functions χ_j the *generalized Lagrange* functions.

Remark 3.2.6.
Considering the limits $\lambda \to 0$ or $\lambda \to \infty$ leads to the following observations: For $\lambda \to 0$, we see that the solution vector α from $(K(X,X) + \lambda I)\alpha = Z$ converges to the solution of $K(X,X)\alpha = Z$ which coincides with the formula for the kernel interpolant of f in the noisy data $z_j = f(x_j) + \varepsilon_j$. Thus, in the limit, the assumption of inexact data is neglected. This motivates us to define $\mathcal{I}_{X,K,0} := \mathcal{I}_{X,K}$; however, we note that for $\lambda = 0$ the minimization problem (3.3) does not have a uniquely determined solution.
On the other hand, for $\lambda \to \infty$, we see directly by the minimization problem that the penalized least squares solution tends to zero. This suggests that a reasonable choice for λ is crucial.
We also want to add that from a numerical perspective, it may be advantageous to use a penalized least squares approach when using exact data $z_j = f(x_j)$ as the kernel matrix $K(X,X)$ can be ill conditioned for large point sets X (see Theorem 3.4.2). This can be mitigated by adding a small regularization term to the matrix.

3.3 Error analysis

In this section, we provide error bounds for the kernel-based approximation methods presented previously. We restrict ourselves to translation invariant positive definite functions $\Phi\colon \mathbb{R}^d \to \mathbb{R}$, whose associated Hilbert spaces are equivalent to a Sobolev space $H^\sigma(\mathbb{R}^d)$ for $\sigma > \frac{d}{2}$. This allows us to use specific advantages of the reproducing kernel Hilbert space, such as the best approximation property, in combination with theory developed for Sobolev spaces. The modern groundwork to prove error estimates for kernel-based approximation methods are so-called *sampling inequalities* (cf. [4], [5], [38]). For a given target function $f \in H^\sigma(\mathcal{D})$, where $\mathcal{D} \subseteq \mathbb{R}^d$ is suitable domain specified below, we can bound weaker norms $\|f\|_{H^\tau(\mathcal{D})}$ with $\tau \leq \sigma$ using pointwise information of the target function at a finite point set $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$. In particular, we provide estimates when X is 'sufficiently dense' in $\mathcal{D}$, which is quantified using the *mesh norm* of X in $\mathcal{D}$ (see [56, Chapter 11]).

Definition 3.3.1 (mesh norm)**.**
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be bounded, $N \in \mathbb{N}$ and $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ a finite point set. Then the *mesh norm* or *fill distance* of X in $\mathcal{D}$ is defined as

$$h_{X,\mathcal{D}} := \sup_{x \in \mathcal{D}} \min_{1 \leq j \leq N} \|x - x_j\|_2.$$

In [40, Definition 6.2], $h_{X,\mathcal{D}}$ is also called the *dispersion* of X.

The mesh norm has a close connection to the Voronoi diagram of X in $\mathcal{D}$. If we define

$$V_k := \{x \in \mathcal{D} \mid \|x - x_k\|_2 \leq \|x - x_i\|_2, 1 \leq i \leq N,\ i \neq k\}, \quad 1 \leq k \leq N,$$

the corresponding Voronoi cells of X in $\mathcal{D}$, then a simple calculation proves that

$$\begin{aligned} h_{X,\mathcal{D}} &= \sup_{x\in\mathcal{D}} \min_{1\le j\le N} \|x - x_j\|_2 = \sup_{x\in\bigcup_{k=1}^N V_k} \min_{1\le j\le N} \|x - x_j\|_2 \\ &= \max_{1\le k\le N} \sup_{x\in V_k} \min_{1\le j\le N} \|x - x_j\|_2 \\ &= \max_{1\le k\le N} \sup_{x\in V_k} \|x - x_k\|_2. \end{aligned}$$

This implies that the mesh norm measures the maximal distance to the boundary of the corresponding Voronoi cells, and therefore describes the radius of the largest ball in $\mathcal{D}$ that contains no point of X.
Using the mesh norm, we can state the following sampling inequality, which summarizes [4, Theorem 4.1] and [5, Theorem 3.2].

Theorem 3.3.2.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. Assume $1 \le p, q, r \le \infty$, and $\sigma \in [0, \infty[$ are given such that the following relations hold:

- *$\sigma \ge d$ if $p = 1$;*
- *$\sigma > \frac{d}{p}$ if $1 < p < \infty$;*
- *$\sigma \in \mathbb{N}$ if $p = \infty$.*

Next define $\ell_0 := \sigma - d\left(\frac{1}{p} - \frac{1}{q}\right)_+$ and set $\gamma := \max\{p, q, r\}$.
Then there exist constants $h_0 = h_0(\mathcal{D}, d, \sigma) > 0$ and $C = C(\mathcal{D}, d, \sigma, p, q, r) > 0$ such that for all finite point sets $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ with mesh norm $h_{X,\mathcal{D}} \le h_0$, every $u \in W^{\sigma,p}(\mathcal{D})$ and

$$\tau \in \begin{cases} [0, \lfloor \ell_0 \rfloor - 1], & 1 \le q < \infty, \\ \tau \in \{0, 1, \dots, \lfloor \ell_0 \rfloor - 1\}, & q = \infty, \end{cases}$$

the estimate

$$|u|_{W^{\tau,q}(\mathcal{D})} \le C \left(h_{X,\mathcal{D}}^{\sigma - \tau - d\left(\frac{1}{p} - \frac{1}{q}\right)_+} |u|_{W^{\sigma,p}(\mathcal{D})} + h_{X,\mathcal{D}}^{\frac{d}{\gamma} - \tau} \|u(X)\|_r \right)$$

holds. If additionally $\sigma \in \mathbb{N}$, then the above bound also holds for $\tau \in [0, \ell_0]$ if $1 \le q < \infty$, or $\tau \in \{0, 1, \dots, \ell_0\}$ if $q = \infty$, when any one of the conditions

- *$p < q < \infty$ and $\ell_0 \in \mathbb{N}$,*
- *$p = q$, or*
- *$p = 1$ and $q = \infty$,*

is fulfilled.

Although this theorem is stated for quite general combinations of indices p, q, r and σ, we consider two special cases, which are summarized in the following corollary.

Corollary 3.3.3.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ a finite point set with mesh norm $h_{X,\mathcal{D}} \leq h_0$. Then, for $\sigma > \frac{d}{2}$, $1 \leq r \leq \infty$ and $u \in H^\sigma(\mathcal{D})$, the estimates

$$\|u\|_{L^2(\mathcal{D})} \leq C \left(h_{X,\mathcal{D}}^\sigma \|u\|_{H^\sigma(\mathcal{D})} + h_{X,\mathcal{D}}^{\frac{d}{\max\{2,r\}}} \|u(X)\|_r \right),$$
$$\|u\|_{C(\mathcal{D})} \leq C \left(h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}} \|u\|_{H^\sigma(\mathcal{D})} + \|u(X)\|_r \right),$$

hold, where u denotes the (Hölder-)continuous representative given by Theorem 2.2.10.

Proof. We want to apply Theorem 3.3.2. Here, $p = 2$ and $\sigma > \frac{d}{2}$, so the prerequisites are satisfied. The first statement is now given with $q = p = 2$ and $\tau = 0$. The second one uses the choices $q = \infty$ and $\tau = 0$. □

These specific error bounds allow us to provide error estimates for the kernel-based interpolation and the penalized least squares approach. The presented result can for example be compared to [33, Theorem 4.7].

Theorem 3.3.4.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary, $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ a finite set of sampling points such that $h_{X,\mathcal{D}} \leq h_0$ with h_0 from Theorem 3.3.2. Furthermore, let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ be a positive definite function with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$. Then, for $\lambda \geq 0$, the interpolation ($\lambda = 0$) and the penalized least squares approach ($\lambda > 0$) error with exact data can be bounded by

$$\begin{aligned} \|f - \mathcal{I}_{X,\Phi,\lambda}(f)\|_{L^2(\mathcal{D})} &\leq C \left(h_{X,\mathcal{D}}^\sigma + h_{X,\mathcal{D}}^{\frac{d}{2}} \sqrt{\lambda} \right) \|f\|_{H^\sigma(\mathcal{D})}, \\ \|f - \mathcal{I}_{X,\Phi,\lambda}(f)\|_{C(\mathcal{D})} &\leq C \left(h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}} + \sqrt{\lambda} \right) \|f\|_{H^\sigma(\mathcal{D})}, \end{aligned} \tag{3.4}$$

for every $f \in H^\sigma(\mathcal{D})$. Assuming $\lambda > 0$ and inexact data values $Z = f(X) + \varepsilon$, where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^T$, the penalized least squares error can be bounded by

$$\begin{aligned} \|f - \mathcal{I}_{X,\Phi,\lambda}(f,\varepsilon)\|_{L^2(\mathcal{D})} &\leq C \left(\left(h_{X,\mathcal{D}}^\sigma + h_{X,\mathcal{D}}^{\frac{d}{2}} \sqrt{\lambda} \right) \|f\|_{H^\sigma(\mathcal{D})} + \left(\frac{h_{X,\mathcal{D}}^\sigma}{\sqrt{\lambda}} + h_{X,\mathcal{D}}^{\frac{d}{2}} \right) \|\varepsilon\|_2 \right), \\ \|f - \mathcal{I}_{X,\Phi,\lambda}(f,\varepsilon)\|_{C(\mathcal{D})} &\leq C \left(\left(h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}} + \sqrt{\lambda} \right) \|f\|_{H^\sigma(\mathcal{D})} + \left(\frac{h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}}}{\sqrt{\lambda}} + 1 \right) \|\varepsilon\|_2 \right). \end{aligned} \tag{3.5}$$

Proof. As an abbreviation, we write $h = h_{X,\mathcal{D}}$ and consider the first case with exact data, i.e., $z_j = f(x_j)$ for $1 \leq j \leq N$. The application of Corollary 3.3.3 yields

$$\begin{aligned} \|f - \mathcal{I}_{X,\Phi,\lambda}(f)\|_{L^2(\mathcal{D})} &\leq C \left(h^\sigma \|f - \mathcal{I}_{X,\Phi,\lambda}(f)\|_{H^\sigma(\mathcal{D})} + h^{\frac{d}{2}} \|f(X) - \mathcal{I}_{X,\Phi,\lambda}(f)(X)\|_2 \right), \\ \|f - \mathcal{I}_{X,\Phi,\lambda}(f)\|_{C(\mathcal{D})} &\leq C \left(h^{\sigma-\frac{d}{2}} \|f - \mathcal{I}_{X,\Phi,\lambda}(f)\|_{H^\sigma(\mathcal{D})} + \|f(X) - \mathcal{I}_{X,\Phi,\lambda}(f)(X)\|_2 \right). \end{aligned} \tag{3.6}$$

Hence, in both cases, we have to bound $\|f-\mathcal{I}_{X,\Phi,\lambda}(f)\|_{H^\sigma(\mathcal{D})}$ and $\|f(X)-\mathcal{I}_{X,\Phi,\lambda}(f)(X)\|_2$. For $\lambda = 0$, the second term vanishes because of the interpolation property, while the first one can be bounded by using the best approximation property of the interpolant by

$$\|f - \mathcal{I}_{X,\Phi}(f)\|_{H^\sigma(\mathcal{D})} \leq C\|f - \mathcal{I}_{X,\Phi}(f)\|_{H_\Phi(\mathcal{D})} \leq C\|f\|_{H_\Phi(\mathcal{D})} \leq C\|f\|_{H^\sigma(\mathcal{D})}.$$

This yields the bound for this case.
For $\lambda > 0$, we consider the first term of (3.6) and estimate

$$\begin{aligned}
&\|f - \mathcal{I}_{X,\Phi,\lambda}(f)\|_{H^\sigma(\mathcal{D})} \\
&\leq \|f\|_{H^\sigma(\mathcal{D})} + \|\mathcal{I}_{X,\Phi,\lambda}(f)\|_{H^\sigma(\mathcal{D})} \\
&\leq \|f\|_{H^\sigma(\mathcal{D})} + C\frac{1}{\sqrt{\lambda}}\left(\|f(X) - \mathcal{I}_{X,\Phi,\lambda}(f)(X)\|_2^2 + \lambda\|\mathcal{I}_{X,\Phi,\lambda}(f)\|_{H_\Phi(\mathcal{D})}^2\right)^{\frac{1}{2}} \\
&\leq C\|f\|_{H^\sigma(\mathcal{D})},
\end{aligned}$$

where the last inequality follows by the minimization property of $\mathcal{I}_{X,\Phi,\lambda}(f)$ given in (3.3). With a similar argument, the second term in (3.6) can also be bounded.

$$\begin{aligned}
\|f(X) - \mathcal{I}_{X,\Phi,\lambda}(f)(X)\|_2 &\leq \left(\|f(X) - \mathcal{I}_{X,\Phi,\lambda}(f)(X)\|_2^2 + \lambda\|\mathcal{I}_{X,\Phi,\lambda}(f)\|_{H_\Phi(\mathcal{D})}^2\right)^{\frac{1}{2}} \\
&\leq \sqrt{\lambda}\|f\|_{H_\Phi(\mathcal{D})} \leq C\sqrt{\lambda}\|f\|_{H^\sigma(\mathcal{D})}.
\end{aligned}$$

Using both estimates in (3.6) leads to the bounds proclaimed for exact data.
Next, we assume $\lambda > 0$ and inexact data $Z = f(X) + \varepsilon$. Using the definition of $\mathcal{I}_{X,\Phi,\lambda}$ allows us to rewrite

$$\begin{aligned}
\mathcal{I}_{X,\Phi,\lambda}(f,\varepsilon) &= \Phi(\cdot,X)\left(\Phi(X,X) + \lambda I_N\right)^{-1}\left(f(X) + \varepsilon\right) \\
&= \Phi(\cdot,X)\left(\Phi(X,X) + \lambda I_N\right)^{-1} f(X) + \Phi(\cdot,X)\left(\Phi(X,X) + \lambda I_N\right)^{-1}\varepsilon \\
&= \mathcal{I}_{X,\Phi,\lambda}(f) + \mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon),
\end{aligned}$$

where the latter summand can be interpreted as a penalized least squares approximant to the zero function using inexact data. Thus, we get

$$\|f - \mathcal{I}_{X,\Phi,\lambda}(f,\varepsilon)\|_{L^2(\mathcal{D})} \leq \|f - \mathcal{I}_{X,\Phi,\lambda}(f)\|_{L^2(\mathcal{D})} + \|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)\|_{L^2(\mathcal{D})},$$

where a similar bound holds for $\|\cdot\|_{C(\mathcal{D})}$ as well. The first term can be estimated by (3.4), and we therefore focus on the second. Again, applying Corollary 3.3.3 yields

$$\begin{aligned}
\|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)\|_{L^2(\mathcal{D})} &\leq Ch^\sigma\|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)\|_{H^\sigma(\mathcal{D})} + h^{\frac{d}{2}}\|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)(X)\|_2, \\
\|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)\|_{C(\mathcal{D})} &\leq Ch^{\sigma-\frac{d}{2}}\|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)\|_{H^\sigma(\mathcal{D})} + \|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)(X)\|_2.
\end{aligned} \tag{3.7}$$

Thus, as before, we have to bound the same terms in both cases. We again use the minimization property of $\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)$ given in (3.3) to obtain

$$\|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)\|_{H^\sigma(\mathcal{D})} \leq C\frac{1}{\sqrt{\lambda}}\left(\|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon) - \varepsilon\|_2^2 + \lambda\,\|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)\|_{H_\Phi(\mathcal{D})}^2\right)^{\frac{1}{2}} \leq C\frac{\|\varepsilon\|_2}{\sqrt{\lambda}}.$$

The second term can be bounded as follows:

$$\begin{aligned}\|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)(X)\|_2 &\leq \|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)(X)-\varepsilon\|_2 + \|\varepsilon\|_2 \\ &\leq \left(\|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)(X)-\varepsilon\|_2^2 + \lambda\,\|\mathcal{I}_{X,\Phi,\lambda}(0,\varepsilon)\|^2_{H_\Phi(\mathcal{D})}\right)^{\frac{1}{2}} + \|\varepsilon\|_2 \\ &\leq 2\|\varepsilon\|_2.\end{aligned}$$

Inserting both estimates back into (3.7) then yields claim for inexact data. □

Remark.
As the previous result suggests, one can mitigate the measurement errors ε_j for $1 \leq j \leq N$ by choosing a regularization parameter $\lambda > 0$. However, there is a certain tradeoff when selecting $\lambda > 0$: On the one hand, a small λ allows us to retain the approximation order, while on the other hand, a large λ alleviates the influence of the measurement error in the approximation. Thus, the choice for a 'suitable' regularization parameter $\lambda > 0$ is crucial. There is a lot of literature addressing this problem, for example [55] and the *cross-validation method*. However, with respect to the previous error bound, the choice $\lambda = h_{X,\mathcal{D}}^{2\sigma-d}$ seems to balance the error estimate reasonably.

In Definition 3.2.2 and Definition 3.2.5, we formulated the interpolation and penalized least squares operator for pointwise defined functions $f\colon \mathcal{D} \to \mathbb{R}$, but proved the error analysis only in the special case where f is an element of $H_\Phi(\mathcal{D})$. However, the correct regularity of the target function f is unknown and usually has to be assumed. This, however, could lead to $f \notin H_\Phi(\mathcal{D})$. In order to relax the assumption $f \in H_\Phi(\mathcal{D})$ and give reasonable error estimates in at least some scenarios, we cite an 'escaping the native space' result, as proven in [39], for example. The result stated there uses two additional quantities, the so-called *separation radius* and the *mesh ratio*:

Definition 3.3.5 (Separation radius and mesh ratio)**.**
Let $X = \{x_1, \ldots, x_N\} \subseteq \mathbb{R}^d$ be a finite point set.

1. The *separation radius* of X is defined as
$$q_X := \frac{1}{2} \min_{\substack{1\leq j,k\leq N \\ j\neq k}} \|x_j - x_k\|_2.$$

2. Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain and $X \subseteq \mathcal{D}$. Then the *mesh ratio* of X in $\mathcal{D}$ is defined as
$$\rho_{X,\mathcal{D}} := \frac{h_{X,\mathcal{D}}}{q_X}.$$

While the separation radius describes the largest radius such that $B_{q_X}(x_j) \cap B_{q_X}(x_k) = 0$ for all $j \neq k$ and therefore measures how far the discrete points of X are spread, the mesh ratio sets this in proportion to the mesh norm. Therefore, $\rho_{X,\mathcal{D}}$ can be seen as a quantity which tells us how 'well' the points are distributed over $\mathcal{D}$. As $\mathcal{D}$ is pathwise connected by our general assumption we see that the separation radius q_X is always less than or equal to the mesh norm $h_{X,\mathcal{D}}$, hence the mesh ratio fulfills $\rho_{X,\mathcal{D}} \geq 1$. Altogether, this results in the following connection of all quantities:

$$q_X \leq h_{X,\mathcal{D}} = \rho_{X,\mathcal{D}}\, q_X. \tag{3.8}$$

We can now cite [39, Theorem 4.2], where an error bound for functions $f \notin H_\Phi(\mathbb{R}^d)$ is provided. By following the steps there and combining them Theorem 3.3.2, one also obtains pointwise error estimates, as it is for example lined out in a more general setting in [17, Theorem 17].

Theorem 3.3.6.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain, $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ a finite point set with mesh norm $h_{X,\mathcal{D}} \leq h_0$, and $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ a positive definite function with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$. Now, let $\tau = k + s$ with $0 \leq s < 1$ and $k > \frac{d}{2}$ such that $\sigma \geq \tau$. Then, for all $f \in H^\tau(\mathcal{D})$ and $0 \leq \mu \leq \tau$, the interpolation error can be estimated by

$$\|f - \mathcal{I}_{X,\Phi}(f)\|_{H^\mu(\mathcal{D})} \leq C h_{X,\mathcal{D}}^{\tau-\mu} \rho_{X,\mathcal{D}}^{\sigma-\tau} \|f\|_{H^\tau(\mathcal{D})},$$
$$\|f - \mathcal{I}_{X,\Phi}(f)\|_{C(\mathcal{D})} \leq C h_{X,\mathcal{D}}^{\tau-\frac{d}{2}} \rho_{X,\mathcal{D}}^{\sigma-\tau} \|f\|_{H^\tau(\mathcal{D})}.$$

3.3.1 Convergence of kernel methods

So far, we only have proved error bounds for the kernel based interpolation and penalized least squares approximation, which are valid for all point sets $X \subseteq \mathcal{D}$ with mesh norm $h_{X,\mathcal{D}} \leq h_0$. In order to properly derive a convergence rate, we have to consider a sequence of point sets $(X_j)_{j\in\mathbb{N}}$ with $X_j \subseteq \mathcal{D}$, $h_{X_j,\mathcal{D}} \leq h_0$ for all $j \in \mathbb{N}$ and an increasing number of elements $|X_j| = N_j$ for $j \in \mathbb{N}$. This way, we obtain a sequence of approximation operators $\mathcal{I}_{X_j,\Phi,\lambda}$ for which the convergence rate can be analyzed. To this end, we want to link the mesh norm of a given point set with the number of points. This can be done as in [56, Proposition 14.1].

Lemma 3.3.7.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain and $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ a finite point set. Then there exist two constants $C_1, C_2 > 0$ independent of X such that

$$C_1 N^{-\frac{1}{d}} \leq h_{X,\mathcal{D}} \leq C_2 \rho_{X,\mathcal{D}} N^{-\frac{1}{d}}.$$

Proof. This proof follows mostly by geometric arguments. First, one can see by the definition of the mesh norm $h_{X,\mathcal{D}}$ that

$$\mathcal{D} \subseteq \bigcup_{j=1}^{N} B_{h_{X,\mathcal{D}}}(x_j).$$

Hence, we compare the volumes of both sets and obtain

$$\mathrm{vol}_d(\mathcal{D}) \leq N h_{X,\mathcal{D}}^d \mathrm{vol}_d(B_1(0)),$$

where vol_d denotes the d-dimensional volume of the given set. Rearranging this inequality leads to

$$h_{X,\mathcal{D}} \geq \left(\frac{\mathrm{vol}_d(\mathcal{D})}{\mathrm{vol}_d(B_1(0))}\right)^{\frac{1}{d}} N^{-\frac{1}{d}},$$

which proves the lower bound with $C_1 := \left(\frac{\mathrm{vol}_d(\mathcal{D})}{\mathrm{vol}_d(B_1(0))}\right)^{\frac{1}{d}}$. For the upper bound, we use that $\mathcal{D} \subseteq \mathbb{R}^d$ is bounded, so that there exists a point $y \in \mathbb{R}^d$ and a radius $R > 0$ such that $\mathcal{D} \subseteq B_R(y)$. If we now consider the balls $B_{q_X}(x_j), 1 \leq j \leq N$, then all of them are disjoint by the definition of the separation radius q_X and are contained in $B_{R+q_X}(y)$. This implies that

$$\bigcup_{j=1}^{N} B_{q_X}(x_j) \subseteq B_{R+q_X}(x_0).$$

Again comparing the volumes of both sets yields the inequality

$$N q_X^d \leq (R + q_X)^d.$$

As, for $x_j, x_k \in X \subseteq \mathcal{D}$ with $j \neq k$, it holds that

$$\|x_j - x_k\|_2 \leq \|x_j - y\|_2 + \|y - x_k\|_2 \leq 2R,$$

we see that $q_X \leq R$. With this estimate, we obtain

$$N \leq \left(1 + \frac{R}{q_X}\right)^d \leq 2^d \left(\frac{R}{q_X}\right)^d = 2^d R^d \left(\frac{h_{X,\mathcal{D}}}{q_X}\right)^d h_{X,\mathcal{D}}^{-d}.$$

Rearranging this inequality yields the upper bound with $C_2 := 2R$. □

When applying the above result to a sequence of point sets $(X_j)_{j\in\mathbb{N}}$, we have to ensure that the mesh ratios given as $\rho_{X_j,\mathcal{D}} = h_{X_j,\mathcal{D}}/q_{X_j}$ are uniformly bounded. Together with (3.8), this leads to the following definition:

Definition 3.3.8 (Quasi-uniformity)**.**
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain and $(X_j)_{j\in\mathbb{N}} \subseteq \mathcal{D}$ a sequence of finite point sets in $\mathcal{D}$. We call $(X_j)_{j\in\mathbb{N}}$ *quasi-uniform* if there exists a constant $c_{qu} > 0$ such that for all $j \in \mathbb{N}$

$$q_{X_j} \leq h_{X_j,\mathcal{D}} \leq c_{qu} q_{X_j}$$

holds, i.e., the mesh ratios $\rho_{X_j,\mathcal{D}}$ are uniformly bounded.

For a sequence of quasi-uniform point sets $(X_j)_{j\in\mathbb{N}}$ with $|X_j| = N_j$, we thus obtain

$$C_1 N_j^{-\frac{1}{d}} \leq h_{X_j,\mathcal{D}} \leq C_2 c_{qu} N_j^{-\frac{1}{d}}. \tag{3.9}$$

We can now, for example, choose kernel-based approximation with exact data and combine Theorem 3.3.4 with (3.9) to obtain the following result:

Theorem 3.3.9.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz-boundary, $(X_j)_{j\in\mathbb{N}}$ a quasi-uniform sequence of finite point sets $X_j \subseteq \mathcal{D}$, mesh norms $h_{X_j,\mathcal{D}} \leq h_0$ and $|X_j| = N_j$ for all $j \in \mathbb{N}$. Let $(\lambda_j)_{j\in\mathbb{N}} \subseteq [0, \infty[$ be a sequence of regularization parameters and $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ a positive definite function with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$. Then, for all $f \in H^\sigma(\mathcal{D})$, the error bound

$$\|f - \mathcal{I}_{X_j,\Phi,\lambda_j}(f)\|_{L^2(\mathcal{D})} \leq C \left(N_j^{-\frac{\sigma}{d}} + \sqrt{\lambda_j} N_j^{-\frac{1}{2}}\right) \|f\|_{H^\sigma(\mathcal{D})}$$

holds. Thus, choosing $\lambda_j \leq N_j^{1-\frac{2\sigma}{d}}$*, we obtain*

$$\|f - \mathcal{I}_{X_j,\Phi,\lambda_j}(f)\|_{L^2(\mathcal{D})} \leq N_j^{-\frac{\sigma}{d}} \|f\|_{H^\sigma(\mathcal{D})}.$$

Proof. Combine Theorem 3.3.4, more specifically (3.4) with (3.9). With the choice of λ_j stated in the theorem, we see that $-\frac{2\sigma-d}{2d} - \frac{1}{2} = -\frac{\sigma}{d}$ and the second claim follows. □

The other results given in Theorem 3.3.4 and Theorem 3.3.6 can of course be expressed in the number of points N_j of X_j as well, which will be omitted here for the sake of brevity.

3.4 Stability of the approximation process

While we mainly showed theoretical properties of our kernel-based approximation methods, we now focus on the numerical properties when explicitly constructing the approximant. In particular, we want to analyze the *condition number* of the kernel matrix $\Phi(X, X)$.

Definition 3.4.1 (Condition number)**.**
Let $A \in \mathbb{R}^{N\times N}$ be an invertible matrix. Then the *condition number* of A is given by

$$\text{cond}_2(A) := \|A\|_2 \|A^{-1}\|_2 = \frac{|\lambda_{max}(A)|}{|\lambda_{min}(A)|},$$

where

$$\|A\|_2 := \|A\|_{2\to 2} = \sup_{\|u\|_2 \leq 1} \|Au\|_2 = \lambda_{\max}(A)$$

denotes the *spectral norm* of A.

In [56, Theorem 12.3], we can find bounds for the condition number of $\Phi(X, X)$.

Theorem 3.4.2.
Let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ *be a positive definite function that possesses a positive Fourier transform* $\widehat{\Phi} \in C(\mathbb{R}^d \setminus \{0\})$*. Let* $X = \{x_1, \ldots, x_N\} \subseteq \mathbb{R}^d$ *be a finite point set and denote by*

$$\varphi_0(M) := \inf_{\|\omega\|_2 \leq 2M} \widehat{\Phi}(\omega),$$

where

$$M \geq \frac{6.38d}{q_X}.$$

Then the largest eigenvalue of $\Phi(X, X)$ *can be bounded from above by*

$$\lambda_{max}(\Phi(X, X)) \leq N\Phi(0),$$

and the smallest eigenvalue from below by

$$\lambda_{min}(\Phi(X, X)) \geq \frac{\varphi_0(M)}{2\Gamma\left(\frac{d}{2}+1\right)} \left(\frac{M}{2^{3/2}}\right)^d,$$

which, in particular, leads to

$$\lambda_{min}(\Phi(X,X)) \geq C_1 q_X^{-d} \inf_{\|\omega\|_2 \leq \frac{C_2}{q_X}} \widehat{\Phi}(\omega),$$

where $C_1, C_2 > 0$ denote suitable constants independent of X.

If we now consider positive definite functions $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ with $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, we get the following result (cf. [56, Corollary 12.8]).

Corollary 3.4.3.
Let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ be a positive definite function that is the reproducing kernel of a Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$. Assume $\mathcal{D} \subseteq \mathbb{R}^d$ is a bounded domain. Then there exists a constant $C = C(\sigma, \mathcal{D}) > 0$ such that for every finite point set $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ with $q_X < C_2$, the smallest eigenvalue of $\Phi(X,X)$ can be bounded from below by

$$\lambda_{min}(\Phi(X,X)) \geq C q_X^{2\sigma - d},$$

and the largest from above by

$$\lambda_{\max}(\Phi(X,X)) \leq C q_X^{-d} \Phi(0).$$

Hence, the condition number is bounded by

$$\mathrm{cond}_2(\Phi(X,X)) \leq C \Phi(0) q_X^{-2\sigma}.$$

Proof. As $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, we know that there exists a constant $c_1 > 0$ such that

$$\widehat{\Phi}(\omega) \geq c_1 (1 + \|\omega\|_2^2)^{-\sigma}.$$

Now, consider the lower bound for the smallest eigenvalue of $\Phi(X,X)$ given in Theorem 3.4.2:

$$\lambda_{min}(\Phi(X,X)) \geq C_1 q_X^{-d} \inf_{\|\omega\| \leq \frac{C_2}{q_X}} \widehat{\Phi}(\omega).$$

If $q_X < C_2$, then $\frac{C_2}{q_X} > 1$, hence we can bound

$$\begin{aligned}
\inf_{\|\omega\|_2 \leq \frac{C_2}{q_X}} \widehat{\Phi}(\omega) &\geq \inf_{\|\omega\|_2 \leq \frac{C_2}{q_X}} c_1 (1 + \|\omega\|_2^2)^{-\sigma} \\
&= c_1 \left(1 + \left(\frac{C_2}{q_X}\right)^2\right)^{-\sigma} \\
&\geq \frac{c_1}{2^\sigma C_2^{2\sigma}} q_X^{2\sigma},
\end{aligned}$$

which yields

$$\lambda_{min}(\Phi(X,X)) \geq C q_X^{2\sigma - d}.$$

For the upper bound of $\lambda_{max}(\Phi(X,X))$, we use Theorem 3.4.2 in combination with Lemma 3.3.7, which leads to

$$\lambda_{\max}(\Phi(X,X)) \leq N \Phi(0) \leq C q_X^{-d} \Phi(0),$$

and therefore the claim. □

Remark.
This result provides a dilemma that is known as the *tradeoff principle*: On the one hand, it would be advisable to use point sets with minimal mesh norm $h_{X,\mathcal{D}}$ in order to get a good error result. On the other hand, when choosing such a point set, the condition number of the kernel matrix $\Phi(X,X)$ deteriorates, as $q_X \leq h_{X,\mathcal{D}}$. Particularly when $q_X \ll h_{X,\mathcal{D}}$, we encounter numerical instabilities when solving the linear equation system. Of course, as already mentioned earlier, one way to mitigate this issue is to choose a regularization parameter $\lambda > 0$ and solve the penalized least squares problem described in Section 3.2.2, as there the smallest eigenvalue of the occurring matrix is given by $\lambda_{min}(\Phi(X,X) + \lambda I_n) = \lambda_{min}(\Phi(X,X)) + \lambda$. In Chapter 6, we will consider another method that circumvents the stability issue using a multilevel scheme based on kernel interpolation.

3.5 Tensorizing approximations

Next, we wish to extend the error estimates of Theorem 3.3.4 to tensorized kernels and the mixed regularity spaces. Of course, one could use the same arguments provided below to generalize the results given in Theorem 3.3.6, but for the sake of brevity, we will omit this case.
First, we see that interpolation with tensorized kernels coincides with the tensor product operator of the individual interpolation operators:

Lemma 3.5.1 (Tensorized interpolant)**.**
Let $\mathcal{D} \subseteq \mathbb{R}^d, \mathcal{T} \subseteq \mathbb{R}^n$ be two domains, $K_1 \colon \mathcal{D}\times\mathcal{D} \to \mathbb{R}$ and $K_2 \colon \mathcal{T}\times\mathcal{T} \to \mathbb{R}$ two reproducing kernels with Hilbert spaces $H_{K_1}(\mathcal{D})$ and $H_{K_2}(\mathcal{T})$, respectively. Denote by

$$K \colon (\mathcal{D}\times\mathcal{T})\times(\mathcal{D}\times\mathcal{T}) \to \mathbb{R}, \quad K((a_1,b_1),(a_2,b_2)) := K_1(a_1,a_2)K_2(b_1,b_2)$$

the tensor product kernel given in Theorem 3.1.11. Then, for all sampling point sets $X = \{x_1,\dots,x_N\} \subseteq \mathcal{D}$ and $Y = \{y_1,\dots,y_M\} \subseteq \mathcal{T}$, the kernel interpolant using K in $X \times Y$ is given by

$$\mathcal{I}_{X\times Y,K} = \mathcal{I}_{X,K_1} \overline{\otimes} \mathcal{I}_{Y,K_2}.$$

Moreover, for $f \in H_{K_1}(\mathcal{D})\overline{\otimes}H_{K_2}(\mathcal{T})$ and $(x,y) \in \mathcal{D}\times\mathcal{T}$, it can be calculated by

$$\mathcal{I}_{X\times Y,K}(f)(x,y) = K_1(x,X)K_1(X,X)^{-1}f(X,Y)K_2(Y,Y)^{-1}K_2(Y,y). \tag{3.10}$$

Proof. If we denote the right-hand side of (3.10) by $\mathcal{I}^*$, i.e.,

$$\mathcal{I}^*(f)(x,y) := K_1(x,X)K_1(X,X)^{-1}f(X,Y)K_2(Y,Y)^{-1}K_2(Y,y), \quad (x,y) \in \mathcal{D}\times\mathcal{T},$$

one can easily see that

$$\mathcal{I}^*(f) \in \operatorname{span}\{K((\cdot,\cdot),(x_j,y_k)) \mid 1 \leq j \leq N,\, 1 \leq k \leq M\} =: V_{N,M}.$$

Additionally, $\mathcal{I}^*(f)$ interpolates f in $X \times Y$, as, for every $1 \leq j \leq N$ and $1 \leq k \leq M$, we obtain

$$\begin{aligned}\mathcal{I}^*(f)(x_j, y_k) &= K_1(x_j, X)K_1(X, X)^{-1} f(X, Y)K_2(Y, Y)^{-1}K_2(Y, y_k) \\ &= \left(e_j^{(N)}\right)^T f(X, Y)e_k^{(M)} \\ &= f(x_j, y_k),\end{aligned}$$

where $e_j^{(N)}$ denotes the j-th unit vector in $\mathbb{R}^N$ and $e_k^{(M)}$ the k-th unit vector in $\mathbb{R}^M$. As the kernel interpolant using K is uniquely determined in $V_{N,M}$ (see Lemma 3.2.1), this leads to the second claim of the theorem.
For the first statement, let $f = \sum_{j=1}^{L} f_1^{(j)} \otimes f_2^{(j)} \in H_{K_1}(\mathcal{D}) \otimes H_{K_2}(\mathcal{T})$. Then

$$\begin{aligned}(\mathcal{I}_{X,K_1} \otimes \mathcal{I}_{Y,K_2})(f) &= \sum_{j=1}^{L} \left(K_1(\cdot, X)K_1(X, X)^{-1} f_1^{(j)}(X)\right) \left(f_2^{(j)}(Y)^T K_2(Y, Y)^{-1}K_2(Y, \cdot)\right) \\ &= K_1(\cdot, X)K_1(X, X)^{-1} f(X, Y)K_2(Y, Y)^{-1}K_2(Y, \cdot) = \mathcal{I}_{X\times Y,K}(f).\end{aligned}$$

Hence, the representation holds for $f \in H_{K_1}(\mathcal{D}) \otimes H_{K_2}(\mathcal{T})$. For an arbitrary function $f \in H_{K_1}(\mathcal{D})\overline{\otimes}H_{K_2}(\mathcal{T})$, there exists a sequence $(f_j)_{j\in\mathbb{N}} \subseteq H_{K_1}(\mathcal{D})\otimes H_{K_2}(\mathcal{T})$ that converges in $H_{K_1}(\mathcal{D})\overline{\otimes}H_{K_2(\mathcal{D})}$ to f, and so by Theorem 3.1.2 also pointwise in $\mathcal{D} \times \mathcal{T}$. Thus, the representation extends to all $f \in H_{K_1}(\mathcal{D})\overline{\otimes}H_{K_2}(\mathcal{T})$ in the tensor product space. □

At this point, it is important to note that a similar statement does not hold for the penalized least squares approach, i.e., $\mathcal{I}_{X\times Y,K,\lambda} \neq \mathcal{I}_{X,K_1,\lambda}\overline{\otimes}\mathcal{I}_{Y,K_2,\lambda}$ for all $\lambda > 0$, as a simple calculation shows. Nonetheless, we can at least derive the same representation for the tensorized penalized least squares approximation.

Corollary 3.5.2.
In the setting of Lemma 3.5.1, let $\lambda_1, \lambda_2 > 0$. Then we can rewrite $\mathcal{I}_{X,K_1,\lambda_1}\overline{\otimes}\mathcal{I}_{Y,K_2,\lambda_2}$ as

$$\begin{aligned}&\mathcal{I}_{X,K_1,\lambda_1}\overline{\otimes}\mathcal{I}_{Y,K_2,\lambda_2}(f)(x, y) \\ &= K_1(x, X) \left(K_1(X, X) + \lambda_1 I_N\right)^{-1} f(X, Y) \left(K_2(Y, Y) + \lambda_2 I_M\right)^{-1} K_2(Y, y),\end{aligned}$$

for $f \in H_{K_1}(\mathcal{D})\overline{\otimes}H_{K_2}(\mathcal{T})$ and $(x, y) \in \mathcal{D} \times \mathcal{T}$.

Proof. Replace the formula of the kernel interpolation with the one of the penalized least squares solution, and proceed as in the proof on Lemma 3.5.1. □

Definition and Remark 3.5.3.
When considering inexact data, we see that

$$\begin{aligned}&\mathcal{I}_{X,K_1,\lambda_1}(\cdot, \varepsilon^{(1)})\overline{\otimes}\mathcal{I}_{Y,K_2,\lambda_2}(\cdot, \varepsilon^{(2)})(f)(x, y) \\ &= K_1(x, X) \left(K_1(X, X) + \lambda_1 I_N\right)^{-1} \left(f(X, Y) + \varepsilon^{(1)} \otimes \varepsilon^{(2)}\right) \left(K_2(Y, Y) + \lambda_2 I_M\right)^{-1} K_2(Y, y),\end{aligned}$$

for $\varepsilon^{(1)} \in \mathbb{R}^N$ and $\varepsilon^{(2)} \in \mathbb{R}^M$ holds. However, if we consider general inexact data values $Z = f(X, Y) + \mathcal{E}$, i.e., $z_{j,k} = f(x_j, y_k) + \varepsilon_{j,k}$ for $1 \leq j \leq N, 1 \leq k \leq M$, no such tensor decomposition holds.

Nonetheless, the above formula yields an intrinsic way to define an approximation operator in this case:

$$\begin{aligned}&\mathcal{I}_{X\times Y,K,\lambda_1,\lambda_2}(f,\mathcal{E})(x,y)\\&:= K_1(x,X)\left(K_1(X,X)+\lambda_1 I_N\right)^{-1}\left(f(X,Y)+\mathcal{E}\right)\left(K_2(Y,Y)+\lambda_2 I_M\right)^{-1}K_2(Y,y),\end{aligned}$$

for $f \in H_{K_1}(\mathcal{D})\overline{\otimes}H_{K_2}(\mathcal{T})$ and $(x,y) \in \mathcal{D}\times\mathcal{T}$.

As before, we can express the kernel approximation using the (generalized) Lagrange functions.

Corollary 3.5.4.
Let $\mathcal{D} \subseteq \mathbb{R}^d$, $\mathcal{T} \subseteq \mathbb{R}^n$, and $K_1\colon \mathcal{D}\times\mathcal{D} \to \mathbb{R}$ and $K_2\colon \mathcal{T}\times\mathcal{T} \to \mathbb{R}$ be reproducing kernels of $H_{K_1}(\mathcal{D})$ and $H_{K_2}(\mathcal{T})$, respectively. Furthermore, let $X = \{x_1,\dots,x_N\} \subseteq \mathcal{D}$ and $Y = \{y_1,\dots,y_M\} \subseteq \mathcal{T}$ be finite point sets and $\lambda_1,\lambda_2 \geq 0$ be given. Then the tensor product of kernel-based interpolations and penalized least squares approximations can be expressed as

$$\begin{aligned}\mathcal{I}_{X,K_1,\lambda_1}\overline{\otimes}\mathcal{I}_{Y,K_2,\lambda_2}(f,\mathcal{E})(x,y) &= (\chi_{1,1},\dots,\chi_{N,1})\left(f(X,Y)+\mathcal{E}\right)(\chi_{1,2},\dots,\chi_{M,2})^T\\&= \sum_{j=1}^{N}\sum_{k=1}^{M}\left(f(x_j,y_k)+\varepsilon_{j,k}\right)\chi_{j,1}(x)\chi_{k,2}(y),\end{aligned}$$

where $\mathcal{E} = (\varepsilon_{j,k}) \in \mathbb{R}^{N\times M}$. For $1 \leq j \leq N$, the functions $\chi_{j,1}$ denote the (generalized) Lagrange functions in X using K_1, and for $1 \leq k \leq M$, the functions $\chi_{k,2}$ the ones in Y using K_2. Hence, the Lagrange functions of this approximation are given by the tensor product $\chi_{j,1}\otimes\chi_{k,2}$ for $1 \leq j \leq N$ and $1 \leq k \leq M$. For $\lambda_1 = \lambda_2 = 0$, these functions are cardinal on $X\times Y$.

Proof. The representation follows immediately from Lemma 3.5.1 and Corollary 3.5.2. The cardinality for $\lambda_1 = \lambda_2 = 0$ of $\chi_{j,1}\otimes\chi_{k,2}$ follows from the one of $\chi_{j,1}$ and $\chi_{k,2}$, respectively, for $1 \leq j \leq N$ and $1 \leq k \leq M$. □

3.5.1 Computational cost

The formulas in Lemma 3.5.1 and Corollary 3.5.2 provide an explicit way of computing the approximants using tensorized kernels. This can be used to analyze the computational cost for their calculation. In the following, we focus on kernel-based interpolation for simplicity, but note that the ideas can easily be generalized to the tensorized penalized least squares approach presented in Corollary 3.5.4 using exact or inexact data.
The setting here is the following: Let $\mathcal{D} \subseteq \mathbb{R}^d$ and $\mathcal{T} \subseteq \mathbb{R}^n$ be two domains and let two positive definite reproducing kernels $K_1\colon \mathcal{D}\times\mathcal{D} \to \mathbb{R}$ and $K_2\colon \mathcal{T}\times\mathcal{T} \to \mathbb{R}$ of $H_{K_1}(\mathcal{D})$ and $H_{K_2}(\mathcal{T})$, respectively, be given. Assume that the kernel evaluation of both kernels in one point can be done in $\mathcal{O}(1)$ time. Let $X = \{x_1,\dots,x_N\} \subseteq \mathcal{D}$ and $Y = \{y_1,\dots,y_M\} \subseteq \mathcal{T}$ be two finite point sets. Our goal is to compute the kernel interpolation of a target function $f\colon \mathcal{D}\times\mathcal{T} \to \mathbb{R}$ using $K := K_1\otimes K_2$, which by Lemma 3.5.1 can be calculated as

$$\mathcal{I}_{X\times Y,K}(f)(x,y) = K_1(x,X)K_1(X,X)^{-1}f(X,Y)K_2(Y,Y)^{-1}K_2(Y,y).$$

We denote the asymptotic time to compute the approximant with $\text{cost}(\mathcal{I}_{X\times Y,K})$. First, we need to calculate the evaluation vectors $K_1(x,X) \in \mathbb{R}^N$ and $K_2(Y,y) \in \mathbb{R}^M$, which, by assumption, can be done in $\mathcal{O}(N)$ and $\mathcal{O}(M)$ time, respectively. Next, we focus on calculating the matrix

$$C := K_1(X,X)^{-1} f(X,Y) K_2(Y,Y)^{-1} \in \mathbb{R}^{N\times M}.$$

Computing this matrix can be accomplished by solving two matrix equations, i.e., multiple linear equation systems. First, we consider the matrix equation

$$K_1(X,X)A = f(X,Y),$$

where $A \in \mathbb{R}^{N\times M}$. This can be solved separately for each row, i.e., for $1 \leq k \leq M$ we solve the linear equation system

$$K_1(X,X)a_k = f(X,y_k) \in \mathbb{R}^N,$$

where $a_k \in \mathbb{R}^N$, and set $A = (a_1,\ldots,a_M) \in \mathbb{R}^{N\times M}$. As $K_1(X,X) \in \mathbb{R}^{N\times N}$ is a positive definite matrix, these linear equation systems can be solved using a CG-solver in at most $\mathcal{O}(N^3)$ time for each right-hand side, which leads to a total cost of $\mathcal{O}(N^3M)$ for this calculation step. The matrix $K_1(X,X)$ itself can be constructed in $\mathcal{O}(N^2)$ time by assumption, yielding asymptotic cost of $\mathcal{O}(N^3M)$ time.
Plugging A back in, we therefore obtain the equation

$$K_1(X,X)CK_2(Y,Y) = f(X,Y) = K_1(X,X)A,$$

which simplifies to

$$CK_2(Y,Y) = A.$$

This can next be transformed into another matrix equation by simply transposing this equation. As $K_2(Y,Y)$ is symmetric, we then need to solve

$$K_2(Y,Y)B = A^T.$$

As before, the matrix $K_2(Y,Y)$ can be constructed in $\mathcal{O}(M^2)$ time. The solution matrix $B \in \mathbb{R}^{M\times N}$ can again be computed separately for every row of $A^T \in \mathbb{R}^{M\times N}$ using a CG-solver, which leads to a computational cost of at most $\mathcal{O}(M^3)$ for each row and, in total, to $\mathcal{O}(NM^3)$ time. Finally, the matrix C is given by $C = B^T$. The total computational cost is therefore given by $\mathcal{O}(N^3M + NM^3)$.
In order to evaluate the kernel interpolation in a fixed point $(x,y) \in \mathcal{D}\times\mathcal{T}$, we have to compute the vector-matrix-vector product

$$\mathcal{I}_{X\times Y,K}(f) = K_1(x,X)CK_2(Y,y),$$

which can be achieved in $\mathcal{O}(NM)$ time, leading to a total computational cost of

$$\text{cost}(\mathcal{I}_{X\times Y,K}) = \mathcal{O}(N^3M + MN^3).$$

3.5.2 Error estimates using tensorized kernels

In order to prove error estimates for tensorized kernels, we seek a way to reformulate the error operator. This can be done as follows:

Lemma 3.5.5.
Let U_1, U_2, V_1 and V_2 be separable Banach spaces, and, for $j = 1, 2$, let there be linear and bounded operators $A_j, B_j \colon U_j \to V_2$. Assume furthermore that $U := U_1 \overline{\otimes} U_2$ and $V := V_1 \overline{\otimes} V_2$ are equipped with norms such that U is compatible with V. Then the following decomposition holds:

$$A_1 \overline{\otimes} A_2 - B_1 \overline{\otimes} B_2 = (A_1 - B_1) \overline{\otimes} A_2 + A_1 \overline{\otimes} (A_2 - B_2) - (A_1 - B_1) \overline{\otimes} (A_2 - B_2).$$

Proof. We first consider functions in $U_1 \otimes U_2$, where the tensor product operator is given by Theorem 2.4.20. There the equality follows by expanding the right-hand side. For arbitrary functions in $U_1 \overline{\otimes} U_2$ the equality stays valid. □

This lemma allows us to prove the following error result.

Theorem 3.5.6.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ and $\mathcal{T} \subseteq \mathbb{R}^n$ be two bounded domains. Assume that two finite point sets $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ and $Y = \{y_1, \dots, y_M\} \subseteq \mathcal{T}$ with mesh norms $h_{X,\mathcal{D}} \leq h_{0,\mathcal{D}}$ and $h_{Y,\mathcal{T}} \leq h_{0,\mathcal{T}}$, respectively, are given. Let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ and $\Psi \in L^1(\mathbb{R}^n) \cap C(\mathbb{R}^n)$ be two positive definite functions such that the native spaces fulfill $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$, and $H_\Psi(\mathbb{R}^n) \cong H^\tau(\mathbb{R}^n)$, where $\tau > \frac{n}{2}$. Then, for all $f \in H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D} \times \mathcal{T})$, the following estimates hold:

1. *For $\lambda_1, \lambda_2 \geq 0$, the interpolation ($\lambda_j = 0$) and penalized least squares ($\lambda_j > 0$) error can be bounded by*

$$\begin{aligned}
&\|f - (\mathcal{I}_{X,\Phi,\lambda_1} \overline{\otimes} \mathcal{I}_{Y,\Psi,\lambda_2})(f)\|_{L^2(\mathcal{D}\times\mathcal{T})} \\
&\leq C \left[\left(h^\sigma_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}} \sqrt{\lambda_1} \right) + \left(h^\tau_{Y,\mathcal{T}} + h^{\frac{n}{2}}_{Y,\mathcal{T}} \sqrt{\lambda_2} \right) \right. \\
&\qquad \left. + \left(h^\sigma_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}} \sqrt{\lambda_1} \right) \left(h^\tau_{Y,\mathcal{T}} + h^{\frac{n}{2}}_{Y,\mathcal{T}} \sqrt{\lambda_2} \right) \right] \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})},
\end{aligned}$$

   and

$$\begin{aligned}
&\|f - (\mathcal{I}_{X,\Phi,\lambda_1} \overline{\otimes} \mathcal{I}_{Y,\Psi,\lambda_2})(f)\|_{C(\mathcal{D}\times\mathcal{T})} \\
&\leq C \left[\left(h^{\sigma-\frac{d}{2}}_{X,\mathcal{D}} + \sqrt{\lambda_1} \right) + \left(h^{\tau-\frac{n}{2}}_{Y,\mathcal{T}} + \sqrt{\lambda_2} \right) \right. \\
&\qquad \left. + \left(h^{\sigma-\frac{d}{2}}_{X,\mathcal{D}} + \sqrt{\lambda_1} \right) \left(h^{\tau-\frac{n}{2}}_{Y,\mathcal{T}} + \sqrt{\lambda_2} \right) \right] \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}(\mathcal{D}\times\mathcal{T})}}.
\end{aligned}$$

2. *For $\lambda_1, \lambda_2 > 0$ and inexact data $z_{j,k} = f(x_j, y_k) + \varepsilon^{(1)}_j \varepsilon^{(2)}_k m$ where $1 \leq j \leq N$ and $1 \leq k \leq M$, hence $Z = f(X, Y) + \varepsilon^{(1)} \otimes \varepsilon^{(2)} \in \mathbb{R}^{N \times M}$, the penalized least squares*

error can be bounded by

$$\begin{aligned}
&\left\| f - \mathcal{I}_{X\times Y,\Phi\otimes\Psi,\lambda_1,\lambda_2}(f, \varepsilon^{(1)} \otimes \varepsilon^{(2)}) \right\|_{L^2(\mathcal{D}\times\mathcal{T})} \\
&\quad \leq \| f - (\mathcal{I}_{X,\Phi,\lambda_1} \overline{\otimes} \mathcal{I}_{Y,\Psi,\lambda_2})(f) \|_{L^2(\mathcal{D}\times\mathcal{T})} \\
&\qquad + C \left(\frac{h_{X,\mathcal{D}}^{\sigma}}{\sqrt{\lambda_1}} + h_{X,\mathcal{D}}^{\frac{d}{2}} \right) \left(\frac{h_{Y,\mathcal{T}}^{\tau}}{\sqrt{\lambda_2}} + h_{Y,\mathcal{T}}^{\frac{n}{2}} \right) \left\| \varepsilon^{(1)} \right\|_2 \left\| \varepsilon^{(2)} \right\|_2 ,
\end{aligned}$$

and similarly a pointwise error estimate holds:

$$\begin{aligned}
&\left\| f - \mathcal{I}_{X\times Y,\Phi\otimes\Psi,\lambda_1,\lambda_2}(f, \varepsilon^{(1)} \otimes \varepsilon^{(2)}) \right\|_{C(\mathcal{D}\times\mathcal{T})} \\
&\quad \leq \| f - (\mathcal{I}_{X,\Phi,\lambda_1} \overline{\otimes} \mathcal{I}_{Y,\Psi,\lambda_2})(f) \|_{C(\mathcal{D}\times\mathcal{T})} \\
&\qquad + C \left(\frac{h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}}}{\sqrt{\lambda_1}} + 1 \right) \left(\frac{h_{Y,\mathcal{T}}^{\tau-\frac{n}{2}}}{\sqrt{\lambda_2}} + 1 \right) \left\| \varepsilon^{(1)} \right\|_2 \left\| \varepsilon^{(2)} \right\|_2 .
\end{aligned}$$

Proof. As an abbreviation, we denote the residuals in $\mathcal{D}$ and $\mathcal{T}$ as

$$\mathcal{R}_{\mathcal{D}} := Id_{\mathcal{D}} - \mathcal{I}_{X,\Phi,\lambda_1}, \quad \mathcal{R}_{\mathcal{T}} := Id_{\mathcal{T}} - \mathcal{I}_{Y,\Psi,\lambda_2},$$

where $Id_{\mathcal{D}}$ denotes the identity operator on $L^2(\mathcal{D})$, $Id_{\mathcal{T}}$ the one on $L^2(\mathcal{T})$, and $Id_{\mathcal{D}\times\mathcal{T}}$ the one on $L^2(\mathcal{D}\times\mathcal{T})$.
Since $Id_{\mathcal{D}\times\mathcal{T}} = Id_{\mathcal{D}} \overline{\otimes} Id_{\mathcal{T}}$, we can decompose the error functional $Id_{\mathcal{D}\times\mathcal{T}} - \mathcal{I}_{X\times Y,\Phi\otimes\Psi,\lambda_1,\lambda_2}$ as

$$Id_{\mathcal{D}\times\mathcal{T}} - \mathcal{I}_{X,\Phi,\lambda_1} \overline{\otimes} \mathcal{I}_{Y,\Psi,\lambda_2} = \mathcal{R}_{\mathcal{D}} \overline{\otimes} Id_{\mathcal{T}} + Id_{\mathcal{D}} \overline{\otimes} \mathcal{R}_{\mathcal{T}} - \mathcal{R}_{\mathcal{D}} \overline{\otimes} \mathcal{R}_{\mathcal{T}}$$

using Lemma 3.5.5.
With this representation, we can split the error in $L^2(\mathcal{D}\times\mathcal{T})$ into separate parts:

$$\begin{aligned}
&\| f - \mathcal{I}_{X,\Phi,\lambda_1} \overline{\otimes} \mathcal{I}_{Y,\Psi,\lambda_2}(f) \|_{L^2(\mathcal{D}\times\mathcal{T})} \\
&\leq \| \mathcal{R}_{\mathcal{D}} \overline{\otimes} Id_{\mathcal{T}}(f) \|_{L^2(\mathcal{D}\times\mathcal{T})} + \| Id_{\mathcal{D}} \overline{\otimes} \mathcal{R}_{\mathcal{T}}(f) \|_{L^2(\mathcal{D}\times\mathcal{T})} + \| \mathcal{R}_{\mathcal{D}} \overline{\otimes} \mathcal{R}_{\mathcal{T}}(f) \|_{L^2(\mathcal{D}\times\mathcal{T})} \\
&\leq \| \mathcal{R}_{\mathcal{D}} \overline{\otimes} Id_{\mathcal{T}} \|_{H^{\sigma,0}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T}) \to L^2(\mathcal{D}\times\mathcal{T})} \| f \|_{H^{\sigma,0}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} \\
&\quad + \| Id_{\mathcal{D}} \overline{\otimes} \mathcal{R}_{\mathcal{T}} \|_{H^{0,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T}) \to L^2(\mathcal{D}\times\mathcal{T})} \| f \|_{H^{0,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} \\
&\quad + \| \mathcal{R}_{\mathcal{D}} \overline{\otimes} \mathcal{R}_{\mathcal{T}} \|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T}) \to L^2(\mathcal{D}\times\mathcal{T})} \| f \|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} .
\end{aligned}$$

As all occurring spaces are tensor product spaces of Hilbert spaces, we can use the compatibility of the used norms given in Proposition 2.4.23 to see for example that

$$\| \mathcal{R}_{\mathcal{D}} \overline{\otimes} Id_{\mathcal{T}} \|_{H^{\sigma,0}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T}) \to L^2(\mathcal{D}\times\mathcal{T})} \leq \| \mathcal{R}_{\mathcal{D}} \|_{H^{\sigma}(\mathcal{D}) \to L^2(\mathcal{D})} \| Id_{\mathcal{T}} \|_{L^2(\mathcal{T}) \to L^2(\mathcal{T})} .$$

While the latter operator norm is equal to one, the first one can be bounded by

$$\begin{aligned}
\| \mathcal{R}_{\mathcal{D}} \|_{H^{\sigma}(\mathcal{D}) \to L^2(\mathcal{D})} &= \sup_{\substack{g \in H^{\sigma}(\mathcal{D}), \\ \|g\|_{H^{\sigma}(\mathcal{D})} \leq 1}} \| g - \mathcal{I}_{X,\Phi,\lambda_1}(g) \|_{L^2(\mathcal{D})} \\
&\leq \sup_{\substack{g \in H^{\sigma}(\mathcal{D}), \\ \|g\|_{H^{\sigma}(\mathcal{D})} \leq 1}} C \left(h_{X,\mathcal{D}}^{\sigma} + h_{X,\mathcal{D}}^{\frac{d}{2}} \sqrt{\lambda_1} \right) \| g \|_{H^{\sigma}(\mathcal{D})} \\
&\leq C \left(h_{X,\mathcal{D}}^{\sigma} + h_{X,\mathcal{D}}^{\frac{d}{2}} \sqrt{\lambda_1} \right)
\end{aligned}$$

using Theorem 3.3.4. Similarly, the other operator norms can be estimated by

$$\begin{aligned}\|Id_{\mathcal{D}}\overline{\otimes}\mathcal{R}_{\mathcal{T}}\|_{H^{0,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})\to L^2(\mathcal{D}\times\mathcal{T})} &\leq C\left(h^{\tau}_{Y,\mathcal{T}}+h^{\frac{n}{2}}_{Y,\mathcal{T}}\sqrt{\lambda}_2\right),\\ \|\mathcal{R}_{\mathcal{D}}\overline{\otimes}\mathcal{R}_{\mathcal{T}}\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})\to L^2(\mathcal{D}\times\mathcal{T})} &\leq C\left(h^{\sigma}_{X,\mathcal{D}}+h^{\frac{d}{2}}_{X,\mathcal{D}}\sqrt{\lambda_1}\right)\left(h^{\tau}_{Y,\mathcal{T}}+h^{\frac{n}{2}}_{Y,\mathcal{T}}\sqrt{\lambda}_2\right).\end{aligned}$$

Next, we can bound the occurring norms of f by $\|f\|_{H^{\sigma,0}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} \leq \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}$ and $\|f\|_{H^{0,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} \leq \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}$, which proves the first claim.
For the pointwise error bound on $\mathcal{D}\times\mathcal{T}$, we proceed in similar fashion using the error representation Lemma 3.5.5 in combination with $C(\mathcal{D}\times\mathcal{T}) = C(\mathcal{D})\overline{\otimes}_{\vee}C(\mathcal{T})$:

$$\begin{aligned}&\|f-\mathcal{I}_{X,\Phi,\lambda_1}\overline{\otimes}\mathcal{I}_{Y,\Psi,\lambda_2}(f)\|_{C(\mathcal{D}\times\mathcal{T})}\\ &\leq \|\mathcal{R}_{\mathcal{D}}\overline{\otimes}Id_{\mathcal{T}}(f)\|_{C(\mathcal{D}\times\mathcal{T})}+\|Id_{\mathcal{D}}\overline{\otimes}\mathcal{R}_{\mathcal{T}}(f)\|_{C(\mathcal{D}\times\mathcal{T})}+\|\mathcal{R}_{\mathcal{D}}\overline{\otimes}\mathcal{R}_{\mathcal{T}}(f)\|_{C(\mathcal{D}\times\mathcal{T})}\\ &\leq \|\mathcal{R}_{\mathcal{D}}\|_{H^{\sigma}(\mathcal{D})\to C(\mathcal{D})}\|f\|_{H^{\sigma}(\mathcal{D})\overline{\otimes}_{\vee}C(\mathcal{T})}+\|\mathcal{R}\|_{H^{\tau}(\mathcal{T})\to C(\mathcal{T})}\|f\|_{C(\mathcal{D})\overline{\otimes}_{\vee}H^{\tau}(\mathcal{D})}\\ &\quad+\|\mathcal{R}_{\mathcal{D}}\|_{H^{\sigma}(\mathcal{D})\to C(\mathcal{D})}\|\mathcal{R}\|_{H^{\tau}(\mathcal{T})\to C(\mathcal{T})}\|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}.\end{aligned}$$

As before, we applied Proposition 2.4.23 to estimate the occurring operator norms on the tensor product spaces by the ones on $\mathcal{D}$ and $\mathcal{T}$, respectively. These, on the other hand, can be estimated as before using the error estimates in Theorem 3.3.4. By the embedding Theorem 2.2.10, we also obtain

$$\begin{aligned}\|f\|_{C(\mathcal{D})\overline{\otimes}_{\vee}H^{\tau}(\mathcal{T})} &\leq \|Id_{\mathcal{D}}\|_{H^{\sigma}(\mathcal{D})\to C(\mathcal{D})}\|Id_{\mathcal{T}}\|_{H^{\tau}(\mathcal{T})\to H^{\tau}(\mathcal{T})}\|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}\\ &\leq C\|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}.\end{aligned}$$

Thus, the last claim concerning exact data follows.
Last, for inexact data values $Z = (z_{j,k}) = \left(f(x_j,y_k)+\varepsilon^{(1)}_j\varepsilon^{(2)}_k\right) \in \mathbb{R}^{N\times M}$ with tensor noise, we can rewrite the approximant as

$$\begin{aligned}&\mathcal{I}_{X\times Y,\Phi\otimes\Psi,\lambda_1,\lambda_2}(f,\varepsilon^{(1)}\otimes\varepsilon^{(2)})(x,y)\\ &=\Phi(x,X)\left(\Phi(X,X)+\lambda_1 I_N\right)^{-1}\left(f(X,Y)+\varepsilon^{(1)}\otimes\varepsilon^{(2)}\right)\left(\Psi(Y,Y)+\lambda_2 I_M\right)^{-1}\Psi(Y,y)\\ &=\Phi(x,X)\left(\Phi(X,X)+\lambda_1 I_N\right)^{-1}f(X,Y)\left(\Psi(Y,Y)+\lambda_2 I_M\right)^{-1}\Psi(Y,y)\\ &\quad+\Phi(x,X)\left(\Phi(X,X)+\lambda_1 I_N\right)^{-1}\left(\varepsilon^{(1)}\otimes\varepsilon^{(2)}\right)\left(\Psi(Y,Y)+\lambda_2 I_M\right)^{-1}\Psi(Y,y)\\ &=\mathcal{I}_{X,\Phi,\lambda_1}\overline{\otimes}\mathcal{I}_{Y,\Psi,\lambda_2}(f)(x,y)\\ &\quad+\left(\Phi(x,X)\left(\Phi(X,X)+\lambda_1 I_N\right)^{-1}\varepsilon^{(1)}\right)\left(\Psi(y,Y)\left(\Psi(Y,Y)+\lambda_2 I_M\right)^{-1}\varepsilon^{(2)}\right).\end{aligned}$$

The latter equation follows from the fact that $v^T(a\otimes b)w = (v^Ta)(b^Tw)$ for arbitrary vectors $v,a\in\mathbb{R}^d$ and $b,w\in\mathbb{R}^n$. While the first term describes the tensorized penalized least squares approximant to f using exact data, the second one can be interpreted as the product of the approximations of $u_1 := 0 \in H^{\sigma}(\mathcal{D})$ and $u_2 := 0 \in H^{\tau}(\mathcal{D})$ with measurement errors $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$, respectively. This yields the formula

$$\begin{aligned}&\mathcal{I}_{X,\Phi,\lambda_1}\overline{\otimes}\mathcal{I}_{Y,\Psi,\lambda_2}\left(f,\varepsilon^{(1)}\otimes\varepsilon^{(2)}\right)\\ &=\mathcal{I}_{X,\Phi,\lambda_1}\overline{\otimes}\mathcal{I}_{Y,\Psi,\lambda_2}(f)(x,y)+\left(\mathcal{I}_{X,\Phi,\lambda_1}\left(0,\varepsilon^{(1)}\right)\right)\otimes\left(\mathcal{I}_{Y,\Psi,\lambda_2}\left(0,\varepsilon^{(2)}\right)\right)(x,y).\end{aligned}$$

With this representation in the error, we estimate further

$$\begin{aligned}
&\|f - \mathcal{I}_{X,\Phi,\lambda_1} \overline{\otimes} \mathcal{I}_{Y,\Psi,\lambda_2} \left(f, \varepsilon^{(1)} \otimes \varepsilon^{(2)}\right) \|_{L^2(\mathcal{D}\times\mathcal{T})} \\
&\leq \|f - \mathcal{I}_{X,\Phi,\lambda_1} \overline{\otimes} \mathcal{I}_{Y,\Psi,\lambda_2}(f)\|_{L^2(\mathcal{D}\times\mathcal{T})} + \left\| \left(\mathcal{I}_{X,\Phi,\lambda_1}(0, \varepsilon^{(1)})\right) \otimes \left(\mathcal{I}_{Y,\Psi,\lambda_2}(0, \varepsilon^{(2)})\right) \right\|_{L^2(\mathcal{D}\times\mathcal{T})}.
\end{aligned}$$

It remains to bound the latter term. The main idea here is to use the cross norm property of $L^2(\mathcal{D} \times \mathcal{T})$ by Theorem 2.4.17 to obtain

$$\begin{aligned}
&\left\| \left(\mathcal{I}_{X,\Phi,\lambda_1}(0, \varepsilon^{(1)})\right) \otimes \left(\mathcal{I}_{Y,\Psi,\lambda_2}(0, \varepsilon^{(2)})\right) \right\|_{L^2(\mathcal{D}\times\mathcal{T})} \\
&= \left\| \mathcal{I}_{X,\Phi,\lambda_1}(0, \varepsilon^{(1)}) \right\|_{L^2(\mathcal{D})} \left\| \mathcal{I}_{Y,\Psi,\lambda_2}(0, \varepsilon^{(2)}) \right\|_{L^2(\mathcal{T})}.
\end{aligned}$$

These norms can now be bounded using (3.5) from Theorem 3.3.4:

$$\begin{aligned}
\|\mathcal{I}_{X,\Phi,\lambda_1}(0, \varepsilon^{(1)})\|_{L^2(\mathcal{D})} &\leq C \left(\frac{h^{\sigma}_{X,\mathcal{D}}}{\sqrt{\lambda_1}} + h^{\frac{d}{2}}_{X,\mathcal{D}} \right) \left\| \varepsilon^{(1)} \right\|_2, \\
\left\| \mathcal{I}_{Y,\Psi,\lambda_2}(0, \varepsilon^{(2)}) \right\|_{L^2(\mathcal{T})} &\leq C \left(\frac{h^{\tau}_{Y,\mathcal{T}}}{\sqrt{\lambda_2}} + h^{\frac{n}{2}}_{Y,\mathcal{T}} \right) \left\| \varepsilon^{(2)} \right\|_2,
\end{aligned}$$

which altogether concludes the stated result. By replacing the $L^2(\mathcal{D} \times \mathcal{T})$ norm with $C(\mathcal{D} \times \mathcal{T})$, we can obtain a similar result by using the second inequality of (3.5) and the cross norm property of the injective norm given in Definition and Remark 2.4.10. □

4 Random Fields

Our aforementioned goal is to apply the approximation methods to *function-valued* random variables, also known as *random fields* in the literature. In this chapter, we first provide a short introduction to random fields and their theoretical background. The theory relies on the well-known theory for real-valued random variables, which will not be covered in this thesis, but can for example be found in [7], [27], as well as on the Bochner integrability and Bochner spaces introduced in Section 2.3. After this brief introduction, we continue in Section 4.2 with a Fourier type expansion for specific random fields which is known as the *Karhunen-Loève expansion*. Finally, in Section 4.3, we give several results concerning the so-called *sample path regularity* of a random field. Throughout this section, we assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\mathcal{D} \subseteq \mathbb{R}^d$ a pathwise connected domain.

4.1 Properties of random fields

We begin this chapter with the definition of a *random field*.

Definition 4.1.1 (Random field)**.**
A function $a\colon \Omega \times \mathcal{D} \to \mathbb{R}$ is called *stochastic process* or *random field* if, for every $x \in \mathcal{D}$, the function $a(\cdot, x)\colon \Omega \to \mathbb{R}$ is a real-valued random variable. In this case, we call the mapping $a(\omega, \cdot)\colon \mathcal{D} \to \mathbb{R}$ for $\omega \in \Omega$ a *realization* or *path* of a.

Remark.
From this point of view, a random field is just a family of random variables indexed over the (usually) uncountable index set $\mathcal{D}$. However, one can also interpret a random field as a *function-valued* random variable $a^*\colon \Omega \to \{f\colon \mathcal{D} \to \mathbb{R}\}$. The latter can be equipped with the so-called *cylindrical σ-algebra* (see [54, Chapter I.1.2] for the definition). Then one can prove that a^* is measurable with respect to this σ-algebra.

In this thesis, we adopt the latter point of view but work with a more restrictive class, namely *second order random fields*.

Definition and Remark 4.1.2 (Second order random field)**.**
A random field $a\colon \Omega \times \mathcal{D} \to \mathbb{R}$ is called *second order random field* if (the equivalence class of) a is an element of $L^2(\Omega; L^2(\mathcal{D}))$. In this case we can define the *mean of a* as the Bochner integral

$$m := \int_\Omega a(\omega, \cdot)\, d\mathbb{P}(\omega) \in L^2(\mathcal{D}),$$

and the *covariance operator of a* as

$$\mathcal{C}\colon L^2(\mathcal{D}) \to L^2(\mathcal{D}), \quad u \mapsto \int_\Omega \langle a(\omega, \cdot) - m, u\rangle_{L^2(\mathcal{D})} (a(\omega, \cdot) - m)\, d\mathbb{P}(\omega).$$

Moreover, a is called *centered* if $m = 0 \in L^2(\mathcal{D})$.

Remark 4.1.3.

1. For the mean and covariance operator of a second order random field, we know by Corollary 2.3.11 and $\mathbb{P}(\Omega) = 1 < \infty$ that $L^2(\Omega; L^2(\mathcal{D})) \subseteq L^1(\Omega; L^2(\mathcal{D}))$ and that the Bochner integrals are thus well-defined. In particular, this implies that the paths of a are $\mathbb{P}$-almost surely in $L^2(\mathcal{D})$.

2. If $a\colon \Omega \times \mathcal{D} \to \mathbb{R}$ is a random field in the sense of Definition 4.1.1, one has to impose additional assumptions on a to obtain a second order random field: First, one has to assume that the paths of a are $\mathbb{P}$-almost surely in $L^2(\mathcal{D})$, additionally that the random field $a\colon \Omega \times \mathcal{D} \to \mathbb{R}$ is measurable with respect to the product measure space $(\Omega \times \mathcal{D}, \mathcal{F} \otimes \mathcal{B}(\mathcal{D}), \mathbb{P} \otimes \lambda)$, where λ denotes the Lebesgue measure on $\mathcal{D}$, and finally that
$$\int_\Omega \int_\mathcal{D} |a(\omega, x)|^2 \, dx \, d\mathbb{P}(\omega) < \infty.$$
The first and second assumption allows us to apply the *Pettis measurability theorem* (see [26, Theorem 1.1.20]) to prove that the random field a, interpreted as an $L^2(\mathcal{D})$-valued random variable, is $\mathbb{P}$-measurable in the sense of Definition 2.3.2 and hence Bochner integrable. The third assumption ensures $a \in L^2(\Omega; L^2(\mathcal{D}))$.

3. In the light of Section 2.3, one can generally speak of *(separable) Banach space-valued* random variables, and a second order random field is just a special case of such a random variable using $B = L^2(\mathcal{D})$. In particular, we will speak of random fields $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ with $\sigma \geq 0$, which are then a special case of second order random fields.

By Fubini's theorem, the covariance operator of a second order random field a is given as an integral operator:

$$\begin{aligned}
\mathcal{C}(u)(y) &= \int_\Omega \int_\mathcal{D} (a(\omega, x) - m(x)) u(x) \, dx \, (a(\omega, y) - m(y)) \, d\mathbb{P}(\omega) \\
&= \int_\mathcal{D} \mathbb{E}\left[(a(\cdot, x) - m(x))(a(\cdot, y) - m(y))\right] u(x) \, dx \\
&=: \int_\mathcal{D} \mathrm{Cov}(x, y) u(x) \, dx,
\end{aligned}$$

where the corresponding kernel Cov is defined as

$$\mathrm{Cov}\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}, \quad (x, y) \mapsto \mathbb{E}\left[(a(\cdot, x) - m(x))\,(a(\cdot, y) - m(y))\right], \tag{4.1}$$

i.e., the pointwise covariance of the real-valued random variables $a(\cdot, x)$ and $a(\cdot, y)$ for $x, y \in \mathcal{D}$. It fulfills several properties:

Definition and Remark 4.1.4 (Covariance function)**.**
Let $\mathcal{D} \subseteq \mathbb{R}^d$ and $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a second order random field. Then we call Cov as defined in (4.1) the *covariance function of* a. It fulfills the following properties:

1. $\mathrm{Cov} \in L^2(\mathcal{D} \times \mathcal{D})$,

2. Cov is positive semi-definite.

Moreover, the covariance operator $\mathcal{C}$ is a positive, self-adjoint, and compact operator.

Proof. For the first property, we use Fubini's theorem and Hölder's inequality to obtain

$$\begin{aligned}
\| \operatorname{Cov} \|^2_{L^2(\mathcal{D}\times\mathcal{D})} &= \int_{\mathcal{D}} \int_{\mathcal{D}} \operatorname{Cov}(x,y)^2 \, dx \, dy \\
&= \int_\Omega \int_\Omega \int_{\mathcal{D}} (a(\omega_1,x) - m(x))(a(\omega_2,x) - m(x)) \, dx \\
&\qquad \int_{\mathcal{D}} (a(\omega_1,y) - m(y))(a(\omega_2,y) - m(y)) \, dy \, d\mathbb{P}(\omega_1) \, d\mathbb{P}(\omega_2) \\
&\leq \int_\Omega \int_\Omega \|a(\omega_1,\cdot) - m\|^2_{L^2(\mathcal{D})} \|a(\omega_2,\cdot) - m\|^2_{L^2(\mathcal{D})} \, d\mathbb{P}(\omega_1) \, d\mathbb{P}(\omega_2) \\
&= \int_\Omega \|a(\omega_1,\cdot) - m\|^2_{L^2(\mathcal{D})} \, d\mathbb{P}(\omega_1) \int_\Omega \|a(\omega_2,\cdot) - m\|^2_{L^2(\mathcal{D})} \, d\mathbb{P}(\omega_2) \\
&= \|a - m\|^4_{L^2(\Omega;L^2(\mathcal{D}))} < \infty.
\end{aligned}$$

For the positive semi-definiteness, we observe that Cov is symmetric. Furthermore, let $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ be a finite point set and $\alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{R}^N$. Then

$$\begin{aligned}
\sum_{k=1}^N \sum_{\ell=1}^N \alpha_k \alpha_\ell \operatorname{Cov}(x_k, x_\ell) &= \sum_{k=1}^N \sum_{\ell=1}^N \alpha_k \alpha_\ell \mathbb{E}[(a(\cdot,x_k) - m(x_k))(a(\cdot,x_\ell) - m(x_\ell))] \\
&= \mathbb{E}\left[\left(\sum_{k=1}^N \alpha_k (a(\cdot,x_k) - m(x_k)) \right) \left(\sum_{\ell=1}^N \alpha_l (a(\cdot,x_\ell) - m(x_\ell)) \right) \right] \\
&= \mathbb{E}\left[\left(\sum_{k=1}^N \alpha_k (a(\cdot,x_k) - m(x_k)) \right)^2 \right] \geq 0.
\end{aligned}$$

The properties of the covariance operator follow directly from Remark 3.1.13 as a Hilbert-Schmidt operator with kernel $\operatorname{Cov} \in L^2(\mathcal{D} \times \mathcal{D})$. □

As $\mathcal{C}$ is compact, positive, and self-adjoint, we can proceed as in Section 3.1.1 to conclude that there exists a sequence of non-negative eigenvalues $(\lambda_j)_{j\in\mathbb{N}} \subseteq [0, \infty[$. Moreover, we obtain (see [14, Theorem 7.2.5]):

Corollary 4.1.5.
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a second order random field with mean m, covariance operator $\mathcal{C}$ and covariance function Cov*. Then $\mathcal{C}$ possesses a decreasing sequence $(\lambda_j)_{j\in\mathbb{N}} \subseteq [0, \infty[$ of eigenvalues with corresponding eigenfunctions $(e_j)_{j\in\mathbb{N}} \subseteq L^2(\mathcal{D})$ and the following statements hold:*

(i) $\mathcal{C}$ is trace-class, i.e., $\sum_{j=1}^\infty \lambda_j = \mathbb{E}\left[\|a - m\|^2_{L^2(\mathcal{D})} \right] < \infty$;

(ii) If additionally either $\operatorname{Cov} \in C(\mathcal{D} \times \mathcal{D})$ or, for all $x \in \mathcal{D}$, the function $\operatorname{Cov}(\cdot, x)$ is continuous and bounded on $\mathcal{D}$, then

$$\operatorname{Cov}(x,y) = \sum_{j=1}^\infty \lambda_j e_j(x) e_j(y), \quad x, y \in \mathcal{D},$$

and the convergence is absolute on $\mathcal{D} \times \mathcal{D}$.

Proof. As we have seen in Section 3.1.1, the integral operator $\mathcal{C}\colon L^2(\mathcal{D}) \to L^2(\mathcal{D})$ is a Hilbert-Schmidt operator, and thus the spectral theory of self-adjoint, compact operators ensures the existence of eigenvalues $(\lambda_j)_{j\in\mathbb{N}}$ and corresponding eigenfunctions $(e_j)_{j\in\mathbb{N}}$ such that

$$\sum_{j=1}^{\infty} \lambda_j^2 < \infty.$$

Additionally, as $\mathcal{C}$ is positive, all eigenvalues λ_j are non-negative. The trace-class result given in [14, Theorem 7.2.5] applies the definition of the covariance function to obtain

$$\begin{aligned}\lambda_j = \langle \mathcal{C}e_j, e_j\rangle_{L^2(\mathcal{D})} &= \int_{\mathcal{D}}\int_{\mathcal{D}}\int_{\Omega} (a(\omega,x)-m(x))\,(a(\omega,y)-m(y))\,e_j(x)e_j(y)\,d\mathbb{P}(\omega)\,dx\,dy \\ &= \mathbb{E}\left[\langle a-m, e_j\rangle_{L^2(\mathcal{D})}\langle a-m, e_j\rangle_{L^2(\mathcal{D})}\right]\end{aligned}$$

by using Fubini's theorem. When we now consider the series of all eigenvalues, we can use dominated convergence to interchange the limit and the expectation to get

$$\sum_{j=1}^{\infty}\lambda_j = \sum_{j=1}^{\infty}\mathbb{E}\left[\langle a-m, e_j\rangle_{L^2(\mathcal{D})}^2\right] = \mathbb{E}\left[\sum_{j=1}^{\infty}\langle a-m, e_j\rangle_{L^2(\mathcal{D})}^2\right] = \mathbb{E}\left[\|a-m\|_{L^2(\mathcal{D})}^2\right] < \infty,$$

where we used that the sequence $(e_j)_{j\in\mathbb{N}}$ forms an orthonormal basis of $L^2(\mathcal{D})$. The last statement is a direct application of Theorem 3.1.14. □

One can also prove the reverse statement to characterize covariance operators.

Lemma 4.1.6.
Let $\mathcal{C}\colon L^2(\mathcal{D}) \to L^2(\mathcal{D})$ be an integral operator with symmetric, positive semi-definite kernel $C \in L^2(\mathcal{D}\times\mathcal{D})$. Then $\mathcal{C}$ is a covariance operator. More specifically, there exists a second order random field $a \in L^2(\Omega; L^2(\mathcal{D}))$ with covariance operator $\mathcal{C}$ if and only if $\mathcal{C}$ is trace-class.

Proof. A proof of this statement can be found in [54, Chapter III.2.3, Corollary]. □

A special class of second order random fields are the so-called *Gaussian processes* (cf. [1, Chapter 1.2]).

Definition 4.1.7 (Gaussian process)**.**
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a second order random field with mean $m\colon \mathcal{D}\to\mathbb{R}$ and covariance function $\mathrm{Cov} \in L^2(\mathcal{D}\times\mathcal{D})$. Then a is called a *Gaussian process* if, for every finite point set $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$, the vector

$$(a(\cdot, x_1), \dots, a(\cdot, x_N)) \sim \mathcal{N}(m_N, K_N)$$

is multivariate normally distributed with mean $m_N = (m(x_1), \dots, m(x_N))^T \in \mathbb{R}^N$ and covariance matrix $K_N = (\mathrm{Cov}(x_i, x_j))_{1\le i,j\le N} \in \mathbb{R}^{N\times N}$. We then write $a \sim \mathcal{N}(m, \mathrm{Cov})$ or equivalently $a \sim GP(m, \mathrm{Cov})$.

[27, Lemma 14.1] shows that a Gaussian process is fully determined by its mean and covariance function, and vice versa, i.e., every mean $m \in L^2(\mathcal{D})$ and positive definite function $K \in L^2(\mathcal{D}\times\mathcal{D})$ uniquely define a Gaussian process $a \sim GP(m, K)$.
In the next chapter, we use the eigenvalues and eigenfunctions of a covariance operator to construct a series expansion for the corresponding random field, the so-called *Karhunen-Loève expansion.*

4.2 Karhunen-Loève expansion

A classical approach to tackle the uncertainty for a given random field a is to expand a in a Fourier-type series. This allows us to *separate* the randomness from the computations on the spatial domain, which hopefully leads to a tractable method of approximating the random field numerically. In the literature, this is achieved by, for example, the *Karhunen-Loève expansion*. The following results can mainly be found in [48] and [1, Chapter 3.1, 3.2]. Throughout this section, we assume $\mathcal{D} \subseteq \mathbb{R}^d$ is bounded. We begin with the definition of the *Cameron-Martin space* ([48, Corollary 3.6]).

Definition 4.2.1 (Cameron-Martin space)**.**
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a second order random field. We call

$$L^2(a) := \overline{\operatorname{span}\{a(\cdot, x) \mid x \in \mathcal{D}\}}^{\|\cdot\|_{L^2(\Omega,\mathbb{P})}} \subseteq L^2(\Omega, \mathbb{P})$$

the *Cameron-Martin space of* a. As a closed subspace of $L^2(\Omega, \mathbb{P})$, it is again a Hilbert space with the induced inner product $\langle \cdot, \cdot \rangle_{L^2(\Omega,\mathbb{P})}$.

If we, as before, denote by Cov the covariance function of a second order random field $a \in L^2(\Omega; L^2(\mathbb{R}^d))$, the Cameron-Martin space $L^2(a)$ turns out to be isometrically isomorphic to $H_{\mathrm{Cov}}(\mathcal{D})$, the reproducing kernel Hilbert space of Cov (note that Cov defines a symmetric and positive semi-definite function by Definition and Remark 4.1.4 and therefore this Hilbert space exists by Theorem 3.1.6).
For convenience, we also assume for the time being that the random field is centered, i.e., $\mathbb{E}[a(\cdot, x)] = 0$ for $x \in \mathcal{D}$. The following result is given in [1, Chapter 3.1]:

Theorem 4.2.2 (Canonical isomorphism)**.**
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a centered second order random field with covariance function $\mathrm{Cov} \in L^2(\mathcal{D} \times \mathcal{D})$. Then $H_{\mathrm{Cov}}(\mathcal{D})$ and $L^2(a)$ are isometrically isomorphic, more specifically, the mapping

$$\psi \colon L^2(a) \to H_{\mathrm{Cov}}(\mathcal{D}), \quad f \mapsto \int_\Omega f(\omega) a(\omega, \cdot)\, d\mathbb{P}(\omega)$$

is linear, bounded, bijective and the unique mapping fulfilling $\psi(a(\cdot, x)) = \mathrm{Cov}(\cdot, x)$ for all $x \in \mathcal{D}$. This mapping ψ is also called the canonical isomorphism.

Proof. The proof will be divided into several steps: First, we consider the dense subspace $V := \operatorname{span}\{a(\cdot, x) \mid x \in \mathcal{D}\}$ of $L^2(a)$ and prove that ψ is well-defined and an isometry on V. Afterwards, we extend these results to all functions in $L^2(a)$. In the second step, we show that ψ is indeed bijective, while in the last step, we prove its uniqueness.
Thus, let $g \in V$ be arbitrary and write $g = \sum_{k=1}^N \alpha_k a(\cdot, x_k)$ with coefficients $\alpha_k \in \mathbb{R}$ and points $x_k \in \mathcal{D}$ for $1 \leq k \leq N$. Using the linearity of the Bochner integral, we can see that

$$\psi(g) = \sum_{k=1}^N \alpha_k \int_\Omega a(\omega, x_k) a(\omega, \cdot)\, d\mathbb{P}(\omega) = \sum_{k=1}^N \alpha_k \operatorname{Cov}(\cdot, x_k) \in H_{\mathrm{Cov}}(\mathcal{D})$$

by the definition of the covariance function and the properties of $H_{\mathrm{Cov}}(\mathcal{D})$. In particular, this shows that $\psi(a(\cdot, x)) = \mathrm{Cov}(\cdot, x)$. Moreover, we see that ψ is an isometry on V since

$$\begin{aligned}\|\psi(g)\|^2_{H_{\mathrm{Cov}}(\mathcal{D})} &= \sum_{k=1}^{N}\sum_{\ell=1}^{N} \alpha_k\alpha_\ell \langle \mathrm{Cov}(\cdot, x_k), \mathrm{Cov}(\cdot, x_\ell)\rangle_{H_{\mathrm{Cov}}(\mathcal{D})} \\ &= \sum_{k,\ell=1}^{N} \alpha_k\alpha_\ell \,\mathrm{Cov}(x_k, x_\ell) \\ &= \int_\Omega g(\omega)^2 \, d\mathbb{P}(\omega) = \|g\|^2_{L^2(\Omega,\mathbb{P})}.\end{aligned}$$

As an isometry, ψ is obviously bounded on V.
We now choose an arbitrary element $g \in L^2(a)$. Since $L^2(a)$ is the completion of V with respect to $\|\cdot\|_{L^2(\Omega,\mathbb{P})}$, there exists a sequence $(g_j)_{j\in\mathbb{N}} \subseteq V$ such that $(g_j)_{j\in\mathbb{N}}$ converges to g for $j \to \infty$ in $L^2(\Omega, \mathbb{P})$. Using the representation of ψ, we obtain that $(\psi(g_j))_{j\in\mathbb{N}}$ converges to $\psi(g)$ pointwise for every $x \in \mathcal{D}$ as

$$|\psi(g)(x) - \psi(g_j)(x)| = \left|\int_\Omega (g(\omega) - g_j(\omega))a(\omega, x)\, d\mathbb{P}(\omega)\right| \le \|g - g_j\|_{L^2(\Omega,\mathbb{P})}\|a(\cdot, x)\|_{L^2(\Omega,\mathbb{P})}.$$

This proves that $\psi(g)$ is actually an element of $H_{\mathrm{Cov}}(\mathcal{D})$, as we see by the first step that

$$\|\psi(g_j) - \psi(g_k)\|_{H_{\mathrm{Cov}}(\mathcal{D})} = \|g_j - g_k\|_{L^2(\Omega,\mathbb{P})} \to 0, \quad j, k \to \infty,$$

i.e., that $(\psi(g_j))_{j\in\mathbb{N}} \subseteq H_{\mathrm{Cov}}(\mathcal{D})$ is a Cauchy sequence. Therefore, there exists an element $h \in H_{\mathrm{Cov}}(\mathcal{D})$ such that $\psi(g_j) \to h$ for $j \to \infty$ in $H_{\mathrm{Cov}}(\mathcal{D})$ as this space is complete. Now the convergence in the reproducing kernel Hilbert space $H_{\mathrm{Cov}}(\mathcal{D})$ implies the pointwise convergence $\psi(g_j)(x) \to h(x)$ for all $x \in \mathcal{D}$ by Theorem 3.1.2, which leads to $\psi(g) = h$. Since $h \in H_{\mathrm{Cov}}(\mathcal{D})$, ψ is well-defined.
Moreover, the isometry can be extended by

$$\|\psi(g)\|_{H_{\mathrm{Cov}}(\mathcal{D})} = \lim_{j\to\infty} \|\psi(g_j)\|_{H_{\mathrm{Cov}}(\mathcal{D})} = \lim_{j\to\infty} \|g_j\|_{L^2(\Omega,\mathbb{P})} = \|g\|_{L^2(\Omega,\mathbb{P})}.$$

This in particular proves that ψ is also bounded on $L^2(a)$. This isometry naturally extends to the inner product, since, for $f, g \in L^2(a)$, we can apply the polarization formula to obtain

$$\begin{aligned}\langle \psi(f), \psi(g)\rangle_{H_{\mathrm{Cov}}(\mathcal{D})} &= \frac{1}{2}\left(\|\psi(f)\|^2_{H_{\mathrm{Cov}}(\mathcal{D})} + \|\psi(g)\|^2_{H_{\mathrm{Cov}}(\mathcal{D})} - \|\psi(f-g)\|^2_{H_{\mathrm{Cov}}(\mathcal{D})}\right) \\ &= \frac{1}{2}\left(\|f\|^2_{L^2(\Omega,\mathbb{P})} + \|g\|^2_{L^2(\Omega,\mathbb{P})} - \|f-g\|^2_{L^2(\Omega,\mathbb{P})}\right) \\ &= \langle f, g\rangle_{L^2(\Omega,\mathbb{P})}.\end{aligned} \tag{4.2}$$

In the second step we want to show that ψ is bijective. As ψ is an isometry, we directly infer that it also is injective, as

$$\|\psi(f) - \psi(g)\|_{H_{\mathrm{Cov}}(\mathcal{D})} = \|f - g\|_{L^2(\Omega,\mathbb{P})},$$

for all $f, g \in L^2(a)$, holds.

For the surjectivity we argue similarly to the first step: Let $h \in H_{\mathrm{Cov}}(\mathcal{D})$ be arbitrary. Then by Theorem 3.1.6 there exists a sequence $(h_j)_{j\in\mathbb{N}} \subseteq \operatorname{span}\{\mathrm{Cov}(\cdot, x) \mid x \in \mathcal{D}\}$ which converges to h for $j \to \infty$ in $H_{\mathrm{Cov}}(\mathcal{D})$. As in the first step, for every h_j there exists a $g_j \in L^2(a)$ such that $\psi(g_j) = h_j$. This time we can show that $(g_j)_{j\in\mathbb{N}} \subseteq L^2(a)$ is a Cauchy sequence, since

$$\|g_j - g_k\|_{L^2(\Omega,\mathbb{P})} = \|\psi(g_j) - \psi(g_k)\|_{H_{\mathrm{Cov}}(\mathcal{D})} = \|h_j - h_k\|_{H_{\mathrm{Cov}}(\mathcal{D})},$$

where the right-hand side tends to zero for $j, k \to \infty$. As $L^2(a)$ is complete as well, there exists an element $g \in L^2(a)$ such that $(g_j)_{j\in\mathbb{N}}$ converges to g in $L^2(\Omega, \mathbb{P})$. To prove that $h = \psi(g)$, we use triangle inequality to obtain

$$\begin{aligned} 0 \leq \|h - \psi(g)\|_{H_{\mathrm{Cov}}(\mathcal{D})} &\leq \|h - h_j\|_{H_{\mathrm{Cov}}(\mathcal{D})} + \|\psi(g_j) - \psi(g)\|_{H_{\mathrm{Cov}}(\mathcal{D})} \\ &= \|h - h_j\|_{H_{\mathrm{Cov}}(\mathcal{D})} + \|g_j - g\|_{L^2(\Omega,\mathbb{P})}. \end{aligned}$$

Again, the right-hand side tends to zero for $j \to \infty$, which proves that ψ is also surjective and therefore bijective.
Finally, we want to show the uniqueness of this mapping. Assume there is another mapping $\varphi\colon L^2(a) \to H_{\mathrm{Cov}}(\mathcal{D})$ with the same properties as ψ and let $g \in L^2(a)$ be arbitrary. Then there exists a sequence $(g_j)_{j\in\mathbb{N}} \subseteq V$ that converges to g in $L^2(\Omega, \mathbb{P})$. By using $\psi(a(\cdot, x)) = \mathrm{Cov}(\cdot, x) = \varphi(a(\cdot, x))$, one can easily check that, for every $j \in \mathbb{N}$, the equality $\psi(g_j) = \varphi(g_j)$ and thus

$$0 \leq \|\psi(g) - \varphi(g)\|_{H_{\mathrm{Cov}}(\mathcal{D})} \leq \|\psi(g) - \psi(g_j)\|_{H_{\mathrm{Cov}}(\mathcal{D})} + \|\varphi(g) - \varphi(g_j)\|_{H_{\mathrm{Cov}}(\mathcal{D})}$$

holds, which proves that $\varphi(g) = \psi(g)$ holds, as the right-hand side tends to zero for $j \to \infty$. Therefore, the mapping is unique. □

Now, we can use the canonical isomorphism to construct the Karhunen-Loève expansion of a random field $a \in L^2(\Omega; L^2(\mathcal{D}))$. The idea of this expansion is to choose an orthonormal basis of $H_{\mathrm{Cov}}(\mathcal{D})$ and then derive the Fourier expansion using this orthonormal basis (see [1, Theorem 3.1.1]):

Lemma 4.2.3.
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a centered second order random field with covariance function $\mathrm{Cov} \in L^2(\mathcal{D} \times \mathcal{D})$ and $(\varphi_j)_{j\in I}$ an orthonormal basis of $H_{\mathrm{Cov}}(\mathcal{D})$, where either $I = \mathbb{N}$ or $I = \{1, \ldots, N\}$, with $N \in \mathbb{N}$, denotes a countable or finite index set, depending on the dimension of $H_{\mathrm{Cov}}(\mathcal{D})$. Then there exists an orthonormal basis $(\xi_j)_{j\in I} \subseteq L^2(a)$ such that

$$a(\cdot, x) = \sum_{j\in I} \xi_j(\cdot)\varphi_j(x),$$

where the series converges in $L^2(\Omega, \mathbb{P})$ for almost all $x \in \mathcal{D}$.

Proof. Using the canonical isomorphism ψ from Theorem 4.2.2, we set the random variables $\xi_j := \psi^{-1}(\varphi_j) \in L^2(a)$ for $j \in I$. This sequence $(\xi_j)_{j\in I}$ forms an orthonormal basis of $L^2(a)$. The orthogonality follows from the orthogonality of $(\varphi_j)_{j\in I}$ and (4.2) by

$$\langle \xi_j, \xi_k \rangle_{L^2(\Omega,\mathbb{P})} = \langle \psi^{-1}(\varphi_j), \psi^{-1}(\varphi_k) \rangle_{L^2(\Omega,\mathbb{P})} = \langle \varphi_j, \varphi_k \rangle_{H_{\mathrm{Cov}}(\mathcal{D})} = \delta_{j,k}. \tag{4.3}$$

Furthermore, if $f \in L^2(a)$ is orthogonal to ξ_j for all $j \in I$, the canonical isomorphism again implies that

$$0 = \langle f, \xi_j \rangle_{L^2(\Omega,\mathbb{P})} = \langle \psi^{-1}(f), \psi^{-1}(\xi_j) \rangle_{H_{\mathrm{Cov}(\mathcal{D})}} = \langle \psi^{-1}(f), \varphi_j \rangle_{H_{\mathrm{Cov}(\mathcal{D})}}$$

for all $j \in I$, which leads to $\psi^{-1}(f) = 0$ and by bijectivity of ψ also to $f = 0$. Thus, $(\xi_j)_{j\in I}$ forms an orthonormal basis of $L^2(a)$.
With Fubini's theorem, we see that

$$\int_{\mathcal{D}} \int_{\Omega} a^2(\omega, x) \, d\mathbb{P}(\omega) \, dx < \infty,$$

which implies that there is a Lebesgue set M of measure zero such that $a(\cdot, x) \in L^2(\Omega, \mathbb{P})$ for $x \in \mathcal{D} \setminus M$. Now, for $x \in \mathcal{D} \setminus M$, we can express $a(\cdot, x)$ with its Fourier series in $L^2(a)$ using $(\xi_j)_{j\in I}$ and thus obtain

$$a(\cdot, x) = \sum_{j \in I} \mathbb{E}[a(\cdot, x)\xi_j]\xi_j,$$

where the series converges in $L^2(\Omega, \mathbb{P})$. Last, we again use the canonical isomorphism ψ to rewrite

$$\mathbb{E}[a(\cdot, x)\xi_j] = \langle a(\cdot, x), \psi^{-1}(\varphi_j) \rangle_{L^2(\Omega,\mathbb{P})} = \langle \mathrm{Cov}(\cdot, x), \varphi_j \rangle_{H_{\mathrm{Cov}(\mathcal{D})}} = \varphi_j(x)$$

for all $j \in I$, which provides the desired result. □

In order to apply this lemma reasonably, we have to construct an orthonormal basis of $H_{\mathrm{Cov}}(\mathcal{D})$, which, by Section 3.1.1, can for example be obtained as the normalized eigenfunctions of the covariance operator $\mathcal{C}$ of a. With this specific basis, we can prove the Karhunen-Loève theorem ([3, Theorem 5.3]):

Theorem 4.2.4 (Karhunen-Loève expansion)**.**
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a centered second order random field with covariance function $\mathrm{Cov} \in L^2(\mathcal{D} \times \mathcal{D})$ and covariance operator $\mathcal{C}$. Further assume that one of the two following conditions holds:

- Cov *is continuous on $\mathcal{D} \times \mathcal{D}$;*
- *For all $x \in \mathcal{D}$, the function $\mathrm{Cov}(\cdot, x)$ is continuous and bounded on $\mathcal{D}$.*

Denote by $(\lambda_j, e_j)_{j\in I} \subseteq]0, \infty[\times L^2(\mathcal{D})$ the finite or countably many eigenpairs of $\mathcal{C}$ with decreasing eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots > 0$. Then the Karhunen-Loève expansion of a is given by

$$a(\cdot, x) = \sum_{j \in I} \sqrt{\lambda_j} \xi_j(\cdot) e_j(x), \quad x \in \mathcal{D},$$

where the series converges in $L^2(\Omega, \mathbb{P})$. The random variables $(\xi_j)_{j\in I}$ are uncorrelated, have mean zero and are given by

$$\xi_j(\omega) = \frac{1}{\sqrt{\lambda_j}} \int_{\mathcal{D}} a(\omega, x) e_j(x) \, dx, \quad j \in I.$$

Proof. Since $a \in L^2(\Omega; L^2(\mathcal{D}))$, one can see that

$$\int_{\mathcal{D}} \operatorname{Cov}(x,x)\,dx = \int_{\mathcal{D}} \int_{\Omega} a(\omega,x)^2\,d\mathbb{P}(\omega)\,dx = \|a\|^2_{L^2(\Omega;L^2(\mathcal{D}))} < \infty.$$

Thus, Lemma 3.1.15 states that $H_{\operatorname{Cov}}(\mathcal{D})$ is compactly embedded in $L^2(\mathcal{D})$. Therefore, we can apply Theorem 3.1.14 to see that $(\sqrt{\lambda_j}e_j)_{j\in I}$ is an orthonormal basis of $H_{\operatorname{Cov}}(\mathcal{D})$. Using this specific basis in Lemma 4.2.3 yields the series

$$a(\cdot,x) = \sum_{j\in I} \sqrt{\lambda_j}\widetilde{\xi}_j(\cdot)e_j(x),$$

which converges in $L^2(\Omega,\mathbb{P})$ for almost all $x \in \mathcal{D}$. The random variables $\widetilde{\xi}_j$ coincide with the ones stated in the result, since

$$\begin{aligned}\widetilde{\xi}_j = \frac{1}{\sqrt{\lambda_j}}\psi^{-1}(\lambda_j e_j) &= \frac{1}{\sqrt{\lambda_j}}\psi^{-1}\left(\int_{\mathcal{D}} \operatorname{Cov}(\cdot,y)e_j(y)\,dy\right)\\ &= \frac{1}{\sqrt{\lambda_j}}\int_{\mathcal{D}} e_j(y)\psi^{-1}(\operatorname{Cov}(\cdot,y))\,dy\\ &= \frac{1}{\sqrt{\lambda_j}}\int_{\mathcal{D}} e_j(y)a(\cdot,y)\,dy.\end{aligned}$$

We additionally see with Fubini's theorem that these random variables are centered since a is centered:

$$\mathbb{E}[\xi_j] = \frac{1}{\sqrt{\lambda_j}}\int_{\Omega}\int_{\mathcal{D}} a(\omega,x)e_j(x)\,dx\,d\mathbb{P}(\omega) = \frac{1}{\sqrt{\lambda_j}}\int_{\mathcal{D}} \mathbb{E}[a(\cdot,x)]e_j(x)\,dx = 0.$$

This implies, using (4.3) with $\mathbb{E}[\xi_j\xi_i] = \delta_{i,j}$, that the ξ_j are indeed uncorrelated.
In order to get the convergence of the series in $L^2(\Omega,\mathbb{P})$ for all $x \in \mathcal{D}$, we observe that

$$\begin{aligned}&\|a(\cdot,x) - \sum_{j=1}^{n}\sqrt{\lambda_j}\xi_j(\cdot)e_j(x)\|^2_{L^2(\Omega,\mathbb{P})}\\ &= \|a(\cdot,x)\|^2_{L^2(\Omega,\mathbb{P})} - 2\sum_{j=1}^{n}\sqrt{\lambda_j}e_j(x)\langle a(\cdot,x),\xi_j\rangle_{L^2(\Omega,\mathbb{P})}\\ &\qquad + \sum_{j,k=1}^{n}\sqrt{\lambda_j}\sqrt{\lambda_k}e_j(x)e_k(x)\langle\xi_j,\xi_k\rangle_{L^2(\Omega,\mathbb{P})}\\ &= \operatorname{Cov}(x,x) - \sum_{j=1}^{n}\lambda_j e_j(x)e_j(x).\end{aligned}$$

For a fixed $x \in \mathcal{D}$, the latter expression converges to zero by Theorem 3.1.14 (choosing $y := x$), and the claim follows. □

To extend this result to non-centered random fields $a \in L^2(\Omega; L^2(\mathcal{D}))$ with mean m, we apply the previous theorem to the centered process $\widetilde{a} := a - m$.

Corollary 4.2.5.
We consider the setting of Theorem 4.2.4, but drop the assumption that $a \in L^2(\Omega; L^2(\mathcal{D}))$ is centered and instead assume that it has a pointwise defined mean $m \in L^2(\mathcal{D})$.
Denote with $(e_j, \lambda_j)_{j\in I} \subseteq L^2(\mathcal{D})\times]0,\infty[$ the finite or countably many eigenvalues and corresponding eigenfunctions of the covariance operator $\mathcal{C}$ of the random field a, where the eigenvalues are sorted in decreasing order $\lambda_1 \geq \lambda_2 \geq \cdots > 0$. Then the Karhunen-Loève expansion of a is given by

$$a(\cdot, x) = m(x) + \sum_{j\in I} \sqrt{\lambda_j}\xi_j(\cdot)e_j(x), \quad x \in \mathcal{D},$$

where, for all $x \in \mathcal{D}$, the series converges in $L^2(\Omega, \mathbb{P})$. The random variables ξ_j for $j \in I$ are given by

$$\xi_j(\omega) = \frac{1}{\sqrt{\lambda_j}} \int_{\mathcal{D}} (a(\omega, x) - m(x))\, e_j(x)\, dx.$$

Proof. We set $\widetilde{a}(\omega, x) := a(\omega, x) - m(x)$ for $\omega \in \Omega$ and $x \in \mathcal{D}$. Then $\widetilde{a}$ is a centered second order random field with covariance function given by

$$\begin{aligned}\widetilde{\mathrm{Cov}}(x, y) &= \mathbb{E}\left[\widetilde{a}(\cdot, x)\widetilde{a}(\cdot, y)\right] \\ &= \mathbb{E}\left[(a(\cdot, x) - m(x))(a(\cdot, y) - m(y))\right] \\ &= \mathrm{Cov}(x, y)\end{aligned}$$

for every $x, y \in \mathcal{D}$. Thus, we can apply Theorem 4.2.4 on $\widetilde{a}$ and add m on both sides to prove the statement. □

While the Karhunen-Loève expansion can be proven for all second order random fields, there is generally no constructive method to determine the distribution of the real-valued random variables ξ_j. A remarkable exception to this issue are Gaussian random fields.

Corollary 4.2.6.
In the setting of Corollary 4.2.5, assume that $a \in L^2(\Omega; L^2(\mathcal{D}))$ is a Gaussian random field. Then the random variables $(\xi_j)_{j\in I}$ in the Karhunen-Loève expansion of the random field a are independent and normally distributed, i.e., $\xi_j \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$ for all $j \in I$, and the Karhunen-Loève expansion is again given by Corollary 4.2.5.

Proof. We know by construction that $\xi_j \in L^2(a)$ for all $j \in I$. Since a is assumed to be Gaussian, one can conclude that all elements of $L^2(a)$ are again Gaussian (since the limit of Gaussian random variables is again Gaussian, see [27, Chapter 6, Exercise 2]). In particular, $\xi_j = \psi^{-1}(e_j) \in L^2(a)$ for all $j \in \mathbb{N}$, thus these are Gaussian random variables. As $(\xi_j)_{j\in I}$ are uncorrelated Gaussian random variables, they are also independent (cf. [8, section 29]). □

As we have seen in the proof of Theorem 4.2.4, the *truncated* Karhunen-Loève expansion $a^{(n)}(\cdot, x) = \sum_{j=1}^{n} \sqrt{\lambda_j}\xi_j e_j(x)$ converges uniformly to a for every $x \in \mathcal{D}$ as a series in $L^2(\Omega, \mathbb{P})$. Moreover, the truncation error can be calculated explicitly (see [48, Theorem 3.5]).

Corollary 4.2.7.
Assume the setting of Corollary 4.2.5 holds, where a random field $a \in L^2(\Omega; L^2(\mathcal{D}))$ is given by

$$a(\cdot, x) = m(x) + \sum_{j \in I} \sqrt{\lambda_j} \xi_j e_j(x)$$

for $x \in \mathcal{D}$, where the convergence is in $L^2(\Omega, \mathbb{P})$, with either $I = \mathbb{N}$ or $I = \{1, \dots, N\}$ for a natural number $N \in \mathbb{N}$. Let $n \in \mathbb{N}$ be fixed and define the truncated Karhunen-Loève expansion *as*

$$a^{(n)}(\cdot, x) := m(x) + \sum_{j=1}^{n} \sqrt{\lambda_j} \xi_j e_j(x), \quad x \in \mathcal{D},$$

where additionally $n \leq N$ in the case where I is finite. Then the truncation error can be expressed as

$$\|a - a^{(n)}\|^2_{L^2(\Omega; L^2(\mathcal{D}))} = \sum_{j \in I \setminus \{1, \dots, n\}} \lambda_j.$$

In particular, this shows that the error reduction using a truncated Karhunen-Loève expansion can be controlled by analyzing the eigenvalue decay of the covariance operator.

4.2.1 Eigenvalue decay

As we have already seen in Definition and Remark 4.1.4, the covariance operator $\mathcal{C}$ of a second order random field $a \in L^2(\Omega; L^2(\mathcal{D}))$ is given as an integral operator, so we can apply the general theory of integral operators to investigate the decay rates of the associated eigenvalues.
These results mainly rely on Weyl's theorem ([31, Theorem 1.b.9]), which can be exploited, as, for example, done in [31, Theorem 3.d.5], to conclude that the sequence of eigenvalues for Hilbert-Schmidt operators lie in specific Lorentz spaces, from which a decay rate can be inferred. Here, we present a more recent and sharper estimate for the eigenvalue decay presented in [22, Theorem 3.2], which is specifically tailored to random fields and covariance functions.

Theorem 4.2.8.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary. Furthermore, let a second order random field $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ with $\sigma > \frac{d}{2}$ be given. As before, we denote by $\mathcal{C} \colon L^2(\mathcal{D}) \to L^2(\mathcal{D})$ the associated covariance operator of the second order random field $a \in L^2(\Omega; L^2(\mathcal{D}))$ and by $(\lambda_n)_{n \in \mathbb{N}}$ the decreasingly sorted eigenvalues of the operator. Then these eigenvalues satisfy the bound

$$\lambda_n \leq C \|a\|^2_{L^2(\Omega; L^2(\mathcal{D}))} n^{-1 - \frac{2\sigma}{d}}, \quad n \in \mathbb{N},$$

where $C > 0$ only depends on $\mathcal{D}, d$ and σ.

Remark.
Note that although it is assumed that Ω is bounded in the cited result, it remains valid when choosing an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$, as already assumed in [22, Chapter 4]. Additionally, the result is formulated for centered random fields $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$, but one can easily extend it to non-centered random fields $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ by simply applying the results to $a - m \in L^2(\Omega; H^\sigma(\mathcal{D}))$, where $m \in H^\sigma(\mathcal{D})$ denotes the mean of a.

This eigenvalue decay can be combined with Corollary 4.2.7 to estimate a viable cutoff index for the Karhunen-Loève expansion of a random field $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$. However, the result relies on knowing the regularity of the paths of a. As we will see in the next chapter, this regularity can be inferred by regularity assumptions on the covariance function Cov of a.

4.3 Sample path regularity

Before we can apply the kernel-based approximation methods stated in Chapter 3 to a random field, we have to ensure that its paths fulfill certain regularity assumptions. In the literature, there is already well-known theory (see for example [35, Chapter 5.5]) that uses the covariance function of this random field to infer regularity of the paths.

Definition 4.3.1.
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a second order stochastic process with mean m. Then a is called *mean-square continuous* if, for all $x \in \mathcal{D}$,

$$\lim_{y \to x} \mathbb{E}\left[(a(\cdot, y) - a(\cdot, x))^2\right] = 0.$$

The following theorem stated in [35, Theorem 5.32] expresses mean-square continuity of a random field using its covariance function:

Lemma 4.3.2.
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a second order random field with mean $m \in C(\mathcal{D})$ and covariance function Cov*. Then a is mean-square continuous if and only if the covariance function* $\mathrm{Cov} \in C(\mathcal{D} \times \mathcal{D})$.

Unfortunately, this result is not enough for our application, as it does not imply that the paths of the random field are continuous almost surely. This can, for example, be established under additional assumptions using the *Kolmogorov continuity theorem* (cf. [35, Theorem 5.40]). In our scenario we are more interested in weak differentiability, in particular, in random fields which are elements of $L^2(\Omega; H^\sigma(\mathcal{D}))$ for $\sigma > \frac{d}{2}$. At this point, it is tempting to assume that the paths of the random field a are elements of the reproducing kernel Hilbert space $H_{\mathrm{Cov}}(\mathcal{D})$ associated to the covariance function of this process. If then for example $a \sim GP(m, \Phi)$ is a Gaussian random field where Φ is a reproducing kernel with Hilbert space equivalent to $H^\sigma(\mathcal{D})$, then the paths of a would also be in this Hilbert space. Unfortunately, as for example proven in [36, Corollary 7.1], this statement does not hold in general.

Theorem 4.3.3.
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a Gaussian process with covariance function $\mathrm{Cov} \in L^2(\mathcal{D} \times \mathcal{D})$ *and mean $m \in H_{\mathrm{Cov}}(\mathcal{D})$. If $H_{\mathrm{Cov}}(\mathcal{D})$ is infinite dimensional, then*

$$\mathbb{P}\left(\{\omega \in \Omega \mid a(\omega, \cdot) \in H_{\mathrm{Cov}}(\mathcal{D})\}\right) = 0,$$

i.e., the paths of the random field a are $\mathbb{P}$-almost surely not contained in $H_{\mathrm{Cov}}(\mathcal{D})$.

While there are also results given in [48, Corollary 4.5] or [36] which make statements on the paths of random fields, we use the following result proven in [46, Theorem 1], which provides a connection between the differentiability of the covariance function of a centered random field to weak differentiability of its sample paths.

Theorem 4.3.4.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain and $a \in L^2(\Omega; L^2(\mathcal{D}))$ a centered random field with covariance function $\operatorname{Cov}\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$. Furthermore, let $k \in \mathbb{N}$ such that, for all multi-indices $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq k$, the derivatives $D_x^\alpha D_y^\alpha \operatorname{Cov}(x, y)$ exist and are continuous on the diagonal $\mathfrak{D} := \{(x, x) \mid x \in \mathcal{D}\}$ in the classical sense. Then the paths of the random field a are $\mathbb{P}$-almost surely in $H^k(\mathcal{D})$.

In [46, Theorem 2], a sufficient condition for paths in Sobolev spaces of fractional orders is also proven. However, we use the techniques applied in this theorem combined with the embedding of Theorem 2.4.16 to link weak differentiability of the covariance function to the weak differentiability of the paths of a in a slightly more general way.

Theorem 4.3.5.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and $a \in L^2(\Omega; L^2(\mathcal{D}))$ a centered random field.
Further assume that the covariance function Cov is an element of $H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D} \times \mathcal{D})$ for $\sigma > \frac{d}{2}$. Then the paths of a are $\mathbb{P}$-almost surely contained in $H^\tau(\mathcal{D})$ for $0 \leq \tau < \sigma - \frac{d}{2}$. Moreover, $a \in L^2(\Omega; H^\tau(\mathcal{D}))$.

Proof. As already mentioned, the proof relies on ideas given in [46, Theorem 2]. To this end, let $0 \leq \tau < \sigma - \frac{d}{2}$ be arbitrary and write $\tau = k + \gamma$ with $k \in \mathbb{N}_0$ and $0 \leq \gamma < 1$. By using Theorem 2.4.25 and Theorem 4.3.4, we can see that the paths of a are $\mathbb{P}$-almost surely contained in $H^k(\mathcal{D})$. Hence, we can focus on the case $\gamma > 0$. There, we have to prove that $|a(\omega, \cdot)|_{H^\sigma(\mathcal{D})}$ is almost surely finite, which, by taking the expectation, is equivalent to the statement that

$$\mathbb{E}\left[|a(\cdot,\cdot)|^2_{H^\sigma(\mathcal{D})}\right] = \mathbb{E}\left[\int_{\mathcal{D}}\int_{\mathcal{D}} \frac{\left(D^\alpha a(\cdot, x) - D^\alpha a(\cdot, y)\right)^2}{\|x - y\|_2^{d+2\gamma}}\, dx\, dy\right] < \infty$$

for all $\alpha \in \mathbb{N}_0^d$ with $|\alpha| = k$. By applying Fubini's theorem, we can interchange the expectation with the inner integrals, which leads to

$$\begin{aligned}
&\int_{\mathcal{D}}\int_{\mathcal{D}} \frac{\mathbb{E}\left[\left(D^\alpha a(\cdot, x) - D^\alpha a(\cdot, y)\right)^2\right]}{\|x - y\|_2^{d+2\gamma}}\, dy\, dx \\
&= \int_{\mathcal{D}}\int_{\mathcal{D}} \frac{D^{\alpha,\alpha} \operatorname{Cov}(x, x) + D^{\alpha,\alpha} \operatorname{Cov}(y, y) - 2D^{\alpha,\alpha} \operatorname{Cov}(x, y)}{\|x - y\|_2^{d+2\gamma}}\, dy\, dx.
\end{aligned}$$

Now, choose $\varepsilon > 0$ such that $\widetilde{\gamma} := \gamma + \varepsilon < 1$ and additionally $\widetilde{\tau} := k + \widetilde{\gamma} < \sigma - \frac{d}{2}$ holds. Then Theorem 2.4.25 implies that the embedding

$$\iota\colon H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D} \times \mathcal{D}) \to C^{\widetilde{\tau},\widetilde{\tau}}_{\mathrm{mix}}(\mathcal{D} \times \mathcal{D}), \quad f \mapsto f$$

is linear and bounded. By Theorem 2.4.16, we therefore conclude that there is a constant $C > 0$ such that

$$\max_{|\alpha|=k} \sup_{\substack{x,y\in\mathcal{D}\\ x\neq y}} \frac{|D^{\alpha,\alpha}\operatorname{Cov}(x,x) + D^{\alpha,\alpha}\operatorname{Cov}(y,y) - 2D^{\alpha,\alpha}\operatorname{Cov}(x,y)|}{\|x-y\|^{2(\gamma+\varepsilon)}} \leq C,$$

which is equivalent to

$$\max_{|\alpha|=k} |D^{\alpha,\alpha}\operatorname{Cov}(x,x) + D^{\alpha,\alpha}\operatorname{Cov}(y,y) - 2D^{\alpha,\alpha}\operatorname{Cov}(x,y)| \leq C\|x-y\|_2^{2\gamma+2\varepsilon}$$

for all $x, y \in \mathcal{D}$ with $x \neq y$. Using this bound yields

$$\begin{aligned}
&\int_{\mathcal{D}}\int_{\mathcal{D}} \frac{D^{\alpha,\alpha}\operatorname{Cov}(x,x) + D^{\alpha,\alpha}\operatorname{Cov}(y,y) - 2D^{\alpha,\alpha}\operatorname{Cov}(x,y)}{\|x-y\|_2^{d+2\gamma}}\,dy\,dx \\
&\leq C\int_{\mathcal{D}}\int_{\mathcal{D}} \|x-y\|_2^{-d+2\varepsilon}\,dy\,dx \\
&= C\int_{\mathcal{D}}\int_{\mathcal{D}-y} \|x\|_2^{-d+2\varepsilon}\,dx\,dy,
\end{aligned}$$

where we used the transformation formula to rewrite the inner integral using the shifted domain $\mathcal{D} - y := \{x - y \mid x \in \mathcal{D}\}$ for $y \in \mathcal{D}$. As we assumed that $\mathcal{D} \subseteq \mathbb{R}^d$ is bounded, there exists a radius $R > 0$ such that $\mathcal{D} \subseteq B_R(0)$. Therefore, $\mathcal{D} - y \subseteq B_{2R}(0)$ for all $y \in \mathcal{D}$ using triangle inequality. This allows us to further bound the expression by

$$\begin{aligned}
&\int_{\mathcal{D}}\int_{\mathcal{D}-y} \|x\|_2^{-d+2\varepsilon}\,dx\,dy \\
&\leq \operatorname{vol}(\mathcal{D})\int_{B_{2R}(0)} \|x\|_2^{-d+2\varepsilon}\,dx.
\end{aligned}$$

Now we apply the polar coordinates transformation, which shows that the latter integral is finite, as

$$\int_{B_{2R}(0)} \|x\|_2^{-d+2\varepsilon}\,dx = \operatorname{vol}(B_1(0))\int_0^{2R} r^{d-1-d+2\varepsilon}\,dr = vol(B_1(0))\int_0^{2R} r^{-1+2\varepsilon}\,dr < \infty.$$

This ultimately yields that the sample paths of a are almost surely in $H^\tau(\mathcal{D})$.
It remains to prove that $a \in L^2(\Omega; H^\tau(\mathcal{D}))$. First, we see that the Borel σ-algebra of $(H^\tau(\mathcal{D}), \|\cdot\|_{H^\tau(\mathcal{D})})$ is a sub-σ-algebra of the Borel σ-algebra of $(L^2(\mathcal{D}), \|\cdot\|_{L^2(\mathcal{D})})$, which follows by the *Lusin-Souslin theorem* proven in [29, Theorem 15.1]. Therefore, the random field a is also measurable with respect to the Borel σ-algebra $\mathcal{B}(H^\tau(\mathcal{D}))$. Using Proposition 2.3.4 then implies that it is $\mathbb{P}$-measurable with respect to $(H^\tau(\mathcal{D}), \|\cdot\|_{H^\tau(\mathcal{D})})$.
Next, we have to show that the following integral is finite:

$$\int_\Omega \|a(\omega,\cdot)\|_{H^\tau(\mathcal{D})}^2\,d\mathbb{P}(\omega) = \int_\Omega \|a(\omega,\cdot)\|_{H^k(\mathcal{D})}^2\,d\mathbb{P}(\omega) + \int_\Omega |a(\omega,\cdot)|_{H^\tau(\mathcal{D})}^2\,d\mathbb{P}(\omega).$$

The finiteness of the latter integral has already been shown, while for the first one we choose $\alpha \in \mathbb{N}_0$ with $|\alpha| \leq k$ and apply Fubini's theorem to obtain

$$\begin{aligned}
\int_\Omega \|D^\alpha a(\omega,\cdot)\|_{L^2(\mathcal{D})}^2\,d\mathbb{P}(\omega) &= \int_\Omega\int_{\mathcal{D}} D^\alpha a(\omega,x) D^\alpha a(\omega,x)\,dx\,d\mathbb{P}(\omega) \\
&= \int_{\mathcal{D}} D^{\alpha,\alpha}\operatorname{Cov}(x,x)\,dx < \infty,
\end{aligned}$$

as $\mathcal{D}$ is bounded and $D^{\alpha,\alpha}\operatorname{Cov} \in C_b(\mathcal{D}\times\mathcal{D})$. This proves the claim. □

This result now relates weak differentiability of a covariance function to weak differentiability of a centered random field. If we want to use this theory for general random fields $a \in L^2(\Omega; L^2(\mathcal{D}))$ with mean $m \neq 0$, we have to assume that the mean has the desired regularity, as we cannot enforce this by assumptions on the covariance function: Consider for example a random process that is actually a deterministic function, i.e., $a(\omega, x) = m(x)$ for all $x \in \mathcal{D}$ and $\omega \in \Omega$. Then the covariance function of a is just given by $\mathrm{Cov}(x, y) = 0$, which is of course infinitely often differentiable, but the smoothness of the sample paths is fully determined by the regularity of m. This behavior can be formalized as follows:

Corollary 4.3.6.
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{D} \subseteq \mathbb{R}^d$. Let $V \subseteq L^2(\mathcal{D})$ be a separable Banach space and $a \in L^2(\Omega; L^2(\mathcal{D}))$ a second order random field. Then the following two statements are equivalent:

1. *$a \in L^2(\Omega; V)$,*
2. *$(a - m) \in L^2(\Omega; V)$ and $m \in V$.*

Proof. If $a \in L^2(\Omega; V)$, the mean $m = \int_\Omega a(\omega, \cdot)\, d\mathbb{P}(\omega)$ is an element of V by the properties of the Bochner integral as well. Hence, we also have $a - m \in L^2(\Omega; V)$, as $\mathbb{P}$ is a probability measure. On the other hand, if $(a - m) \in L^2(\Omega; V)$ and $m \in V$, we have

$$\begin{aligned}\int_\Omega \|a(\omega, \cdot)\|_V^2 \, d\mathbb{P}(\omega) &\leq \int_\Omega \left(\|a(\omega, \cdot) - m\|_V + \|m\|_V\right)^2 \, d\mathbb{P}(\omega) \\ &\leq 2\|a - m\|^2_{L^2(\Omega;V)} + 2\|m\|_V^2 < \infty.\end{aligned}$$

Hence, the random field a is an element in $L^2(\Omega; V)$. □

5 Approximation of random fields

In this chapter, we use the concept of kernel-based approximation introduced earlier to reconstruct random fields using only information on a finite number of sampling points. It is organized as follows: In Section 5.1, we prove a general result which allows us to apply approximation or other methods pathwise on a given random field. In Section 5.2, we then apply kernel-based approximation in this way and prove error results in different settings.

5.1 Pathwise approximation

To illustrate the idea we investigate in the following, assume an underlying, but generally unknown, random field $a\colon \Omega \times \mathcal{D} \to \mathbb{R}$ is given. However, similar to the deterministic case, we only assume to know the real-valued random variables $a(\cdot, x_j)$ for sampling points $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$. We now seek a way to reconstruct the random field a using only this pointwise information. Since we usually do not have access to the random variables themselves, but only to realizations of them, we can interpret every realization $a(\omega, X) \in \mathbb{R}^N$ for $\omega \in \Omega$ as a data point set of the *path* $a(\omega, \cdot)\colon \mathcal{D} \to \mathbb{R}$ of a. This path itself is a deterministic function. Hence, we can apply reconstruction operators, such as the kernel-based ones presented Chapter 3, to this specific path of a. When we proceed in this manner for every realization of $a(\cdot, X)$, the question arises whether we again obtain a (second order) random field. Additionally, if this is the case, we would also like to have representations for the mean and covariance function of this new random field. To answer this question, we use the following two lemmata.

Lemma 5.1.1.
Let $U \subseteq L^2(\mathcal{D})$ be a separable Banach space endowed with the norm $\|\cdot\|_U$. Then, for every second order random field $a \in L^2(\Omega; U)$, the mean m is an element of U and the covariance function Cov *is an element of $U \overline{\otimes} U$, where the completion is with respect to a cross norm on $U \overline{\otimes} U$.*

Proof. It is clear by the definition of the Bochner integral that $m \in U$. For the second statement, we observe that the covariance function Cov of a can be interpreted as the Bochner integral of the function

$$c\colon \Omega \to U \otimes U, \quad \omega \mapsto (a(\omega, \cdot) - m) \otimes (a(\omega, \cdot) - m).$$

As a is $\mathbb{P}$-measurable, this can be extended to the mapping c using the cross norm property. Moreover, it is Bochner integrable by

$$\int_\Omega \|(a(\omega, \cdot) - m) \otimes (a(\omega, \cdot) - m)\|_{U \overline{\otimes} U} \, d\mathbb{P}(\omega) \leq \int_\Omega \|a(\omega, \cdot) - m\|_U^2 \, d\mathbb{P}(\omega) < \infty,$$

thus it is Bochner integrable as a $U \overline{\otimes} U$ valued function and the claim follows. □

The previous lemma can be used in the following general statement, which allows us to apply the desired approximation methods pathwise:

Lemma 5.1.2.
Let $U \subseteq L^2(\mathcal{D})$, endowed with the norm $\|\cdot\|_U$, and $V \subseteq L^2(\mathcal{T})$, endowed with $\|\cdot\|_V$, be two separable Banach spaces. Additionally, assume there is a constant $C > 0$ such that, for all $f \in V$, the bound $\|f\|_{L^2(\mathcal{T})} \leq C\|f\|_V$ holds, i.e., V is continuously embedded in $L^2(\mathcal{T})$. Further suppose that $U \overline{\otimes} U$ and $V \overline{\otimes} V$ are endowed with norms such that $\|\cdot\|_{U\overline{\otimes}U}$ is compatible with $\|\cdot\|_{V\overline{\otimes}V}$. Then, for all $B \in \mathcal{L}(U, V)$, the pathwise application to a second order random field $a \in L^2(\Omega; U)$,

$$Ba\colon \Omega \times \mathcal{T} \to \mathbb{R}, \quad (\omega, y) \mapsto B(a(\omega, \cdot))(y),$$

yields a second order random field. In particular, we have $Ba \in L^2(\Omega; V)$. Additionally, the mean of Ba is given by

$$\mathbb{E}[Ba] = Bm \in V,$$

and the covariance function is given by

$$\mathbb{E}[(Ba(\cdot, x) - Bm(x))(Ba(\cdot, y) - Bm(y))] = (B\overline{\otimes}B)(\mathrm{Cov})(x, y)$$

for all $x, y \in \mathcal{T}$.

Proof. We first note that Ba is measurable as a V-valued function, as by assumption the random field a is measurable as a U-valued function and B is continuous. This can be used to apply Proposition 2.3.4, since $\mathbb{P}$ is a finite measure, to see that Ba is $\mathbb{P}$-measurable in the sense of Definition 2.3.2 as a V-valued function. Moreover, $Ba \in L^2(\Omega; L^2(\mathcal{T}))$ is a second order random field by Proposition 2.3.6, as

$$\begin{aligned}\int_\Omega \|Ba(\omega, \cdot)\|^2_{L^2(\mathcal{T})}\, d\mathbb{P}(\omega) &\leq C^2 \int_\Omega \|Ba(\omega, \cdot)\|^2_V\, d\mathbb{P}(\omega) \\ &\leq C^2\|B\|^2_{U\to V} \int_\Omega \|a(\omega, \cdot)\|^2_U\, d\mathbb{P}(\omega) < \infty,\end{aligned}$$

which additionally proves that $Ba \in L^2(\Omega; V)$. This allows us to apply Lemma 5.1.1, which proves that the mean and covariance function of Ba are in V and $V\overline{\otimes}V$, respectively. This, on the other hand, can be used together with Remark 2.4.22 to see that $B\colon U \to V$ and $B\overline{\otimes}B\colon U\overline{\otimes}U \to V\overline{\otimes}V$ are linear and bounded. By Corollary 2.3.7, we can interchange the integral with the respective operator, which yields

$$\mathbb{E}[Ba(\cdot, x)] = \int_\Omega Ba(\omega, x)\, d\mathbb{P}(\omega) = Bm(x)$$

for the mean of Ba and

$$\begin{aligned}&\mathbb{E}[(Ba(\cdot, x) - Bm(x))\,(Ba(\cdot, y) - Bm(y))] \\ &= \int_\Omega (B(a(\omega, \cdot) - m)(x))\,(B(a(\omega, \cdot) - m)(y))\; d\mathbb{P}(\omega) \\ &= \int_\Omega B\overline{\otimes}B\,((a(\omega, \cdot) - m)\,(a(\omega, \cdot) - m))\,(x, y)\, d\mathbb{P}(\omega) \\ &= (B\overline{\otimes}B) \int_\Omega (a(\omega, x) - m(x))\,(a(\omega, y) - m(y))\; d\mathbb{P}(\omega) \\ &= (B\overline{\otimes}B)(\mathrm{Cov})(x, y),\end{aligned}$$

for the covariance function of Ba, and the claim follows. □

5.2 Kernel-based pathwise approximation

The main use of Lemma 5.1.2 in our setting is the application of kernel-based interpolation or penalized least squares approximation to a second order random field $a \in L^2(\Omega; L^2(\mathcal{D}))$. As before, we denote by $m(x) \coloneqq \mathbb{E}[a(\cdot, x)]$ the mean and by $\mathrm{Cov} \in L^2(\mathcal{D} \times \mathcal{D})$ the covariance function of a. Additionally, we suppose that m, Cov and every path of the random field a are pointwise defined as deterministic functions. With the tools proven before, the mean and covariance function of the pathwise application can be calculated explicitly.

Corollary 5.2.1.
Let $K\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ be a symmetric and positive definite function with Hilbert space $H_K(\mathcal{D}) \subseteq L^2(\mathcal{D})$. Suppose $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ is a finite point set and $\lambda \geq 0$. Then, for every $x, y \in \mathcal{D}$, the mean of the pathwise application of the approximation operator $\mathcal{I}_{X,\Phi,\lambda}$ to the random field a is given by

$$\mathbb{E}[\mathcal{I}_{X,K,\lambda}(a)](x) = K(x, X)\,(K(X, X) + \lambda I_N)^{-1}\, m(X),$$

and the covariance function by

$$\begin{aligned}&(\mathcal{I}_{X,K,\lambda} \overline{\otimes} \mathcal{I}_{X,K,\lambda})\,\mathrm{Cov}(x, y)\\ &= K(x, X)\,(K(X, X) + \lambda I_N)^{-1}\,\mathrm{Cov}(X, X)\,(K(X, X) + \lambda I_N)^{-1}\,K(X, y).\end{aligned}$$

Proof. This proof is a direct consequence of Lemma 3.5.1 and Corollary 3.5.2 in combination with Lemma 5.1.2. □

Now the question for a suitable choice of the kernel K arises. In the deterministic case considered in Chapter 3, this choice was based on regularity assumptions on the target function. However, when including randomness, we consider different cases that will be structured the following way:

The numerical point of view: In this case, we use certain smoothness properties of the random field. With respect to Section 4.3, there are two distinct cases we can consider: On the one hand, we assume the regularity of $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$, where $\sigma > \frac{d}{2}$, is known. This allows us to choose a positive definite function $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$. Here, we can directly apply the approximation methods.
On the other hand, assume $a \in L^2(\Omega; L^2(\mathcal{D}))$ is a second order random field, where the smoothness of its covariance function $\mathrm{Cov} \in H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D} \times \mathcal{D})$ is known. Then the paths of the underlying random field possess lesser regularity, which is determined by the smoothness of the centered random field $a - m$ and the smoothness of the mean m. In this case, there are again two seperate approaches we can consider: First, we can choose the kernel Φ suitable to the smoothness of the covariance function Cov or, second, suitable to the infered smoothness of the random field a. All proposed cases will be covered in the following.

The statistical point of view: From a statistical point of view, we consider so-called *linear predictors*, i.e., random fields given as

$$s(\omega, x) := \sum_{j=1}^{N} \beta_j(x) a(\omega, x_j), \quad \beta_j(x) \in \mathbb{R}, \quad x \in \mathcal{D}, \quad \omega \in \Omega.$$

It is now our goal to find the best linear predictor in the *(centered) mean square sense*, i.e., the linear predictor s^* with mean m^* such that, for all linear predictors $\widetilde{s}$ with mean $\widetilde{m}$ and $x \in \mathcal{D}$, the estimate

$$\mathbb{E}\left[((a(\cdot, x) - m(x)) - (s^*(\cdot, x) - m^*(x)))^2\right] \leq \mathbb{E}\left[((a(\cdot, x) - m(x)) - (\widetilde{s}(\cdot, x) - \widetilde{m}(x)))^2\right]$$

holds. We note that the kernel approximation with a reproducing kernel K yields a linear predictor via the (generalized) Lagrange representation as well:

$$\mathcal{I}_{X,K,\lambda} a(\cdot, x) = \sum_{j=1}^{N} a(\cdot, x_j) \chi_j(x).$$

To find the best predictor, we apply Lemma 5.1.2 to see that the mean of $\widetilde{s}$ is given by $\widetilde{m}(x) = \sum_{j=1}^{N} \beta_j(x) m(x_j)$. Thus, the mean square error can be written as

$$\begin{aligned}
&\mathbb{E}\left[((a(\cdot, x) - m(x)) - (\widetilde{s}(\cdot, x) - \widetilde{m}(x)))^2\right] \\
&= \mathbb{E}\left[(a(\cdot, x) - m(x))^2\right] - 2 \sum_{j=1}^{N} \beta_j(x) \mathbb{E}\left[(a(\cdot, x) - m(x))(a(\cdot, x_j) - m(x_j))\right] \\
&\qquad + \sum_{j=1}^{N} \sum_{k=1}^{N} \beta_j(x) \beta_k(x) \mathbb{E}\left[(a(\cdot, x_j) - m(x_j))(a(\cdot, x_k) - m(x_k))\right] \\
&= \operatorname{Cov}(x, x) - 2\beta(x)^T \operatorname{Cov}(X, x) + \beta(x)^T \operatorname{Cov}(X, X) \beta(x).
\end{aligned}$$

If we assume that the matrix $\operatorname{Cov}(X, X)$ is positive definite, then this minimization problem has a unique minimizer for every $x \in \mathcal{D}$ that is determined by the equation system

$$\operatorname{Cov}(X, X)\beta(x) = \operatorname{Cov}(X, x),$$

i.e., the minimizer is given by $\beta(x) = \operatorname{Cov}(X, X)^{-1} \operatorname{Cov}(X, x)$, which coincides with the Lagrange functions given by the kernel $K = \operatorname{Cov}$. This procedure is also known as *Kriging* in the statistical literature (c.f, [55], [47]).
Thus, the Kriging estimator chooses the covariance function of the random field a as its kernel. The main drawback of this method in our setting is that we do not know the covariance function everywhere on $\mathcal{D} \times \mathcal{D}$, hence we can not evaluate the random field that stems from the Kriging estimator outside the sampling points of X which makes this choice infeasible.
This motivates us to focus on the *numerical point of view*. In our case, it will be advantageous to consider either the *mean integrated square error* (MISE)

$$\operatorname{MISE}(a) := \mathbb{E}\left[\|a - \mathcal{I}_{X,K,\lambda}(a)\|_{L^2(\mathcal{D})}^2\right]^{\frac{1}{2}} = \|a - \mathcal{I}_{X,K,\lambda}(a)\|_{L^2(\Omega; L^2(\mathcal{D}))},$$

or the *uniform mean square error*

$$\mathrm{UMSE}(a) \coloneqq \mathbb{E}\left[\|a - \mathcal{I}_{X,K,\lambda}(a)\|^2_{C(\mathcal{D})}\right]^{\frac{1}{2}} = \|a - \mathcal{I}_{X,K,\lambda}(a)\|_{L^2(\Omega;C(\mathcal{D}))}$$

Note that the latter bounds the mean square error uniformly on $\mathcal{D}$:

$$\begin{aligned}
&\sup_{x\in\mathcal{D}} \mathbb{E}\left[(a(\cdot,x) - \mathcal{I}_{X,K,\lambda}(a)(\cdot,x))^2\right] \\
&= \sup_{x\in\mathcal{D}} \int_\Omega (a(\omega,x) - \mathcal{I}_{X,K,\lambda}(a)(\omega,x))^2 \, d\mathbb{P}(\omega) \\
&\leq \int_\Omega \left(\sup_{x\in\mathcal{D}} |a(\omega,x) - \mathcal{I}_{X,K,\lambda}(a)(\omega,x)|\right)^2 d\mathbb{P}(\omega) \\
&= \mathbb{E}\left[\|a - \mathcal{I}_{X,K,\lambda}(a)\|^2_{C(\mathcal{D})}\right].
\end{aligned}$$

5.2.1 Error estimates

We are now ready to prove error estimates in the previously mentioned cases.

First case: Known regularity of the random field

Here, we assume the regularity of the random field $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ is known. In this case, we have already seen in Lemma 5.1.1 that the mean m and covariance function Cov are also elements of $H^\sigma(\mathcal{D})$ and $H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D})$, respectively. This leads to the following result for the mean integrated square error:

Theorem 5.2.2.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ a positive definite function with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$ for $\sigma > \frac{d}{2}$. Then, for every finite point set $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ with mesh norm $h_{X,\mathcal{D}} \leq h_0$ and second order random field $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ with mean m and covariance function Cov*, the following statements hold:*

1. *For exact data values $z_j = a(\cdot, x_j)$ with $1 \leq j \leq N$ and regularization parameter $\lambda \geq 0$, the mean integrated square error can be bounded by*

$$\mathbb{E}\left[\|a - \mathcal{I}_{X,\Phi,\lambda}(a)\|^2_{L^2(\mathcal{D})}\right]^{\frac{1}{2}} \leq C\left(h^\sigma_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}}\sqrt{\lambda}\right)\|a\|_{L^2(\Omega;H^\sigma(\mathcal{D}))}.$$

   If we denote the covariance function of $\mathcal{I}_{X,\Phi,\lambda}(a)$ by Cov^**, we obtain the bound*

$$\begin{aligned}
&\|\operatorname{Cov} - \operatorname{Cov}^*\|_{L^2(\mathcal{D}\times\mathcal{D})} \\
&\quad\leq C\left(h^\sigma_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}}\sqrt{\lambda} + \left(h^\sigma_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}}\sqrt{\lambda}\right)^2\right)\|\operatorname{Cov}\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})}.
\end{aligned}$$

2. *For inexact data values $z_j = a(\cdot, x_j) + \varepsilon_j(\cdot)$ for $1 \leq j \leq N$, where the second moments $\mathbb{E}[\varepsilon_j^2] < \infty$ exist, the mean integrated square error can be bounded by*

$$\begin{aligned}
&\|a - \mathcal{I}_{X,\Phi,\lambda}(a,\varepsilon)\|_{L^2(\Omega;L^2(\mathcal{D}))} = \mathbb{E}\left[\|a - \mathcal{I}_{X,\Phi,\lambda}(a,\varepsilon)\|^2_{L^2(\mathcal{D})}\right]^{\frac{1}{2}} \\
&\leq C\left(h^\sigma_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}}\sqrt{\lambda}\right)\|a\|_{L^2(\Omega;H^\sigma(\mathcal{D}))} + \left(\frac{h^\sigma_{X,\mathcal{D}}}{\sqrt{\lambda}} + h^{\frac{d}{2}}_{X,\mathcal{D}}\right)\mathbb{E}\left[\|\varepsilon\|_2^2\right]^{\frac{1}{2}}.
\end{aligned} \tag{5.1}$$

If we denote by Cov^* *the covariance function of* $\mathcal{I}_{X,\Phi,\lambda}(a,\varepsilon)$*, then the error of the covariance functions can be bounded by*

$$\begin{aligned}
&\| \mathrm{Cov} - \mathrm{Cov}^* \|_{L^2(\mathcal{D}\times\mathcal{D})} \\
&\leq C\|a-m\|_{L^2(\Omega;L^2(\mathcal{D}))} \\
&\quad \cdot \left[\left(h^{\sigma}_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}}\sqrt{\lambda} \right) \|a-m\|_{L^2(\Omega;H^\sigma(\mathcal{D}))} + \left(\frac{h^{\sigma}_{X,\mathcal{D}}}{\sqrt{\lambda}} + h^{\frac{d}{2}}_{X,\mathcal{D}} \right) \mathbb{E}\left[\|\varepsilon - \mathbb{E}[\varepsilon]\|_2^2 \right]^{\frac{1}{2}} \right] \\
&\quad + C \left[\left(h^{\sigma}_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}}\sqrt{\lambda} \right) \|a-m\|_{L^2(\Omega;H^\sigma(\mathcal{D}))} + \left(\frac{h^{\sigma}_{X,\mathcal{D}}}{\sqrt{\lambda}} + h^{\frac{d}{2}}_{X,\mathcal{D}} \right) \mathbb{E}\left[\|\varepsilon - \mathbb{E}[\varepsilon]\|_2^2 \right]^{\frac{1}{2}} \right]^2 .
\end{aligned}$$

In all cases, $C = C(\mathcal{D},\sigma) > 0$ *denotes a constant independent of* a *and* X*.*

Proof. We first consider exact data, i.e., $\varepsilon_j = 0$, and a regularization parameter $\lambda \geq 0$. In this case, we directly apply estimate (3.4) of Theorem 3.3.4, which yields the upper bound for the mean square error:

$$\begin{aligned}
\mathbb{E}\left[\|a - \mathcal{I}_{X,\Phi,\lambda}(a)\|^2_{L^2(\mathcal{D})} \right] &= \int_\Omega \|a(\omega,\cdot) - \mathcal{I}_{X,\Phi,\lambda}(a(\omega,\cdot))\|^2_{L^2(\mathcal{D})}\, d\mathbb{P}(\omega) \\
&\leq C^2 \left(h^{\sigma}_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}}\sqrt{\lambda} \right)^2 \int_\Omega \|a(\omega,\cdot)\|^2_{H^\sigma(\mathcal{D})}\, d\mathbb{P}(\omega) \\
&= C^2 \left(h^{\sigma}_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}}\sqrt{\lambda} \right)^2 \|a\|^2_{L^2(\Omega;H^\sigma(\mathcal{D}))}.
\end{aligned}$$

Next, we consider the bound for the covariance functions. Here, we can apply Lemma 5.1.2 to see that $\mathrm{Cov}^* = \mathcal{I}_{X,\Phi,\lambda}\overline{\otimes}\mathcal{I}_{X,\Phi,\lambda}(\mathrm{Cov})$. Therefore, the estimate given in Theorem 3.5.6 directly yields the stated bound.
For inexact data values $z_j = a(\cdot,x_j) + \varepsilon_j(\cdot)$ with $\lambda > 0$, we denote by $\varepsilon := (\varepsilon_1,\dots,\varepsilon_N)$ and use the bound (3.5) of Theorem 3.3.4 to obtain

$$\begin{aligned}
&\|a - \mathcal{I}_{X,\Phi,\lambda}(a,\varepsilon)\|^2_{L^2(\Omega;L^2(\mathcal{D}))} \\
&\leq C^2 \int_\Omega \left(\left(h^{\sigma}_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}}\sqrt{\lambda} \right) \|a(\omega,\cdot)\|_{H^\sigma(\mathcal{D})} + \left(\frac{h^{\sigma}_{X,\mathcal{D}}}{\sqrt{\lambda}} + h^{\frac{d}{2}}_{X,\mathcal{D}} \right) \|\varepsilon(\omega)\|_2 \right)^2 d\mathbb{P}(\omega) \\
&\leq C \left(h^{\sigma}_{X,\mathcal{D}} + h^{\frac{d}{2}}_{X,\mathcal{D}}\sqrt{\lambda} \right)^2 \|a\|^2_{L^2(\Omega;H^\sigma(\mathcal{D}))} + C \left(\frac{h^{\sigma}_{X,\mathcal{D}}}{\sqrt{\lambda}} + h^{\frac{d}{2}}_{X,\mathcal{D}} \right)^2 \mathbb{E}\left[\|\varepsilon\|_2^2 \right].
\end{aligned}$$

However, in contrast to before, we cannot apply Lemma 5.1.2 to obtain a representation of the covariance function of $\mathcal{I}_{X,\Phi,\lambda}(a,\varepsilon)$, as the operator $\mathcal{I}_{X,\Phi,\lambda}(\cdot,\varepsilon)$ itself depends on $\omega \in \Omega$ via ε. This forces us to calculate the emerging covariance function directly. With the Lagrange representation of the approximant, the mean of $\mathcal{I}_{X,\Phi,\lambda}(a,\varepsilon)$ can be written as

$$\mathbb{E}\left[\mathcal{I}_{X,\Phi,\lambda}(a,\varepsilon) \right](x) = \sum_{j=1}^N \chi_j(x) \int_\Omega a(\omega,x_j) + \varepsilon_j(\omega)\, d\mathbb{P}(\omega) = \mathcal{I}_{X,\Phi,\lambda}(m,\mathbb{E}[\varepsilon])(x).$$

For fixed $\omega \in \Omega$ and $x \in \mathcal{D}$, we therefore obtain

$$\begin{aligned}
&\mathcal{I}_{X,\Phi,\lambda}(a,\varepsilon)(\omega,x) - \mathcal{I}_{X,\Phi,\lambda}(m,\mathbb{E}[\varepsilon])(x) \\
&= \sum_{j=1}^{N} \chi_j(x)\left(a(\omega,x_j) - m(x_j) + \varepsilon_j(\omega) - \mathbb{E}(\varepsilon_j)\right) \\
&= \mathcal{I}_{X,\Phi,\lambda}(a(\omega,\cdot) - m, \varepsilon(\omega) - \mathbb{E}[\varepsilon])(x) \\
&=: \mathcal{R}_{X,\Phi,\lambda}(\omega,x).
\end{aligned} \tag{5.2}$$

Similarly, we denote by $\mathcal{R}(\omega,x) := a(\omega,x) - m(x)$. Using Lemma 3.5.5, this allows us to express the error of the covariance functions as

$$\begin{aligned}
&\mathrm{Cov}(x,y) - \int_\Omega \mathcal{R}_{X,\Phi,\lambda}(\omega,x)\mathcal{R}_{X,\Phi,\lambda}(\omega,y)\,d\mathbb{P}(\omega) \\
&= \int_\Omega \mathcal{R}(\omega,x)\mathcal{R}(\omega,y) - \mathcal{R}_{X,\Phi,\lambda}(\omega,x)\mathcal{R}_{X,\Phi,\lambda}(\omega,y)\,d\mathbb{P}(\omega) \\
&= \int_\Omega \left(\mathcal{R}(\omega,x) - \mathcal{R}_{X,\Phi,\lambda}(\omega,x)\right)\mathcal{R}(\omega,y) + \mathcal{R}(\omega,x)\left(\mathcal{R}(\omega,y) - \mathcal{R}_{X,\Phi,\lambda}(\omega,y)\right) \\
&\qquad - \left(\mathcal{R}(\omega,x) - \mathcal{R}_{X,\Phi,\lambda}(\omega,x)\right)\left(\mathcal{R}(\omega,y) - \mathcal{R}_{X,\Phi,\lambda}(\omega,y)\right)\,d\mathbb{P}(\omega).
\end{aligned}$$

These integrals can be bounded separately using Hölder's inequality to obtain

$$\begin{aligned}
&\left|\mathrm{Cov}(x,y) - \int_\Omega \mathcal{R}_{X,\Phi,\lambda}(\omega,x)\mathcal{R}_{X,\Phi,\lambda}(\omega,y)\,d\mathbb{P}(\omega)\right| \\
&\le \|\mathcal{R}(\cdot,x) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,x)\|_{L^2(\Omega,\mathbb{P})}\|\mathcal{R}(\cdot,y)\|_{L^2(\Omega,\mathbb{P})} \\
&\qquad + \|\mathcal{R}(\cdot,x)\|_{L^2(\Omega,\mathbb{P})}\|\mathcal{R}(\cdot,y) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,y)\|_{L^2(\Omega,\mathbb{P})} \\
&\qquad + \|\mathcal{R}(\cdot,x) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,x)\|_{L^2(\Omega,\mathbb{P})}\|\mathcal{R}(\cdot,y) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,y)\|_{L^2(\Omega,\mathbb{P})},
\end{aligned}$$

for fixed $x, y \in \mathcal{D}$. The $L^2(\mathcal{D}\times\mathcal{D})$ error of this function can then be estimated by

$$\begin{aligned}
&\left\|\mathrm{Cov} - \int_\Omega \mathcal{R}_{X,\Phi,\lambda}(\omega,\cdot)\mathcal{R}_{X,\Phi,\lambda}(\omega,\cdot)\,d\mathbb{P}(\omega)\right\|^2_{L^2(\mathcal{D}\times\mathcal{D})} \\
&\le \int_\mathcal{D}\int_\mathcal{D} \Big(\|\mathcal{R}(\cdot,x) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,x)\|_{L^2(\Omega,\mathbb{P})}\|\mathcal{R}(\cdot,y)\|_{L^2(\Omega,\mathbb{P})} \\
&\qquad + \|\mathcal{R}(\cdot,x)\|_{L^2(\Omega,\mathbb{P})}\|\mathcal{R}(\cdot,y) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,y)\|_{L^2(\Omega,\mathbb{P})} \\
&\qquad + \|\mathcal{R}(\cdot,x) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,x)\|_{L^2(\Omega,\mathbb{P})}\|\mathcal{R}(\cdot,y) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,y)\|_{L^2(\Omega,\mathbb{P})}\Big)^2\,dy\,dx \\
&\le 3\int_\mathcal{D}\int_\mathcal{D} \|\mathcal{R}(\cdot,x) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,x)\|^2_{L^2(\Omega,\mathbb{P})}\|\mathcal{R}(\cdot,y)\|^2_{L^2(\Omega,\mathbb{P})} \\
&\qquad + \|\mathcal{R}(\cdot,x)\|^2_{L^2(\Omega,\mathbb{P})}\|\mathcal{R}(\cdot,y) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,y)\|^2_{L^2(\Omega,\mathbb{P})} \\
&\qquad + \|\mathcal{R}(\cdot,x) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,x)\|^2_{L^2(\Omega,\mathbb{P})}\|\mathcal{R}(\cdot,y) - \mathcal{R}_{X,\Phi,\lambda}(\cdot,y)\|^2_{L^2(\Omega,\mathbb{P})}\,dy\,dx.
\end{aligned}$$

We have now separated the arguments in the integrand and thus obtain the product of the integrals over $\mathcal{D}$. After taking the square root and using Fubini's theorem, this yields the bound

$$\begin{aligned}
&\|\mathrm{Cov} - \mathrm{Cov}^*\|_{L^2(\mathcal{D}\times\mathcal{D})} \\
&\le 2\sqrt{3}\|\mathcal{R} - \mathcal{R}_{X,\Phi,\lambda}\|_{L^2(\Omega;L^2(\mathcal{D}))}\|\mathcal{R}\|_{L^2(\Omega;L^2(\mathcal{D}))} + \sqrt{3}\|\mathcal{R} - \mathcal{R}_{X,\Phi,\lambda}\|^2_{L^2(\Omega;L^2(\mathcal{D}))}.
\end{aligned}$$

For this upper bound, we can now resubstitute $\mathcal{R}$ and $\mathcal{R}_{X,\Phi,\lambda}$, and apply the previously proven results, as $\mathcal{R}_{X,\Phi,\lambda}$ denotes the penalized least squares approximation of $\mathcal{R} = a - m$ with inexact data $\widetilde{\varepsilon} := \varepsilon - \mathbb{E}[\varepsilon]$. After rearranging, the claim follows. □

Remark.
If we consider the second case of Theorem 5.2.2 and additionally assume that the random variables ε_j are independent of $a(\cdot, x_k)$ for all $1 \leq j, k \leq N$, the covariance function Cov^* of $\mathcal{I}_{X,\Phi,\lambda}(a, \varepsilon)$ can be expressed as

$$\begin{aligned}\mathrm{Cov}^*(x, y) &= \mathcal{I}_{X,\Phi,\lambda} \overline{\otimes} \mathcal{I}_{X,\Phi,\lambda}(\mathrm{Cov})(x, y) + \chi(x)\mathcal{E}\chi(y)^T \\ &= \mathcal{I}_{X,\Phi,\lambda} \overline{\otimes} \mathcal{I}_{X,\Phi,\lambda}(\mathrm{Cov})(x, y) + \mathcal{I}_{X,\Phi,\lambda} \overline{\otimes} \mathcal{I}_{X,\Phi,\lambda}(0, \mathcal{E})(x, y),\end{aligned}$$

where $\mathcal{E}$ denotes the covariance matrix of ε, i.e.,

$$\mathcal{E} = \int_\Omega (\varepsilon(\omega) - \mathbb{E}[\varepsilon]) \otimes (\varepsilon(\omega) - \mathbb{E}[\varepsilon])\, d\mathbb{P}(\omega) \in \mathbb{R}^{N \times N}.$$

This representation follows directly by (5.2), the Lagrange representation, and the independence of the random variables.

Using analogous techniques allows us to bound the uniform mean square error of the pathwise approximation as well.

Corollary 5.2.3.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ be a positive definite function with $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$ for $\sigma > \frac{d}{2}$. Then, for every finite point set $X = \{x_1, \ldots, x_N\} \subseteq \mathcal{D}$ with $h_{X,\mathcal{D}} \leq h_0$ and every random field $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ with mean m and covariance function Cov*, the following estimates hold:*

1. *For exact data values $z_j = a(\cdot, x_j)$ and regularization parameter $\lambda \geq 0$, we obtain the bound*

$$\mathbb{E}\left[\|a - \mathcal{I}_{X,\Phi,\lambda}(a)\|^2_{C(\mathcal{D})}\right]^{\frac{1}{2}} \leq C\left(h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}} + \sqrt{\lambda}\right)\|a\|_{L^2(\Omega;H^\sigma(\mathcal{D}))}.$$

   Denoting the covariance function of $\mathcal{I}_{X,\Phi,\lambda}(a)$ by Cov^**, we obtain the bound*

$$\begin{aligned}&\|\,\mathrm{Cov} - \mathrm{Cov}^*\,\|_{C(\mathcal{D}\times\mathcal{D})} \\ &\quad\leq C\left(h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}} + \sqrt{\lambda} + \left(h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}} + \sqrt{\lambda}\right)^2\right)\|\,\mathrm{Cov}\,\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})}.\end{aligned}$$

2. *For inexact data values $z_j = a(\cdot, x_j) + \varepsilon_j(\cdot)$ for $1 \leq j \leq N$, where for every $1 \leq j \leq N$ the second moments $\mathbb{E}[\varepsilon_j^2] < \infty$ exist, the uniform mean square error can be bounded by*

$$\begin{aligned}&\|a - \mathcal{I}_{X,\Phi,\lambda}(a, \varepsilon)\|_{L^2(\Omega;C(\mathcal{D}))} = \mathbb{E}\left[\|a - \mathcal{I}_{X,\Phi,\lambda}(a)\|^2_{C(\mathcal{D})}\right] \\ &\leq C\left(h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}} + \sqrt{\lambda}\right)\|a\|_{L^2(\Omega;H^\sigma(\mathcal{D}))} + \left(\frac{h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}}}{\sqrt{\lambda}} + 1\right)\mathbb{E}\left[\|\varepsilon\|_2^2\right]^{\frac{1}{2}}.\end{aligned}$$

Denoting the covariance function of $\mathcal{I}_{X,\Phi,\lambda}(a,\varepsilon)$ *by* Cov^**, the error of the covariance functions can be bounded by*

$$\begin{aligned}
&\| \operatorname{Cov} - \operatorname{Cov}^* \|_{C(\mathcal{D}\times\mathcal{D})} \\
&\leq C\|a-m\|_{L^2(\Omega;C(\mathcal{D}))} \\
&\qquad \cdot \left[\left(h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}} + \sqrt{\lambda} \right) \|a-m\|_{L^2(\Omega;H^\sigma(\mathcal{D}))} + \left(\frac{h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}}}{\sqrt{\lambda}} + 1 \right) \mathbb{E}\left[\|\varepsilon - \mathbb{E}[\varepsilon]\|_2^2 \right]^{\frac{1}{2}} \right] \\
&\quad + C \left[\left(h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}} + \sqrt{\lambda} \right) \|a-m\|_{L^2(\Omega;H^\sigma(\mathcal{D}))} + \left(\frac{h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}}}{\sqrt{\lambda}} + 1 \right) \mathbb{E}\left[\|\varepsilon - \mathbb{E}[\varepsilon]\|_2^2 \right]^{\frac{1}{2}} \right]^2 .
\end{aligned}$$

In both cases, $C = C(\mathcal{D},\sigma) > 0$ *denotes a constant independent of* a *and* X*.*

Proof. The errors for exact data are proven completely analogously to Theorem 5.2.2 using Theorem 3.3.4 and (3.4). For the error with inexact data, we also proceed similarly to before using (3.5). □

Remark 5.2.4.
In order to interpret the error result concerning inexact data given in (5.1), assume that the measurement errors ε_j for $1 \leq j \leq N$ are all independent and identically distributed with mean zero and variance $\mathbb{E}[\varepsilon_j^2] = \gamma^2 > 0$. As already noted in the remark after Theorem 3.3.4, a reasonable choice for the regularization parameter is $\lambda = h_{X,\mathcal{D}}^{2\sigma-d}$, as this balances the occurring terms in the presented error bounds, which leads to the simplified bound

$$\|a - \mathcal{I}_{X,\Phi,\lambda}(a,\varepsilon)\|_{L^2(\Omega;L^2(\mathcal{D}))} \leq C \left(h_{X,\mathcal{D}}^{\sigma} \|a\|_{L^2(\Omega;H^\sigma(\mathcal{D}))} + h_{X,\mathcal{D}}^{\frac{d}{2}} \left(\sum_{j=1}^{N} \mathbb{E}[\varepsilon_j^2] \right)^{\frac{1}{2}} \right).$$

If we focus on the latter term controlling the measurement error, we can use Lemma 3.3.7 to set the mesh norm $h_{X,\mathcal{D}}$ of X in relation to the number of points N. Therefore, this term can further be bounded by

$$h_{X,\mathcal{D}}^{\frac{d}{2}} \left(\sum_{j=1}^{N} \mathbb{E}[\varepsilon_j^2] \right)^{\frac{1}{2}} \leq C \rho_{X,\mathcal{D}}^{\frac{d}{2}} N^{-\frac{1}{2}} \left(\sum_{j=1}^{N} \gamma^2 \right)^{\frac{1}{2}} = C \rho_{X,\mathcal{D}}^{\frac{d}{2}} \gamma.$$

With this bound, we can expect to approximate the given random field in $L^2(\mathcal{D})$ up to the standard deviation of the measurement errors. However, for the pointwise error result given in Corollary 5.2.3, we can not infer a similar a result, since the measurement errors in the upper bound can not be controlled by a suitable choice of λ or the mesh norm.

As we have seen, we can give various error bounds for different scenarios, including the use of a regularization parameter λ and the possibility of inexact data. For the sake of simplicity, we focus on interpolation with exact data in the following, but again emphasize that using the presented ideas, other scenarios can be handled as well.

Second case: Known regularity of the covariance function

Next, we consider the case where we assume the regularity $\sigma > \frac{d}{2}$ of $\mathrm{Cov} \in H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D} \times \mathcal{D})$ is known. This allows us to apply Theorem 4.3.5 to the centered random field $a - m$ and conclude that $a - m \in L^2(\Omega; H^\tau(\mathcal{D}))$ for $0 \leq \tau < \sigma - \frac{d}{2}$. In order to infer the regularity of a, we additionally have to assume that $m \in H^\beta(\mathcal{D})$ for $\beta > \frac{d}{2}$, which leads to $a \in L^2(\Omega; H^\tau(\mathcal{D}))$ for $0 \leq \tau < \min\{\sigma - \frac{d}{2}, \beta\}$. On the other hand, if we want to apply Theorem 3.3.6, we have to additionally demand that $\tau > \frac{d}{2}$. This forces us to make the more restrictive assumption $\sigma > d$. In this case, we obtain the following result.

Theorem 5.2.5.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ a positive definite function with $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$ for $\sigma > d$. Now, let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a second order random field with mean $m \in H^\beta(\mathcal{D})$ for $\beta > \frac{d}{2}$ and covariance function $\mathrm{Cov} \in H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D})$. Then, for every finite point set $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ with mesh norm $h_{X,\mathcal{D}} \leq h_0$ and every $\frac{d}{2} < \tau < \min\left\{\beta, \sigma - \frac{d}{2}\right\}$, we obtain the bounds

$$\mathbb{E}\left[\|a - \mathcal{I}_{X,\Phi}(a)\|^2_{L^2(\mathcal{D})}\right]^{\frac{1}{2}} \leq C h^\tau_{X,\mathcal{D}} \rho^{\sigma-\tau}_{X,\mathcal{D}} \|a\|_{L^2(\Omega; H^\tau(\mathcal{D}))},$$
$$\mathbb{E}\left[\|a - \mathcal{I}_{X,\Phi}(a)\|^2_{C(\mathcal{D})}\right]^{\frac{1}{2}} \leq C h^{\tau-\frac{d}{2}}_{X,\mathcal{D}} \rho^{\sigma-\tau}_{X,\mathcal{D}} \|a\|_{L^2(\Omega; H^\tau(\mathcal{D}))}.$$

Denoting the covariance function of $\mathcal{I}_{X,\Phi}(a)$ by Cov^*, the estimates

$$\|\,\mathrm{Cov} - \mathrm{Cov}^*\,\|_{L^2(\mathcal{D}\times\mathcal{D})} \leq C h^\sigma_{X,\mathcal{D}} \|\,\mathrm{Cov}\,\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})},$$
$$\|\,\mathrm{Cov} - \mathrm{Cov}^*\,\|_{C(\mathcal{D}\times\mathcal{D})} \leq C h^{\sigma-\frac{d}{2}}_{X,\mathcal{D}} \|\,\mathrm{Cov}\,\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})}$$

hold. In both cases, $C = C(\mathcal{D}, \sigma, \tau) > 0$ denotes a constant independent of X.

Proof. We only prove the $L^2(\mathcal{D})$ error estimate, as the other one can be proven analogously. As already sketched beforehand, we use Theorem 4.3.5 to conclude that the centered random field $a - m$ is an element of $L^2(\Omega; H^\tau(\mathcal{D}))$ for all $\frac{d}{2} < \tau < \sigma - \frac{d}{2}$. Since, by assumption, m is an element of $H^\beta(\mathcal{D})$, we can infer by Corollary 4.3.6 that the random field $a \in L^2(\Omega; H^\tau(\mathcal{D}))$ for all $\frac{d}{2} < \tau < \min\left\{\beta, \sigma - \frac{d}{2}\right\}$. If we then apply the kernel interpolation with Φ pathwise on a, we can thus use Theorem 3.3.6 to bound the mean square error by

$$\begin{aligned}
\mathbb{E}\left[\|a - \mathcal{I}_{X,\Phi}(a)\|^2_{L^2(\mathcal{D})}\right] &= \int_\Omega \|a(\omega,\cdot) - \mathcal{I}_{X,\Phi}(a(\omega,\cdot))\|^2_{L^2(\mathcal{D})}\, d\mathbb{P}(\omega) \\
&\leq C^2 h^{2\tau}_{X,\mathcal{D}} \rho^{2(\sigma-\tau)}_{X,\mathcal{D}} \int_\Omega \|a(\omega,\cdot)\|^2_{H^\tau(\mathcal{D})}\, d\mathbb{P}(\omega) \\
&= C^2 h^{2\tau}_{X,\mathcal{D}} \rho^{2(\sigma-\tau)}_{X,\mathcal{D}} \|a\|^2_{L^2(\Omega; H^\tau(\mathcal{D}))}.
\end{aligned}$$

For the covariance function, we again apply Lemma 5.1.2 for the representation, afterwards Theorem 3.5.6 with $\lambda_1 = \lambda_2 = 0$ for the error estimate and immediately get the claim, as $h^{2\sigma}_{X,\mathcal{D}} \leq h^\sigma_{X,\mathcal{D}}$ if the mesh norm is less than or equal to one. □

As seen in the proof beforehand, when choosing the kernel Φ to match the smoothness of the covariance function, we may *overestimate* the smoothness of the paths of the random field a. In order to prevent this, one could first use Theorem 4.3.5 in order to infer the smoothness of the paths of $a - m$ and afterwards choose the kernel Φ according to the deduced regularity. This idea is covered in the next theorem.

Theorem 5.2.6.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ a finite point set with mesh norm $h_{X,\mathcal{D}} \leq h_0$. Assume that $a \in L^2(\Omega; L^2(\mathcal{D}))$ is a second order random field with mean $m \in H^\beta(\mathcal{D})$, where $\beta > \frac{d}{2}$, and covariance function Cov *that is an element of $H^{\sigma,\sigma}_{\text{mix}}(\mathcal{D} \times \mathcal{D})$, where $\sigma > d$. Let $\frac{d}{2} < \tau < \min\left\{\beta, \sigma - \frac{d}{2}\right\}$ be fixed, and choose a positive definite function $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ with $H_\Phi(\mathbb{R}^d) \cong H^\tau(\mathbb{R}^d)$. Then $a \in L^2(\Omega; H^\tau(\mathcal{D}))$ and there is a constant $C = C(\mathcal{D}, \tau) > 0$ such that we can bound the occurring interpolation error of a and $\mathcal{I}_{X,\Phi}(a)$ by*

$$\mathbb{E}\left[\|a - \mathcal{I}_{X,\Phi}(a)\|^2_{L^2(\mathcal{D})}\right]^{\frac{1}{2}} \leq C h^\tau_{X,\mathcal{D}} \|a\|_{L^2(\Omega; H^\tau(\mathcal{D}))},$$
$$\mathbb{E}\left[\|a - \mathcal{I}_{X,\Phi}(a)\|^2_{C(\mathcal{D})}\right]^{\frac{1}{2}} \leq C h^{\tau - \frac{d}{2}}_{X,\mathcal{D}} \|a\|_{L^2(\Omega; H^\tau(\mathcal{D}))}.$$

If we denote the covariance function of $\mathcal{I}_{X,\Phi}(a)$ by Cov**, the error can be estimated by*

$$\| \operatorname{Cov} - \operatorname{Cov}^* \|_{L^2(\mathcal{D}\times\mathcal{D})} \leq C h^\tau_{X,\mathcal{D}} \| \operatorname{Cov} \|_{H^{\tau,\tau}_{\text{mix}}(\mathcal{D}\times\mathcal{D})},$$
$$\| \operatorname{Cov} - \operatorname{Cov}^* \|_{C(\mathcal{D}\times\mathcal{D})} \leq C h^{\tau - \frac{d}{2}}_{X,\mathcal{D}} \| \operatorname{Cov} \|_{H^{\tau,\tau}_{\text{mix}}(\mathcal{D}\times\mathcal{D})}.$$

Proof. As in Theorem 5.2.5 we can conclude that $a \in L^2(\Omega; H^\tau(\mathcal{D}))$. Now, by assumption, we additionally have $\tau > \frac{d}{2}$ and Cov $\in H^{\sigma,\sigma}_{\text{mix}}(\mathcal{D} \times \mathcal{D}) \subseteq H^{\tau,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{D})$, hence we are able to apply Theorem 5.2.2 and Corollary 5.2.3, respectively, and the claim follows. □

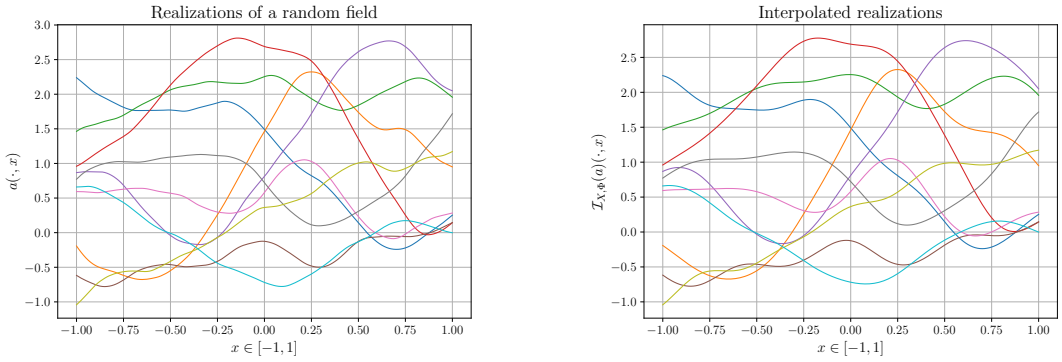


Figure 5.1: A total of 10 realizations of a Gaussian random field with mean $m(x) = 1$ and covariance function using the Matérn kernel $\Phi_{2.5}$. The left-hand side shows the true realizations, while the right-hand side shows the paths interpolated at 11 equidistant points in $[-1, 1]$ using the Wendland kernel $\Phi = \Phi_{1,2}$.

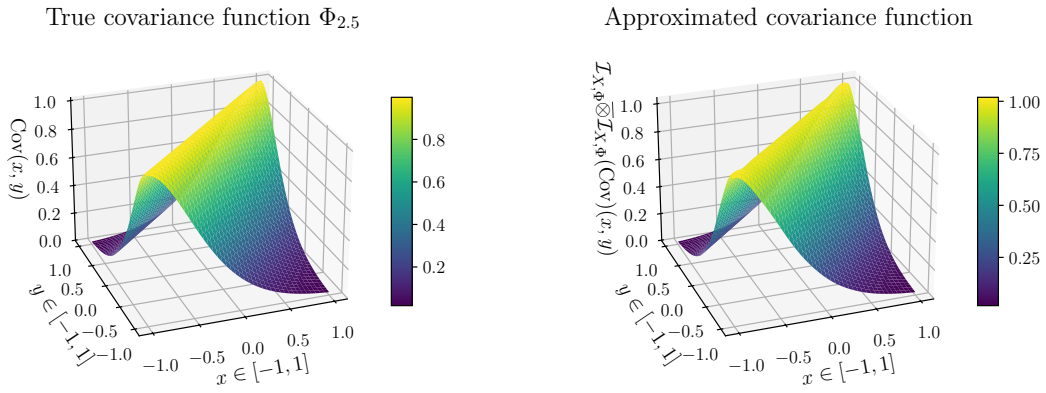


Figure 5.2: The corresponding covariance functions of the underlying random fields considered in Figure 5.1. In this example, the covariance function is given by the Matérn kernel $\Phi_{2.5}$, which reproduces $H^3(\mathbb{R}^d)$. With the use of Mercer's theorem given in Theorem 3.1.14, we can see that $\Phi_{2.5}$ permits a series expansion with eigenfunctions $e_j \in H_{\Phi_{2.5}}([0,1]) \cong H^3([0,1])$. This implies that $\Phi_{2.5} \in H^{3,3}_{\mathrm{mix}}([-1,1] \times [-1,1])$. The Wendland kernel $\Phi_{1,2}$ reproduces a Hilbert space that is equivalent to $H^3([-1,1])$ as well, thus we are in the setting of Theorem 5.2.5.

5.3 Karhunen-Loève expansion of the interpolated process

In Chapter 4, we discussed a Fourier type expansion of a given second order random field $a \in L^2(\Omega; L^2(\mathcal{D}))$, the so-called *Karhunen-Loève* expansion. In particular, we proved under some mild assumptions that a can be expanded as

$$a(\omega, x) = m(x) + \sum_{j \in I} \sqrt{\lambda_j} \xi_j(\omega) e_j(x),$$

where I is a finite or countable index set. Furthermore, if we assume without loss of generality that the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ are sorted in decreasing order, we can proceed as in Corollary 4.2.7 and reduce the complexity of the random field a by truncating this representation after at most $L \in \mathbb{N}$ terms. If I is finite, we have to choose $L < |I|$ to properly reduce the dimensionality. A major downside to this theorem is that the representation is barely constructive, as the (possible infinitely many) eigenfunctions and eigenvalues of the induced integral operator $\mathcal{C}$ by the covariance function Cov of the random field a have to be calculated. Additionally, although we have an explicit representation of the random variables ξ_j given by

$$\xi_j(\omega) = \frac{1}{\sqrt{\lambda_j}} \int_{\mathcal{D}} \left(a(\omega, x) - m(x)\right) e_j(x)\, dx,$$

the distributions of these random variables are (except for Gaussian processes) usually unknown and therefore these integrals must be approximated using, for example, quadrature rules.
However, if we want to calculate the expansion for the interpolated process $\mathcal{I}_{X,K}(a)$ using a symmetric and positive definite function $K\colon \mathcal{D}\times\mathcal{D}\to\mathbb{R}$ in a fixed point set $X\subseteq\mathcal{D}$, we can give a more explicit representation of the Karhunen-Loève expansion.
Again, for the sake of simplicity, we will not consider the penalized least squares approximation, as our results can easily be transferred to this case. Throughout this section, we assume that $a\in L^2(\Omega;L^2(\mathcal{D}))$ is a second order random field with pointwise defined paths, mean m and covariance function Cov. Moreover, $X=\{x_1,\dots,x_N\}\subseteq\mathcal{D}$ denotes a finite set of sampling points and $K\colon\mathcal{D}\times\mathcal{D}\to\mathbb{R}$ a symmetric and positive definite function. If we now denote by $a^*\coloneqq\mathcal{I}_{X,K}(a)$ the interpolated random field with mean m^* and covariance function Cov^*, we make the following crucial observation (cf. [23, Chapter 4]).

Proposition 5.3.1.
There are at most N strictly positive eigenvalues (counted with multiplicity) of the induced integral operator

$$\mathcal{C}^*\colon L^2(\mathcal{D})\to L^2(\mathcal{D}),\quad f\mapsto\int_{\mathcal{D}}\mathrm{Cov}^*(\cdot,y)f(y)\,dy.$$

In particular, the image of $\mathcal{C}^*$ fulfills $\mathcal{C}^*(L^2(\mathcal{D}))\subseteq V_N=\operatorname{span}\{K(\cdot,x_j)\mid 1\le j\le N\}$.

Proof. As we know by Corollary 5.2.1, the covariance function of $\mathcal{I}_{X,K}(a)$ is given by

$$\begin{aligned}\mathrm{Cov}^*(x,y)&=\mathcal{I}_{X,K}\overline{\otimes}\mathcal{I}_{X,K}(\mathrm{Cov})(x,y)\\&=K(x,X)K(X,X)^{-1}\,\mathrm{Cov}(X,X)K(X,X)^{-1}K(X,y)\end{aligned}$$

for $x,y\in\mathcal{D}$. If we denote the inner matrix by

$$C\coloneqq K(X,X)^{-1}\,\mathrm{Cov}(X,X)K(X,X)^{-1}=(c_{i,j})_{1\le i,j\le N}\in\mathbb{R}^{N\times N},\tag{5.3}$$

then, for arbitrary $f\in L^2(\mathcal{D})$, the integral operator $\mathcal{C}^*$ can be written as

$$\begin{aligned}\mathcal{C}^*(f)&=\int_{\mathcal{D}}f(y)\,\mathrm{Cov}^*(\cdot,y)\,dy\\&=\int_{\mathcal{D}}f(y)\sum_{j,k=1}^{N}K(\cdot,x_j)c_{j,k}K(y,x_k)\,dy\\&=\sum_{j=1}^{N}\left(\sum_{k=1}^{N}c_{j,k}\int_{\mathcal{D}}f(y)K(y,x_k)\,dy\right)K(\cdot,x_j),\end{aligned}$$

which proves that $\mathcal{C}^*(L^2(\mathcal{D}))\subseteq V_N$. Now, if $u\in L^2(\mathcal{D})$ is an eigenfunction of $\mathcal{C}^*$ associated to an eigenvalue $\lambda>0$, then

$$\mathcal{C}^*\left(\frac{1}{\lambda}u\right)=u,$$

hence, every eigenfunction also lies in V_N. Thus, there are at most N different possible eigenfunctions, as different eigenfunctions are linearly independent and $\dim(V_N)=N$. □

We also want to emphasize that we are only interested in the eigenfunctions whose corresponding eigenvalues are positive, as the others vanish in the Karhunen-Loève expansion. This result allows us to calculate the eigenvalues and eigenfunctions of this operator explicitly. They are obtained by solving the system of equations

$$\sum_{j=1}^{N} b_j \langle \mathcal{C}^* K(\cdot, x_j), K(\cdot, x_k) \rangle_{L^2(\mathcal{D})} = \lambda \sum_{j=1}^{N} b_j \langle K(\cdot, x_j), K(\cdot, x_k) \rangle_{L^2(\mathcal{D})}$$

for $1 \leq k \leq N$, as every eigenfunction u of $\mathcal{C}^*$ admits a representation $u = \sum_{j=1}^{N} b_j K(\cdot, x_j)$ and $V_N = \operatorname{span}\{K(\cdot, x_k) \mid 1 \leq k \leq N\}$. If we now define the matrices

$$A = (a_{j,k})_{1 \leq j,k \leq N} \coloneqq \Big(\langle \mathcal{C}^* K(\cdot, x_j), K(\cdot, x_k) \rangle_{L^2(\mathcal{D})} \Big)_{1 \leq j,k \leq N} \in \mathbb{R}^{N \times N}$$

and

$$M = (m_{j,k})_{1 \leq j,k \leq N} \coloneqq \Big(\langle K(\cdot, x_j), K(\cdot, x_k) \rangle_{L^2(\mathcal{D})} \Big)_{1 \leq j,k \leq N} \in \mathbb{R}^{N \times N},$$

we can write this system of equations as

$$Ab = \lambda M b, \quad b = (b_1, \ldots, b_N) \in \mathbb{R}^N,$$

which yields a generalized eigenvalue problem in $\mathbb{R}^N$. In order to transform this to an ordinary eigenvalue problem, we first observe that $A = (a_{j,k})_{1 \leq j,k \leq N}$ can be rewritten as

$$\begin{aligned}
a_{j,k} &= \langle \mathcal{C}^* K(\cdot, x_j), K(\cdot, x_k) \rangle_{L^2(\mathcal{D})} \\
&= \int_{\mathcal{D}} \int_{\mathcal{D}} \operatorname{Cov}^*(x, y) K(y, x_j)\, dy\, K(x, x_k)\, dx \\
&= \sum_{\ell,m=1}^{N} c_{\ell,m} \int_{\mathcal{D}} \int_{\mathcal{D}} K(x_\ell, x) K(y, x_m) K(y, x_j) K(x, x_k)\, dy\, dx \\
&= \sum_{\ell,m=1}^{N} c_{\ell,m} \left(\int_{\mathcal{D}} K(y, x_m) K(y, x_j)\, dy \right) \left(\int_{\mathcal{D}} K(x, x_l) K(x, x_k)\, dx \right) \\
&= M_j^T C M_k,
\end{aligned}$$

where M_j denotes the j-th row of M, and C is given as in (5.3). Hence, A is simply given by $A = MCM$.
We will prove in Lemma 5.3.2 below that M is a positive definite matrix. Hence, we can compute the Cholesky decomposition $M = LL^T$ of M. If we set $c \coloneqq L^T b$, the generalized eigenvalue problem can equivalently be written as

$$\begin{aligned}
& LL^T C L L^T b = \lambda L L^T b \\
\Leftrightarrow\ & L^T C L c = \lambda c.
\end{aligned}$$

Thus, we reformulated the generalized eigenvalue Problem $Ab = \lambda M b$ as an ordinary eigenvalue problem using the positive semi-definite matrix $L^T C L$. This eigenvalue problem can be solved to obtain eigenpairs $(\lambda_j, c^{(j)}) \in [0, \infty[\times \mathbb{R}^N$ with orthonormal eigenvectors. By backwards substitution for $L^T b^{(j)} = c^{(j)}, 1 \leq j \leq N$, this immediately yields the

original eigenvectors of $Ab^{(j)} = \lambda_j M b^{(j)}$. For fixed index $1 \leq j \leq N$, we can construct the eigenfunction e_j of $\mathcal{C}^*$ using the vector $b^{(j)} = (b_1^{(j)}, \dots, b_N^{(j)})^T \in \mathbb{R}^N$ by setting

$$e_j(x) = \sum_{k=1}^{N} b_k^{(j)} K(x, x_j) = \left(c^{(j)}\right)^T L^{-1} K(X, x), \quad 1 \leq j \leq N, \quad x \in \mathcal{D}.$$

With this explicit expression for the eigenfunctions of the covariance operator $\mathcal{C}^*$, we can examine the occurring random variables ξ_j in more detail. By Lemma 5.1.2, the mean of $\mathcal{I}_{X,K}(a)$ is given by $\mathcal{I}_{X,K}(m)$, thus, for every eigenvalue $\lambda_j > 0$ with $1 \leq j \leq N$, the random variable ξ_j of the Karhunen-Loève expansion Corollary 4.2.5 is given by

$$\xi_j(\omega) = \frac{1}{\sqrt{\lambda_j}} \int_{\mathcal{D}} \left(\mathcal{I}_{X,K}(a(\omega, \cdot))(x) - \mathcal{I}_{X,K}(m)(x)\right) e_j(x)\, dx.$$

When inserting the formulas for $\mathcal{I}_{X,K}(a(\omega, \cdot) - m)$ and $e_j = K(\cdot, X) L^{-T} c^{(j)}$, we see that

$$\begin{aligned}
\xi_j(\omega) &= \frac{1}{\sqrt{\lambda_j}} \int_{\mathcal{D}} \left(a(\omega, X) - m(X)\right)^T K(X, X)^{-1} K(X, x) K(x, X) L^{-T} c^{(j)}\, dx \\
&= \frac{1}{\sqrt{\lambda_j}} \int_{\mathcal{D}} \left(a(\omega, X) - m(X)\right)^T K(X, X)^{-1} K(X, x) \otimes K(X, x) L^{-T} c^{(j)}\, dx,
\end{aligned}$$

where $K(X, x) \otimes K(X, x) \coloneqq (K(x_i, x) K(x_k, x))_{1 \leq i,k \leq N}$ denotes the dyadic product of these vectors. When we expand the term inside the integral we can apply the integral componentwise on $K(X, x) \otimes K(X, x)$, which then just yields the matrix $M = LL^T$. This leads to the final expression for ξ_j as

$$\xi_j(\omega) = \frac{1}{\sqrt{\lambda_j}} \left(a(\omega, X) - m(X)\right)^T K(X, X)^{-1} L c^{(j)}, \quad 1 \leq j \leq N.$$

Before we summarize these considerations, we need to show that M is indeed a positive definite matrix.

Lemma 5.3.2.
Let $K \colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ be a symmetric and positive definite function. Then the mapping

$$G \colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}, \quad (x, y) \mapsto \langle K(\cdot, x), K(\cdot, y) \rangle_{L^2(\mathcal{D})}$$

is positive definite as well.

Proof. Obviously, G is symmetric. Next, Let $\alpha = (\alpha_1, \dots, \alpha_N) \in \mathbb{R}^N$ and a finite point set $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ be given. Then the quadratic form of G can be rewritten as

$$\begin{aligned}
\sum_{j,k=1}^{N} \alpha_j \alpha_k G(x_j, x_k) &= \int_{\mathcal{D}} \sum_{j,k=1}^{N} \alpha_j \alpha_k K(x_j, y) K(x_k, y)\, dy \\
&= \int_{\mathcal{D}} f(y)^2\, dy \geq 0,
\end{aligned}$$

where

$$f \coloneqq \sum_{j=1}^{N} \alpha_j K(\cdot, x_j).$$

Hence, G is positive semi-definite. Since K is positive definite, we also know that the functions $\{K(\cdot, x) \mid x \in X\}$ are linearly independent and thus f can only vanish when $\alpha = 0$. This proves the positive definiteness of G. □

Combining the representations of the eigenfunctions e_j and the random variables ξ_j yields:

Theorem 5.3.3.
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a second order random field, $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ a finite point set, and $K \in C(\mathcal{D} \times \mathcal{D})$ a symmetric and positive definite function on $\mathcal{D}$. Define the matrices

$$M \coloneqq (\langle K(\cdot, x_j), K(\cdot, x_k)\rangle_{L^2(\mathcal{D})})_{1 \le j,k \le N} = LL^T,$$

and

$$C \coloneqq K(X,X)^{-1} \operatorname{Cov}(X,X) K(X,X)^{-1},$$

and denote by $(\lambda_j, c^{(j)}) \in]0, \infty[\times \mathbb{R}^N$ for $1 \le j \le N$ the eigenpairs with positive eigenvalues and normalized eigenvectors of the positive semi-definite matrix $L^T C L$. Furthermore, let $b^{(j)} \in \mathbb{R}^N$ be the solutions of $L^T b^{(j)} = c^{(j)}$ for $1 \le j \le N$. Then the Karhunen-Loève expansion of $\mathcal{I}_{X,K}(a)$ can be written as

$$\mathcal{I}_{X,K}(a)(\cdot, x) = \mathcal{I}_{X,K}(m)(x) + \sum_{j=1}^{N} \sqrt{\lambda_j} \xi_j(\cdot) e_j(x),$$

where

$$e_j(x) = \left(b^{(j)}\right)^T K(X,x) = \left(c^{(j)}\right)^T L^{-1} K(X,x), \quad x \in \mathcal{D}$$

and

$$\xi_j(\omega) = \frac{1}{\sqrt{\lambda_j}} \left(a(\omega, X) - m(X)\right)^T K(X,X)^{-1} L c^{(j)}, \quad \omega \in \Omega$$

for all $1 \le j \le N$.

Remark 5.3.4.
Without truncation, we can recover the kernel interpolation from the expansion by

$$\begin{aligned}
&\mathcal{I}_{X,K}(m)(x) + \sum_{j=1}^{N} \sqrt{\lambda_j} \xi_j(\omega) e_j(x) \\
&= \mathcal{I}_{X,K}(m)(x) + \left(a(\omega, X) - m(X)\right) K(X,X)^{-1} L \left(\sum_{j=1}^{N} c^{(j)} \otimes c^{(j)} \right) L^{-1} K(X,x) \\
&= \mathcal{I}_{X,K}(m)(x) + \mathcal{I}_{X,K}(a-m)(\omega, x) = \mathcal{I}_{X,K}(a)(\omega, x),
\end{aligned}$$

as the eigenvectors $c^{(j)}$ form an orthonormal basis of $\mathbb{R}^n$, thus, $\sum_{j=1}^{N} c^{(j)} \otimes c^{(j)} = I_N$.

Decay of eigenvalues

As we have seen in Corollary 4.2.7, the truncation error of the Karhunen-Loève expansion of a random field $a \in L^2(\Omega; L^2(\mathcal{D}))$ can be expressed using the eigenvalues of the associated covariance operator $\mathcal{C}$ of a. In our specific case, the Karhunen-Loève expansion of $\mathcal{I}_{X,K}(a)$ only consists of at most $N = |X|$ distinct eigenvalues. Thus, we choose $L < N$ and denote the truncated expansion of $\mathcal{I}_{X,K}(a)$ by

$$\mathcal{I}_{X,K}^{(L)}(a)(\omega, x) := \mathcal{I}_{X,K}(m)(x) + \sum_{j=1}^{L} \sqrt{\widetilde{\lambda_j}}\xi_j(\omega)e_j(x),$$

where $x \in \mathcal{D}$ and $\omega \in \Omega$ and $\widetilde{\lambda_j}$ denote the eigenvalues of the covariance operator corresponding to $\mathcal{I}_{X,K}(a)$ for $1 \leq j \leq N$. The resulting error is then given as

$$\left\|\mathcal{I}_{X,K}(a) - \mathcal{I}_{X,K}^{(L)}(a)\right\|^2_{L^2(\Omega;L^2(\mathcal{D}))} = \sum_{j=L+1}^{N} \widetilde{\lambda}_j.$$

Yet, as there are now finitely many eigenvalues, we can not use the decay rates for eigenvalues given in Theorem 4.2.8 directly. We therefore have to compare the eigenvalues of the approximated random field with the ones of the covariance operator of a. We obtain the following abstract bound.

Lemma 5.3.5.
Let $K_j \in L^2(\mathcal{D} \times \mathcal{D})$, for $j = 1, 2$, be two symmetric kernels with induced integral operator

$$\mathcal{K}_j \colon L^2(\mathcal{D}) \to L^2(\mathcal{D}), \quad f \mapsto \int_{\mathcal{D}} K_j(\cdot, y) f(y)\, dy.$$

Denote by $\left(\lambda_k^{(j)}\right)_{k\in\mathbb{N}}$ the decreasing sequence of eigenvalues of $\mathcal{K}_j$ for $j = 1, 2$. Then, for all $k \in \mathbb{N}$, the bound

$$\left|\lambda_k^{(1)} - \lambda_k^{(2)}\right| \leq \|K_1 - K_2\|_{L^2(\mathcal{D}\times\mathcal{D})} \tag{5.4}$$

holds.

Proof. Using the Courant-Fischer principle for eigenvalues ([28, Chapter 1, § 6.10]), we can estimate for $k \in \mathbb{N}$

$$\begin{aligned}
\lambda_k^{(1)} &= \max_{\substack{U \subseteq L^2(\mathcal{D}) \\ \dim(U)=k}} \min_{\substack{g \in U \\ \|g\|_{L^2(\mathcal{D})}=1}} \langle \mathcal{K}_1(g), g\rangle_{L^2(\mathcal{D})} \\
&= \max_{\substack{U \subseteq L^2(\mathcal{D}) \\ \dim(U)=k}} \min_{\substack{g \in U \\ \|g\|_{L^2(\mathcal{D})}=1}} \langle \mathcal{K}_2(g), g\rangle_{L^2(\mathcal{D})} + \langle \mathcal{K}_1(g) - \mathcal{K}_2(g), g\rangle_{L^2(\mathcal{D})} \\
&\leq \lambda_k^{(2)} + \|\mathcal{K}_1 - \mathcal{K}_2\|_{L^2(\mathcal{D})\to L^2(\mathcal{D})}.
\end{aligned}$$

We now subtract $\lambda_k^{(2)}$ on both sides. By swapping the roles $\mathcal{K}_1$ and $\mathcal{K}_2$ we obtain

$$\left|\lambda_k^{(1)} - \lambda_k^{(2)}\right| \leq \|\mathcal{K}_1 - \mathcal{K}_2\|_{L^2(\mathcal{D})\to L^2(\mathcal{D})}.$$

This can further be bound by $\|K_1 - K_2\|_{L^2(\mathcal{D}\times\mathcal{D})}$ using the integral representations of $\mathcal{K}_1$ and $\mathcal{K}_2$, and the claim follows. □

A major downside of Lemma 5.3.5 is that the estimate in (5.4) is uniform in $k \in \mathbb{N}$ for all eigenvalues $\lambda_k^{(1)}$ and $\lambda_k^{(2)}$, which indicates a pessimistic upper bound.

Next, we apply this lemma with $K_1 = \mathcal{I}_{X,K}\overline{\otimes}\mathcal{I}_{X,k}(\mathrm{Cov})$ and $K_2 = \mathrm{Cov}$. If we denote the eigenvalues of the latter with $\lambda_j, j \in \mathbb{N}$, the truncation error can be estimated by

$$\begin{aligned}
&\left\|\mathcal{I}_{K,X}(a) - \mathcal{I}_{K,X}^{(L)}(a)\right\|^2_{L^2(\Omega;L^2(\mathcal{D}))} \\
&= \sum_{k=L+1}^{N} \widetilde{\lambda}_k = \sum_{k=L+1}^{N} \lambda_k + \left(\widetilde{\lambda}_k - \lambda_k\right) \\
&\leq \left(\sum_{k=L+1}^{N} \lambda_k\right) + (N-L)\left\|\mathrm{Cov} - \mathcal{I}_{X,K}\overline{\otimes}\mathcal{I}_{X,k}(\mathrm{Cov})\right\|_{L^2(\mathcal{D}\times\mathcal{D})}.
\end{aligned}$$

This allows us to use the decay estimates given in Theorem 4.2.8 combined with error estimates for the covariance function to estimate the occurring truncation error. This theoretical bound can then be used a-priori to choose a suitable truncation parameter $L < N$.

5.4 Covariance estimation

Using kernel-based methods for the approximation of random fields, we have already seen that we can directly calculate the resulting covariance function of the process. In particular, if $a \in L^2(\Omega; L^2(\mathcal{D}))$ is a second order random field, $K\colon \mathcal{D}\times\mathcal{D} \to \mathbb{R}$ the reproducing kernel of $H_K(\mathcal{D}) \subseteq L^2(\mathcal{D})$ and $X = \{x_1, \ldots, x_N\} \subseteq \mathcal{D}$ a finite point set, then, by Corollary 5.2.1, the covariance function of $\mathcal{I}_{X,K,\lambda}(a)$ with $\lambda \geq 0$ is given by

$$\begin{aligned}
&(\mathcal{I}_{X,K,\lambda}\overline{\otimes}\mathcal{I}_{X,K,\lambda})\,\mathrm{Cov}(x,y) \\
&= K(x,X)\,(K(X,X)+\lambda I_N)^{-1}\,\mathrm{Cov}(X,X)\,(K(X,X)+\lambda I_N)^{-1}\,K(X,y).
\end{aligned}$$

To compute this approximation, we therefore need to know the matrix $\mathrm{Cov}(X,X)$, which is not covered by the assumptions we made throughout this thesis so far. Of course, the simplest method would be to assume that this matrix is given. However, by definition, $\mathrm{Cov}(X,X)$ is given as the *covariance matrix* of the $\mathbb{R}^N$-valued random variable $a(\cdot,X)$, which allows us to use standard statistical estimators to approximate $\mathrm{Cov}(X,X)$. In our setting, usually the dimension N is large, which is why we seek to use estimators tailored to this high-dimensional problem.
This motivates us to cite results given in [10], which prove optimality results for estimators on a specific subclass of multivariate random variables ([10, Definition 1]).

Definition 5.4.1 (Sub-Gaussian)**.**
Let $A\colon \Omega \to \mathbb{R}^N$ be a random variable. We call A *sub-Gaussian* if there exists a constant $\rho > 0$ such that, for all $v \in \mathbb{R}^N$ with $\|v\|_2 = 1$ and $t > 0$,

$$\mathbb{P}\left(|v^T(A - \mathbb{E}[A])| > t\right) \leq \exp\left(-t^2\frac{\rho}{2}\right)$$

holds.

For a given random variable $A\colon \Omega \to \mathbb{R}^N$, we denote with $\Sigma(A) = (\sigma_{i,j}) \in \mathbb{R}^{N\times N}$ the corresponding covariance matrix of A. The following result is a summary of [11, Chapter 3.3] or given in [10, Theorem 1].

Theorem 5.4.2.
Let $T_0, T > 0$, define

$$\begin{aligned}&\mathcal{F}_\alpha(T_0, T)\\ &:= \left\{ \Sigma = (\sigma_{i,j}) \in \mathbb{R}^{N\times N} \,\middle|\, \max_{1\le j\le N} \sum_{\substack{i=1\\ |j-i|>c}}^{N} |\sigma_{i,j}| \le T c^{-\alpha} \text{ for } 1 \le c \le N,\ \lambda_{\max}(\Sigma) \le T_0 \right\}\end{aligned}$$

and denote by $\mathcal{P}_\alpha = \mathcal{P}_\alpha(T_0, T)$ the set of all random variables $A\colon \Omega \to \mathbb{R}^N$ such that A is sub-Gaussian and the corresponding covariance matrix $\Sigma(A)$ is an element of $\mathcal{F}_\alpha(T_0, T)$. Now, let $M \in \mathbb{N}$ denote the number of independent samples of a random variable and assume there exists a constant $\varepsilon > 0$ such that $N < \exp(\varepsilon M)$. Then, the so-called minimax rate *of all estimators $\widehat{\Sigma} \in \mathbb{R}^{N\times N}$ estimating the covariance matrix of a given $\mathbb{R}^N$-valued random variable using M independent samples is given by*

$$\inf_{\widehat{\Sigma}} \sup_{A\in\mathcal{P}_\alpha} \|\widehat{\Sigma} - \Sigma(A)\|_{2\to 2} \in \begin{cases} \Theta\left(M^{-\frac{2\alpha}{2\alpha+1}} + \frac{\ln(N)}{M}\right), & N \ge M^{\frac{1}{2\alpha+1}}, \\ \Theta\left(\frac{N}{M}\right), & N < M^{\frac{1}{2\alpha+1}}. \end{cases}$$

We are interested in constructive ways to obtain these convergence rates. Although the following results also hold for arbitrary multivariate random variables $A\colon \Omega \to \mathbb{R}^N$, we tailor the subsequent results to our setting $A = a(\cdot, X)$. First, we consider the *sample covariance matrix*.

Definition 5.4.3.
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a second order random field with covariance function Cov, and let $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ be a finite point set. Let $a_1(\cdot, X), \dots, a_M(\cdot, X)$ be $M \in \mathbb{N}$ independent copies of $a(\cdot, X)$.

1. The *sample mean* or *empirical mean* of $a(\cdot, X)$ is defined as

$$\overline{a(\cdot, X)} := \frac{1}{M} \sum_{k=1}^{M} a_k(\cdot, X).$$

2. The *sample covariance matrix* of $a(\cdot, X)$ is given as

$$\frac{1}{M-1} \sum_{k=1}^{M} \left(a_k(\cdot, X) - \overline{a(\cdot, X)}\right) \otimes \left(a_k(\cdot, X) - \overline{a(\cdot, X)}\right).$$

In [11], the following rate-optimal estimators are given:

Theorem 5.4.4.
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a second order random field, $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$, and let $a_1(\cdot, X), \dots, a_M(\cdot, X)$ be $M \in \mathbb{N}$ independent copies of $a(\cdot, X)$. First, assume that the dimension fulfills $N \geq M^{\frac{1}{2\alpha+1}}$. For an even natural number $k \in \mathbb{N}$, we define weights $\omega_{i,j}$ for $1 \leq i, j \leq N$ as

$$\omega_{i,j} := \begin{cases} 1, & |i-j| \leq \frac{k}{2}, \\ 2\left(1 - \frac{|i-j|}{k}\right), & \frac{k}{2} < |i-j| < k, \\ 0, & |i-j| \geq k. \end{cases}$$

If we denote by $\widetilde{\Sigma} = (\widetilde{\sigma}_{i,j})$ the sample covariance matrix of $a(\cdot, X)$, then

$$\widehat{\Sigma} := (\omega_{i,j}\widetilde{\sigma}_{i,j})_{1 \leq i,j \leq N}$$

is the so-called tapering estimator *of* $\mathrm{Cov}(X, X)$ *and satisfies*

$$\left\| \widehat{\Sigma} - \mathrm{Cov}(X, X) \right\|_{2 \to 2} \in \Theta\left(M^{-\frac{2\alpha}{2\alpha+1}} + \frac{\ln(N)}{M} \right)$$

when choosing $k \in \mathbb{N}$ proportional to $M^{1/(2\alpha+1)} \in \mathbb{N}$. For $N < M^{\frac{1}{2\alpha+1}}$, a rate optimal estimator is given by the (untapered) sample covariance matrix given in Definition 5.4.3.

We note that although the tapering estimator yields a rate optimal estimator for the covariance matrix $\mathrm{Cov}(X, X)$, the estimator itself is not in general a positive (semi-) definite matrix (cf. [11, Remark 3]). As suggested there, to obtain a suitable approximation of the covariance matrix, one first has to diagonalize the tapering estimator and then replace all negative eigenvalues by zero.

6 Kernel-based multilevel method

6.1 Motivation

In the last two chapters, we proposed a method using positive definite functions to calculate an interpolant or a penalized least squares approximant to a deterministic function and generalized this idea to random fields. A main drawback of the stated methods has been indicated in Corollary 3.4.3, as we may expect ill-conditioned kernel matrices whenever the mesh norm $h_{X,\mathcal{D}}$ and therefore also the separation radius q_X become small, which ultimately leads to numerical instabilities when computing the approximation. To overcome these instabilities, we introduce a different method based on kernel interpolation. In the following, we restrict ourselves to translation invariant positive definite functions $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$, but whose support is compact, as for example the Wendland kernels. To motivate the following idea, we assume for the time being that $\operatorname{supp}(\Phi) = \overline{B_1(0)}$. For a fixed point set $X = \{x_1, \dots, x_N\} \subseteq \mathbb{R}^d$, this implies that all entries $\Phi(x_j - x_k)$ with $\|x_j - x_k\|_2 > 1$ vanish, leading to a certain sparsity of the interpolation matrix $\Phi(X, X)$. We now want to use this property in the following way: To generate more vanishing entries in the kernel matrix, we will *scale* the function Φ by a factor $0 < \delta < 1$ such that the support of the scaled kernel Φ_δ is $\overline{B_\delta(0)}$. For this, we first state the following lemma, which allows us to rescale Φ without algebraically changing the resulting reproducing Hilbert space (see [57, Lemma 1]).

Lemma 6.1.1.
Let $0 < \delta \leq 1$ and $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ be a positive definite function with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$. Define

$$\Phi_\delta := \delta^{-d} \Phi\left(\frac{\cdot}{\delta}\right).$$

If we denote by c_1 and c_2 the equivalence constants of $\widehat{\Phi}$, i.e,

$$c_1 \left(1 + \|\omega\|_2^2\right)^{-\sigma} \leq \widehat{\Phi}(\omega) \leq c_2 \left(1 + \|\omega\|_2^2\right)^{-\sigma},$$

then the Fourier transform of Φ_δ satisfies

$$c_1 \left(1 + \delta^2\|\omega\|_2^2\right)^{-\sigma} \leq \widehat{\Phi_\delta}(\omega) \leq c_2 \left(1 + \delta^2\|\omega\|_2^2\right)^{-\sigma},$$

which in particular implies that

$$c_1 \left(1 + \|\omega\|_2^2\right)^{-\sigma} \leq \widehat{\Phi_\delta}(\omega) \leq c_2 \delta^{-2\sigma} \left(1 + \|\omega\|_2^2\right)^{-\sigma}.$$

Therefore, $H_{\Phi_\delta}(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, and the norm equivalence

$$\sqrt{c_1}(2\pi)^{\frac{d}{4}} \|f\|_{H_{\Phi_\delta}(\mathbb{R}^d)} \leq \|f\|_{H^\sigma(\mathbb{R}^d)} \leq \sqrt{c_2}(2\pi)^{\frac{d}{4}} \delta^{-\sigma} \|f\|_{H_{\Phi_\delta}(\mathbb{R}^d)}$$

holds for all $f \in H^\sigma(\mathbb{R}^d)$.

Proof. We first observe that

$$\widehat{\Phi_\delta} = \widehat{\Phi}(\delta\,\cdot),$$

which immediately proves the first statement:

$$c_1 \left(1 + \delta^2 \|\omega\|_2^2\right)^{-\sigma} \leq \widehat{\Phi_\delta}(\omega) \leq c_2 \left(1 + \delta^2 \|\omega\|_2^2\right)^{-\sigma}.$$

As $\delta \leq 1$, this estimate can be used to obtain

$$\begin{aligned} c_1 \left(1 + \|\omega\|_2^2\right)^{-\sigma} &\leq c_1 \left(1 + \delta^2 \|\omega\|_2^2\right)^{-\sigma} \leq \widehat{\Phi_\delta}(\omega) \leq c_2 \left(1 + \delta^2 \|\omega\|_2^2\right)^{-\sigma} \\ &= c_2 \delta^{-2\sigma} \left(\delta^{-2} + \|\omega\|_2^2\right)^{-\sigma} \leq c_2 \delta^{-2\sigma} \left(1 + \|\omega\|_2^2\right)^{-\sigma}. \end{aligned}$$

Thus, Φ_δ induces a Hilbert space which is equivalent to $H^\sigma(\mathbb{R}^d)$. The norm equivalence now follows directly by plugging the above estimate into the norm representation given in Theorem 3.1.7. □

In order to obtain bounds on the condition number of the scaled kernel matrix $\Phi_\delta(X, X)$, we have to couple the scaling parameter δ to the mesh norm $h_{X,\mathcal{D}}$ of X. This leads to the following bound.

Theorem 6.1.2.
Let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ be a positive definite function with compact support $\overline{B_1(0)}$ and native space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$. Suppose $\mathcal{D} \subseteq \mathbb{R}^d$ is a bounded domain and $X = \{x_1, \ldots, x_N\} \subseteq \mathcal{D}$ a finite point set. Furthermore, let there be a constant $c^* \geq \rho_{X,\mathcal{D}}$, i.e., c^* fulfills $h_{X,\mathcal{D}} \leq c^* q_X$. Choose a parameter $\nu > 1$ such that $\delta := \nu h_{X,\mathcal{D}} < 1$, and set $\Phi_\delta := \delta^{-d} \Phi\left(\frac{\cdot}{\delta}\right)$ as in Lemma 6.1.1.
Then the maximum eigenvalue of $\Phi_\delta(X, X)$ can be bounded from above by

$$\lambda_{max}(\Phi_\delta(X, X)) \leq C \left(1 + \nu c^*\right)^d \delta^{-d} \Phi(0),$$

and the minimum eigenvalue from below by

$$\lambda_{min}(\Phi_\delta(X, X)) \geq C \left(\nu c^*\right)^{d-2\sigma} \delta^{-d}.$$

Thus, the condition number of the matrix $\Phi_\delta(X, X)$ is bounded independently of δ by

$$\mathrm{cond}_2(\Phi_\delta(X, X)) \leq C\Phi(0)(1 + \nu c^*)^d (\nu c^*)^{2\sigma - d},$$

where in each estimate $C = C(\Phi, \mathcal{D}, \sigma) > 0$ denotes a constant independent of X.

Proof. The estimates for the minimum and maximum eigenvalue are both stated in the proof of [52, Proposition 3.4]. For brevity, we omit the bound for the largest eigenvalue of $\Phi_\delta(X, X)$.
The idea for the minimal eigenvalue presented in the aforementioned reference is to apply Theorem 3.4.2 or, more specifically, Corollary 3.4.3 as follows: We consider the matrix

$$\widetilde{A} := \delta^d \Phi_\delta(X, X) = \left(\Phi\left(\frac{x_j - x_k}{\delta}\right)\right)_{1 \leq j,k \leq N}.$$

Then, for the minimum eigenvalues, the relation $\lambda_{\min}(\Phi_\delta(X,X)) = \delta^{-d}\lambda_{\min}(\widetilde{A})$ holds. Moreover, $\widetilde{A}$ can be interpreted as the kernel matrix of Φ evaluated at the point set $X_\delta := \{\frac{x_1}{\delta}, \dots, \frac{x_N}{\delta}\}$, i.e., $\widetilde{A} = \Phi(X_\delta, X_\delta)$. As a simple calculation shows, the separation radius of X_δ is given by

$$q_{X_\delta} = \frac{q_X}{\delta} \geq (\nu c^*)^{-1}.$$

We apply Corollary 3.4.3 (note that the lower bound stated there is independent of $\mathcal{D}$ and purely relies on the separation radius of the point set) to obtain the bound for the smallest eigenvalue by

$$\lambda_{\min}(\widetilde{A}) \geq C q_{X_\delta}^{2\sigma-d} \geq C(\nu c^*)^{d-2\sigma}.$$

This implies the bounds on the smallest eigenvalue of $\Phi_\delta(X,X)$ stated in the theorem and the bound for the condition number of $\Phi_\delta(X,X)$ follows immediately. □

This result allows us to control the condition number of the kernel matrix $\Phi_\delta(X,X)$ with the parameter ν. The error estimates for these scaled kernels are given in [58, Theorem 7] and read as follows.

Corollary 6.1.3.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary, $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ with mesh norm $h_{X,\mathcal{D}} \leq h_0$, and $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ a positive definite function with $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$. Then, for any $0 < \delta \leq 1$, we obtain the bound

$$\|f - \mathcal{I}_{X,\Phi_\delta}(f)\|_{L^2(\mathcal{D})} \leq C \left(\frac{h_{X,\mathcal{D}}}{\delta}\right)^\sigma \|f\|_{H^\sigma(\mathcal{D})}.$$

Proof. Using the sampling inequality of Theorem 3.3.2 yields

$$\|f - \mathcal{I}_{X,\Phi_\delta}(f)\|_{L^2(\mathcal{D})} \leq C h_{X,\mathcal{D}}^\sigma \|f - \mathcal{I}_{X,\Phi_\delta}(f)\|_{H^\sigma(\mathcal{D})}.$$

Now, consider the extension Ef of f to $\mathbb{R}^d$ using Theorem 2.2.9, and observe that the interpolation satisfies $\mathcal{I}_{X,\Phi_\delta}(f) = \mathcal{I}_{X,\Phi_\delta}(Ef)$. Then

$$\begin{aligned}
\|f - \mathcal{I}_{X,\Phi_\delta}(f)\|_{H^\sigma(\mathcal{D})} &\leq \|Ef - \mathcal{I}_{X,\Phi_\delta}(Ef)\|_{H^\sigma(\mathbb{R}^d)} \\
&\leq \sqrt{c_2}\delta^{-\sigma}\|Ef - \mathcal{I}_{X,\Phi_\delta}(Ef)\|_{H_{\Phi_\delta}(\mathbb{R}^d)} \\
&\leq \sqrt{c_2}\delta^{-\sigma}\|Ef\|_{H_{\Phi_\delta}(\mathbb{R}^d)} \\
&\leq C_E \frac{\sqrt{c_2}}{\sqrt{c_1}}\delta^{-\sigma}\|f\|_{H^\sigma(\mathcal{D})}
\end{aligned}$$

holds, where C_E denotes the constant of the embedding operator E. Putting both estimates together yields the claim. □

Here, the following problem occurs: When coupling $\delta = \nu h_{X,\mathcal{D}}$ with a fixed constant $\nu > 0$, we lose the approximation power provided by $h_{X,\mathcal{D}}^\sigma$, as this leads to

$$\|f - \mathcal{I}_{X,\Phi_\delta}(f)\|_{L^2(\mathcal{D})} \leq C \left(\frac{1}{\nu}\right)^\sigma \|f\|_{H^\sigma(\mathcal{D})}.$$

So again, while we can now work with better conditioned matrices, we cannot expect any error reduction and therefore no suitable approximations. The key to overcome this issue lies in the idea to consider a sequence $(X_j)_{j\in\mathbb{N}} \subseteq \mathcal{D}$ of point sets, where we choose scaling parameters $\delta_j = \nu h_{X_j,\mathcal{D}}$. On each *level* $L \in \mathbb{N}$, we combine recursively the interpolants $\mathcal{I}_{X_j,\Phi_{\delta_j}}$ in a suitable way. This leads to approximation operators $A_L(f)$ for all $L \in \mathbb{N}$ such that $(A_L(f))_{L\in\mathbb{N}}$ converges to f in a suitable space. This is the main idea behind the kernel-based *multilevel method* (see [57],[16]), which we now discuss in greater detail.

6.2 Multilevel method – setup

In the following, we formalize the recursive idea of the multilevel method, provide a convergence result and afterwards analyze various properties and representations of the scheme.
Although we will present the following scheme in the context of kernel-based interpolation, we want to note that, in principle, different approximation methods could be used as well. We consider the following setting:

Setting 6.2.1.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and $(X_j)_{j\in\mathbb{N}} \subseteq \mathcal{D}$ a sequence of point sets consisting of $N_j \in \mathbb{N}$ points with mesh norms $h_j := h_{X_j,\mathcal{D}}$.
Additionally, suppose $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ is a positive definite function with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$. For $j \in \mathbb{N}$, we set $\delta_j := \nu h_j$, where $\nu \geq 1$ will be specified later on, and define $\Phi_j := \delta_j^{-d}\Phi\left(\frac{\cdot}{\delta_j}\right)$ as in Lemma 6.1.1.

The main idea in the following is a residual correction scheme: In the first step, we use the kernel Φ_1 to interpolate f in the points of X_1. This yields the approximation on the *first level*. After that, we consider the *residual* $f - \mathcal{I}_{X_1,\Phi_1}(f)$ and correct the error by interpolating this residual in the points of X_2 using Φ_2. The approximation on the *second level* is now given by the one on the first level added to the residual correction. This scheme now can be applied iteratively, and leads to the following algorithm:

Algorithm 1 Multilevel method

Input: $f_0 := 0,\ e_0 := f$
for $j = 1, 2, \ldots$ **do**
  $s_j \leftarrow \mathcal{I}_{X_j,\Phi_j}(e_{j-1})$
  $f_j \leftarrow f_{j-1} + s_j$
  $e_j \leftarrow e_{j-1} - s_j$
end for

Here, e_j denotes the residual, f_j the multilevel approximation after j levels and s_j the kernel approximation on the j-th level using the residual of the preceding level. When implementing this scheme, we only use a finite number $L \in \mathbb{N}$ of levels and then denote the emerging multilevel approximant by f_L.

Definition and Remark 6.2.2.
Assume Setting 6.2.1 holds and let $L \in \mathbb{N}$. The *multilevel operator* with L levels using Algorithm 1 is defined as

$$A_L \colon C(\mathcal{D}) \to L^2(\mathcal{D}), \quad A_L(f) := f_L.$$

The operator has the following properties:

1. $A_L(f) = f_L = \sum_{j=1}^{L} s_j = A_{L-1}(f) + \mathcal{I}_{X_L,\Phi_L}(f - A_{L-1}(f))$,
2. $e_L = f - A_L(f)$,
3. A_L is linear, and
4. $A_L(f)(X_L) = f(X_L)$, i.e., the multilevel operator interpolates f on the highest level X_L.

Proof. Inductively, one can see that

$$A_L(f) = f_L = f_{L-1} + s_L = f_{L-2} + s_{L-1} + s_L = \cdots = f_0 + \sum_{j=1}^{L} s_j = \sum_{j=1}^{L} s_j,$$

which proves the first representation of $A_L(f)$. With the same argument, one can also see that

$$e_L = e_{L-1} - s_L = \cdots = e_0 - \sum_{j=1}^{N} s_j = f - f_L,$$

implying the second statement and the second representation in the first statement, as

$$A_L(f) = f_{L-1} + s_L = A_{L-1}(f) + \mathcal{I}_{X_L,\Phi_L}(e_{L-1}).$$

The linearity of A_L is clear, as all occurring interpolation operators and the residual are linear. If we now rewrite e_L, we see that

$$f - A_L(f) = e_L = e_{L-1} - \mathcal{I}_{X_L,\Phi_L}(e_{L-1}),$$

and therefore $e_L(X_L) = 0$. □

In order to calculate the multilevel approximant of a function f, we therefore only use the information of f evaluated at all points in $\bigcup_{j=1}^{L} X_j$. Moreover, we can equivalently calculate the coefficients $\boldsymbol{\alpha}_j \in \mathbb{R}^{N_j}$ of $s_j = \sum_{k=1}^{N_j} \boldsymbol{\alpha}_k^{(j)} \Phi_j(\cdot - x_k^{(j)})$ using the following linear equation system (see [57]):

$$\begin{pmatrix} \Phi_1(X_1,X_1) & 0 & \dots & 0 \\ \Phi_2(X_2,X_1) & \Phi_2(X_2,X_2) & \dots & 0 \\ \vdots & & \ddots & 0 \\ \Phi_L(X_L,X_1) & \Phi_L(X_L,X_2) & \dots & \Phi_L(X_L,X_L) \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_L \end{pmatrix} = \begin{pmatrix} f(X_1) \\ f(X_2) \\ \vdots \\ f(X_L) \end{pmatrix}, \tag{6.1}$$

where the matrix is invertible as a lower block diagonal matrix with invertible diagonal blocks. In particular, the multilevel method given in Algorithm 1 can be interpreted as block-wise backwards substitution of this linear equation system. If we now assume that the sequence of point sets is quasi-uniform, we can combine this with Theorem 6.1.2 to obtain well-conditioned kernel matrices $\Phi_j(X_j, X_j)$, i.e., their condition number is bounded independently of the level $j \in \mathbb{N}$.

Corollary 6.2.3.
Assume Setting 6.2.1 holds, where the sequence of point sets $(X_j)_{j\in\mathbb{N}}$ is additionally quasi-uniform. Then for each $j \in \mathbb{N}$, the condition number of the kernel matrices $\Phi_j(X_j, X_j)$ is bounded by

$$\operatorname{cond}_2(\Phi_j(X_j, X_j)) \leq C \frac{\Phi(0)\,(1+\nu c_{qu})^d}{(\nu c_{qu})^{d-2\sigma}}.$$

Proof. Choose, for $j \in \mathbb{N}$, the constant $c^* = c_{qu}$ in Theorem 6.1.2. □

6.3 Convergence of the multilevel method

We now provide an error estimate for the multilevel method (cf. [57, Theorem 1]).

Theorem 6.3.1 (Multilevel method)**.**
Assume Setting 6.2.1 holds and the point sets $(X_j)_{j\in\mathbb{N}}$ are refined in such a way that there are constants $\mu \in]0,1[$ and $c \in]0,1]$ such that for every $j \in \mathbb{N}$, the mesh norms $(h_j)_{j\in\mathbb{N}}$ satisfy

$$c\mu h_j \leq h_{j+1} \leq \mu h_j, \quad j \in \mathbb{N},$$

where $0 < h_1 < 1$ is small enough to fulfill the assumptions made in Theorem 3.3.4. Furthermore, choose $\nu \geq 1$ such that $\nu h_1 \leq 1$. Then there exist two constants $C > 0$ and $C_1 = C_1(\mathcal{D}, \sigma, \Phi) > 0$ with $\alpha := C_1 \left(\nu^{-2\sigma} + \mu^{2\sigma}\right)^{\frac{1}{2}}$ such that for every $f \in H^\sigma(\mathcal{D})$ the estimate

$$\|f - A_L(f)\|_{L^2(\mathcal{D})} \leq C\alpha^L \|f\|_{H^\sigma(\mathcal{D})}$$

holds. Thus, the multilevel method converges linearly in $L^2(\mathcal{D})$ if $\alpha < 1$.

In order to prove this theorem, we will use the following lemma, which provides a link between the residuals e_j and e_{j-1}.

Lemma 6.3.2.
In the setting of Theorem 6.3.1, for all $j \in \mathbb{N}$, the estimate

$$\|Ee_j\|_{H_{\Phi_{j+1}}(\mathbb{R}^d)} \leq C_1 \left(\left(\frac{1}{\nu}\right)^{2\sigma} + \mu^{2\sigma} \right)^{\frac{1}{2}} \|Ee_{j-1}\|_{H_{\Phi_j}(\mathbb{R}^d)}$$

holds, where E denotes the extension operator given in Theorem 2.2.9.

Proof. We first observe that

$$\delta_{j+1} = \nu h_{j+1} \leq \nu\mu h_j \leq \nu h_j = \delta_j \leq \cdots \leq \delta_1 \leq 1,$$

i.e., Lemma 6.1.1 is applicable. Together with Theorem 3.1.7, this yields the estimate

$$\begin{aligned}
\|Ee_j\|^2_{H_{\Phi_{j+1}}(\mathbb{R}^d)} &= (2\pi)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \frac{\left|\widehat{Ee_j}(\omega)\right|^2}{\widehat{\Phi_{j+1}}(\omega)}\, d\omega \\
&\leq \frac{1}{c_1(2\pi)^{\frac{d}{2}}} \int_{\mathbb{R}^d} \left|\widehat{Ee_j}(\omega)\right|^2 \left(1 + \delta_{j+1}^2 \|\omega\|_2^2\right)^\sigma \, d\omega \\
&=: \frac{1}{c_1(2\pi)^{\frac{d}{2}}} \left(I_1 + I_2\right),
\end{aligned}$$

where

$$I_1 = \int_{\|\omega\|_2 \leq \frac{1}{\delta_{j+1}}} \left|\widehat{Ee_j}(\omega)\right|^2 \left(1 + \delta_{j+1}^2 \|\omega\|_2^2\right)^\sigma \, d\omega,$$
$$I_2 = \int_{\|\omega\|_2 > \frac{1}{\delta_{j+1}}} \left|\widehat{Ee_j}(\omega)\right|^2 \left(1 + \delta_{j+1}^2 \|\omega\|_2^2\right)^\sigma \, d\omega.$$

We now want to bound both terms separately. For I_1, we use $\delta_{j+1}\|\omega\|_2 \leq 1$ to obtain

$$\begin{aligned} I_1 &\leq \int_{\|\omega\|_2 \leq \frac{1}{\delta_{j+1}}} \left|\widehat{Ee_j}(\omega)\right|^2 (1+1)^\sigma \, d\omega \\ &\leq 2^\sigma \int_{\mathbb{R}^d} \left|\widehat{Ee_j}(\omega)\right|^2 \, d\omega \\ &= 2^\sigma \|Ee_j\|_{L^2(\mathbb{R}^d)}^2, \end{aligned}$$

using the Plancherel theorem in the last step. We now want to apply the sampling inequality of Corollary 3.3.3, as $e_j(X_j) = 0$. For this, we first use the boundedness of the extension operator E to estimate

$$I_1 \leq 2^\sigma \|Ee_j\|_{L^2(\mathbb{R}^d)}^2 \leq 2^\sigma C_E^2 \|e_j\|_{L^2(\mathcal{D})}^2 \leq 2^\sigma C_E^2 C_S^2 h_j^{2\sigma} \|e_j\|_{H^\sigma(\mathcal{D})}^2,$$

where C_E denotes the constant from the extension operator and C_S the one of the sampling inequality. The norm of e_j can now be bounded using the recursive representation to obtain

$$\begin{aligned} \|e_j\|_{H^\sigma(\mathcal{D})} &= \|e_{j-1} - \mathcal{I}_{X_j,\Phi_j}(e_{j-1})\|_{H^\sigma(\mathcal{D})} \\ &= \|Ee_{j-1} - \mathcal{I}_{X_j,\Phi_j}(Ee_{j-1})\|_{H^\sigma(\mathcal{D})} \\ &\leq \|Ee_{j-1} - \mathcal{I}_{X_j,\Phi_j}(Ee_{j-1})\|_{H^\sigma(\mathbb{R}^d)} \\ &\leq \sqrt{c_2}(2\pi)^{\frac{d}{4}} \delta_j^{-\sigma} \|Ee_{j-1} - \mathcal{I}_{X_j,\Phi_j} Ee_{j-1}\|_{H_{\Phi_j}(\mathbb{R}^d)} \\ &\leq \sqrt{c_2}(2\pi)^{\frac{d}{4}} \delta_j^{-\sigma} \|Ee_{j-1}\|_{H_{\Phi_j}(\mathbb{R}^d)}, \end{aligned} \tag{6.2}$$

i.e., the estimate for I_1 is given by

$$I_1 \leq 2^\sigma C_E^2 C_S^2 c_2 (2\pi)^{\frac{d}{2}} \frac{h_j^{2\sigma}}{\delta_j^{2\sigma}} \|Ee_{j-1}\|_{H_{\Phi_j}(\mathbb{R}^d)}^2. \tag{6.3}$$

For I_2, we use $\delta_j\|\omega\|_2 > 1$ to see that

$$\left(1 + \delta_{j+1}^2\|\omega\|_2^2\right)^\sigma \leq \left(2\delta_{j+1}^2\|\omega\|_2^2\right)^\sigma = 2^\sigma \delta_{j+1}^{2\sigma}\|\omega\|_2^{2\sigma} \leq 2^\sigma \delta_{j+1}^{2\sigma} \left(1 + \|\omega\|_2^2\right)^\sigma.$$

With this we can bound I_2 in the following manner:

$$\begin{aligned} I_2 &\leq 2^\sigma \delta_{j+1}^{2\sigma} \int_{\mathbb{R}^d} \left|\widehat{Ee_j}(\omega)\right|^2 \left(1 + \|\omega\|_2^2\right)^\sigma \, d\omega \\ &= 2^\sigma \delta_{j+1}^{2\sigma} \|Ee_j\|_{H^\sigma(\mathbb{R}^d)}^2 \\ &\leq 2^\sigma C_E^2 \delta_{j+1}^{2\sigma} \|e_j\|_{H^\sigma(\mathcal{D})}^2. \end{aligned}$$

The occurring norm of e_j can again be estimated by using (6.2), which leads to

$$I_2 \leq 2^\sigma C_E^2 c_2 (2\pi)^{\frac{d}{2}} \frac{\delta_{j+1}^{2\sigma}}{\delta_j^{2\sigma}} \|Ee_{j-1}\|^2_{H_{\Phi_j}(\mathbb{R}^d)}. \tag{6.4}$$

If we now collect both estimates (6.3) and (6.4), we obtain

$$\begin{aligned}
\|Ee_j\|^2_{H_{\Phi_{j+1}}(\mathbb{R}^d)} &\leq \frac{1}{c_1}(I_1 + I_2) \\
&\leq \frac{1}{c_1 (2\pi)^{\frac{d}{2}}} \left(2^\sigma C_E^2 C_S^2 c_2 (2\pi)^{\frac{d}{2}} \frac{h_j^{2\sigma}}{\delta_j^{2\sigma}} + 2^\sigma C_E^2 c_2 (2\pi)^{\frac{d}{2}} \frac{\delta_{j+1}^{2\sigma}}{\delta_j^{2\sigma}} \right) \|Ee_{j-1}\|^2_{H_{\Phi_j}(\mathbb{R}^d)} \\
&\leq C_1^2 \left(\left(\frac{h_j}{\delta_j} \right)^{2\sigma} + \left(\frac{\delta_{j+1}}{\delta_j} \right)^{2\sigma} \right) \|Ee_{j-1}\|^2_{H_{\Phi_j}(\mathbb{R}^d)},
\end{aligned}$$

where

$$C_1 := \frac{\sqrt{c_2}}{\sqrt{c_1}} 2^{\frac{\sigma}{2}} C_E \max\{C_S, 1\}. \tag{6.5}$$

As

$$\frac{h_j}{\delta_j} = \frac{1}{\nu}, \quad \frac{\delta_{j+1}}{\delta_j} = \frac{h_{j+1}}{h_j} \leq \mu, \quad j \in \mathbb{N},$$

the claim follows. □

We can now use this result to prove Theorem 6.3.1:

Proof. As $A_L(f)$ interpolates f in X_L, we can use Corollary 3.3.3 and the norm equivalence given in Lemma 6.1.1 to obtain

$$\begin{aligned}
\|f - A_L(f)\|_{L^2(\mathcal{D})} &\leq C_S h_L^\sigma \|e_L\|_{H^\sigma(\mathcal{D})} \\
&= C_S h_L^\sigma \|Ee_{L-1} - \mathcal{I}_{X_L,\Phi_L}(Ee_{L-1})\|_{H^\sigma(\mathcal{D})} \\
&\leq C_S \sqrt{c_2} \left(\frac{h_L}{\delta_L} \right)^\sigma \|Ee_{L-1} - \mathcal{I}_{X_L,\Phi_L}(Ee_{L-1})\|_{H_{\Phi_L}(\mathbb{R}^d)} \\
&\leq \alpha \|Ee_{L-1}\|_{H_{\Phi_L}(\mathbb{R}^d)}.
\end{aligned}$$

We note that the last estimate holds, as the constant C_1 obtained in (6.5) of Lemma 6.3.2 fulfills

$$C_S \sqrt{c_2} \leq \frac{\sqrt{c_2}}{\sqrt{c_1}} 2^{\frac{\sigma}{2}} C_E \max\{C_S, 1\} = C_1.$$

Next, we can apply Lemma 6.3.2 recursively to obtain

$$\begin{aligned}
\|f - A_L(f)\|_{L^2(\mathcal{D})} &\leq \alpha \|Ee_{L-1}\|_{H_{\Phi_L}(\mathbb{R}^d)} \leq \alpha^2 \|Ee_{L-2}\|_{H_{\Phi_{L-1}}(\mathbb{R}^d)} \\
&\leq \cdots \leq \alpha^L \|Ee_0\|_{H_{\Phi_1}(\mathbb{R}^d)} \leq \frac{C_E (2\pi)^{-\frac{d}{4}}}{\sqrt{c_1}} \alpha^L \|f\|_{H^\sigma(\mathcal{D})},
\end{aligned}$$

where the step follows by Lemma 6.1.1 and the boundedness of the extension operator. This yields the claim. □

Remark 6.3.3.
We emphasize the condition that in order to obtain convergence of the method, we have to assume that $\alpha = C_1 \left(\left(\frac{1}{\nu}\right)^{2\sigma} + \mu^{2\sigma} \right)^{\frac{1}{2}} < 1$, which can be achieved by choosing μ sufficiently small and ν sufficiently large. Unfortunately, the constant C_1 is not known explicitly and depends on various other (and not explicitly known) constants, for example the equivalence constants c_1 and c_2 of $H_\Phi(\mathbb{R}^d)$ and $H^\sigma(\mathbb{R}^d)$ or the operator norm of the extension operator E. Moreover, the bound suggests choosing $\frac{1}{\nu} \approx \mu$.
The theorem additionally implies that the multilevel operator $A_L \colon H^\sigma(\mathcal{D}) \to L^2(\mathcal{D})$ is bounded, as

$$\|A_L(f)\|_{L^2(\mathcal{D})} \leq \|A_L(f) - f\|_{L^2(\mathcal{D})} + \|f\|_{L^2(\mathcal{D})} \leq (C\alpha^L + 1)\|f\|_{H^\sigma(\mathcal{D})}, \quad f \in H^\sigma(\mathcal{D}).$$

Next, we want to prove that a pointwise estimate can also be obtained with the use of Lemma 6.3.2 and Corollary 3.3.3, but note that, in principle, also other estimates can be obtained by Theorem 3.3.2. In the pointwise estimate, we have to explicitly couple μ and ν.

Corollary 6.3.4.
Consider the setting of Theorem 6.3.1, where we additionally assume that there exists a constant $\gamma > 0$ such that

$$\frac{\gamma}{\nu} \leq \mu$$

holds. Then there are two constants $C > 0$ and $\widetilde{C}_1 = \widetilde{C}_1(\mathcal{D}, \sigma, \gamma)$ such that, for all $f \in H^\sigma(\mathcal{D})$, the estimate

$$\|f - A_L(f)\|_{C(\mathcal{D})} \leq C\beta^L \|f\|_{H^\sigma(\mathcal{D})}$$

holds, where β is given as

$$\beta = \widetilde{C}_1 \left(\left(\frac{1}{\nu}\right)^{2\sigma - d} + \mu^{2\sigma - d} \right)^{\frac{1}{2}}. \tag{6.6}$$

Thus, the multilevel method converges pointwise in $\mathcal{D}$ if $\beta < 1$.

Proof. We first observe that

$$\|f - A_L(f)\|_{C(\mathcal{D})} = \|Id_{\mathcal{D}}\,(Ee_L)\,\|_{C(\mathcal{D})} \leq \|Id\|_{H_{\Phi_{L+1}}(\mathbb{R}^d) \to C(\mathcal{D})} \|Ee_L\|_{H_{\Phi_{L+1}}(\mathbb{R}^d)}. \tag{6.7}$$

On the right-hand side we now can apply the recursive estimate given in Lemma 6.3.2, similarly to before. Therefore, it remains to bound the occurring operator norm of the inclusion. For this, let $g \in H_{\Phi_{L+1}(\mathbb{R}^d)}$ be fixed. Then, for all $x \in \mathcal{D}$, the estimate

$$|g(x)| = |\langle g, \Phi_{L+1}(\cdot - x)\rangle_{H_{\Phi_{L+1}(\mathbb{R}^d)}}| \leq \|g\|_{H_{\Phi_{L+1}(\mathbb{R}^d)}} \|\Phi(\cdot - x)\|_{H_{\Phi_{L+1}(\mathbb{R}^d)}}$$

holds. The latter can be directly evaluated by

$$\|\Phi(\cdot - x)\|^2_{H_{\Phi_{L+1}(\mathbb{R}^d)}} = \Phi_{L+1}(x - x) = \Phi_{L+1}(0) = \delta_{L+1}^{-d}\Phi(0),$$

for every $x \in \mathcal{D}$. Thus, the operator norm permits the bound

$$\|Id\|_{H_{\Phi_{L+1}}(\mathbb{R}^d)\to C(\mathcal{D})} \leq \delta_{L+1}^{-\frac{d}{2}} \sqrt{\Phi(0)}. \tag{6.8}$$

One can even prove equality by choosing $g = \Phi_{L+1}(\cdot - x^*)$ for fixed $x^* \in \mathcal{D}$. With this bound applied to (6.7), we obtain

$$\begin{aligned}
\|f - A_L(f)\|_{C(\mathcal{D})} &\leq \sqrt{\Phi(0)}\delta_{L+1}^{-\frac{d}{2}} \|Ee_L\|_{\Phi_{L+1}(\mathbb{R}^d)} \\
&\leq \cdots \leq \sqrt{\Phi(0)}\delta_{L+1}^{-\frac{d}{2}} \alpha^L \|Ee_0\|_{H_{\Phi_1}(\mathbb{R}^d)} \\
&\leq \frac{C_E \sqrt{\Phi(0)}(2\pi)^{-\frac{d}{4}}}{\sqrt{c_1}} \delta_{L+1}^{-\frac{d}{2}} \alpha^L \|f\|_{H^\sigma(\mathcal{D})}.
\end{aligned}$$

Next, we have to bound δ_{L+1} from below. This can be accomplished via

$$\delta_{L+1} = \nu h_{L+1} \geq \nu c \mu h_L \geq \cdots \geq \nu c^L \mu^L h_1 = \mu^L c^L \delta_1.$$

Together with the definition of α, we obtain

$$\begin{aligned}
\delta_{L+1}^{-\frac{d}{2}} \alpha^L &\leq \mu^{-\frac{d}{2}L} c^{-\frac{d}{2}L} \delta_1^{-\frac{d}{2}} \left(C_1 \left(\mu^{2\sigma} + \left(\frac{1}{\nu}\right)^2 \right)^{\frac{1}{2}} \right)^L \\
&\leq \delta_1^{-\frac{d}{2}} \left(C_1 c^{-\frac{d}{2}} \left(\nu^{-2\sigma}\mu^{-d} + \mu^{2\sigma-d} \right)^{\frac{1}{2}} \right)^L \\
&\leq \delta_1^{-\frac{d}{2}} \left(C_1 c^{-\frac{d}{2}} \max\left(1, \gamma^{-\frac{d}{2}}\right) \left(\left(\frac{1}{\nu}\right)^{2\sigma-d} + \mu^{2\sigma-d} \right)^{\frac{1}{2}} \right)^L.
\end{aligned}$$

Thus, we can choose $\widetilde{C_1} = \max\left(1, C_1 c^{-\frac{d}{2}} \max\left(1, \gamma^{-\frac{d}{2}}\right)\right)$ to prove the statement with

$$\|f - A_L(f)\|_{C(\mathcal{D})} \leq \frac{C_E \sqrt{\Phi(0)}(2\pi)^{-\frac{d}{4}}}{\sqrt{c_1}} \delta_{L+1}^{-\frac{d}{2}} \alpha^L \|f\|_{H^\sigma(\mathcal{D})} \leq \frac{C_E \sqrt{\Phi(0)}(2\pi)^{-\frac{d}{4}}}{\sqrt{c_1 \delta_1^d}} \beta^L \|f\|_{H^\sigma(\mathcal{D})}.$$

□

6.4 Properties of the multilevel scheme

6.4.1 Representations in the general case

Similar as in Section 5.2, we want to apply the multilevel method pathwise on a given second order random field. As we already have seen that the multilevel operator A_L is linear and bounded, we can use Lemma 5.1.2 to obtain a representation of the mean and covariance function of this random field, where the latter is given by the tensor product operator $A_L \overline{\otimes} A_L$ applied to the true covariance function. We want to give an

explicit representation of this tensor product operator. For this, similar to Lemma 3.5.1, a Lagrange type representation of the multilevel method,

$$A_L(f) = \sum_{j=1}^{N} f(x_j) b_j,$$

with some natural number $N \in \mathbb{N}$ and functions b_j, may be useful.
It turns out that such a representation can be obtained explicitly and moreover leads to structural insights about the method, in particular, a characterization of the approximation space. The following ideas have by now been published in [20]. Although most of these results are not directly necessary for the pathwise application of the multilevel method, we provide them in this section.
We again assume that Setting 6.2.1 holds throughout this section.
To characterize the approximation space, we observe that

$$A_L(f) = s_1 + \cdots + s_L, \quad f \in C(\mathcal{D}),$$

where $s_j = \mathcal{I}_{X_j,\Phi_j}(e_{j-1})$. This implies that every interpolant s_j is an element of the subspace $W_j := \operatorname{span}\{\Phi_j(\cdot - x) \mid x \in X_j\}$ and the whole approximation therefore lies in

$$A_L(f) \in V_L := W_1 + \cdots + W_L.$$

For specific kernels, we can additionally see that this sum is direct:

Corollary 6.4.1.
Let $X_1, \dots, X_L \subseteq \mathcal{D}$ be finite point sets with pairwise distinct mesh norms $h_1, \dots, h_L$ and $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ a Wendland kernel with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$. Let $\delta_j := \nu h_j$ with $\nu \geq 1$ and

$$\Phi_j := \delta_j^{-d} \Phi\left(\frac{\cdot}{\delta_j}\right), \quad 1 \leq j \leq L,$$

as in Theorem 6.3.1. Then the family of scaled and shifted kernel functions

$$\{\Phi_j(\cdot - x) \mid x \in X_j,\ 1 \leq j \leq L\}$$

is linearly independent. Therefore, $W_j \subseteq H^\sigma(\mathcal{D})$ is a finite dimensional, hence closed, subspace of $H^\sigma(\mathcal{D})$ and the sum

$$V_L = W_1 \oplus \cdots \oplus W_L$$

is direct.

Proof. The linear independence can be shown as in [19, Lemma 5.1]. The main idea is to use that the kernel functions $\Phi_j(\cdot - x_k)$ have finite regularity on $\partial B_{\delta_j}(x_k)$. If the kernel functions were linearly dependent, this finite regularity would vanish, which leads to a contradiction. The direct sum follows immediately. □

Although we write $A_L \colon C(\mathcal{D}) \to L^2(\mathcal{D})$ as a linear operator between two infinite dimensional spaces, we have seen that the image is a finite dimensional subspace of V_L. Additionally, this operator does not use the values of $f \in C(\mathcal{D})$ everywhere on $\mathcal{D}$ but only the pointwise information $f(X)$, where $X := \bigcup_{j=1}^{L} X_j$. If we write $N := |X|$, we can decompose the multilevel operator $A_L = T \circ \mathcal{A}_L$ where

- $T \colon C(\mathcal{D}) \to \mathbb{R}^N, f \mapsto f(X)$ is the *sampling* operator;
- $\mathcal{A}_L \colon \mathbb{R}^N \to V_L$ is the *synthesis operator* such that $A_L = \mathcal{A}_L \circ T$.

This allows us to characterize the approximation space of A_L by using the operator $\mathcal{A}_L$, which is a linear and bounded operator between *finite* dimensional spaces. Hence, we can use linear algebra to obtain the following theorem.

Theorem 6.4.2.
In the setting of Theorem 6.3.1, we define $X := \bigcup_{j=1}^{L} X_j$ and $N := |X|$. Then, we can decompose the multilevel operator A_L using the sampling operator T and the synthesis operator $\mathcal{A}_L$ as $A_L = \mathcal{A}_L \circ T$. Moreover, T is surjective, i.e., $A_L(C(\mathcal{D})) = \mathcal{A}_L(\mathbb{R}^N)$. If we additionally assume that the sum $V_L = W_1 \oplus \cdots \oplus W_L$ is direct, $\mathcal{A}_L$ is injective. Hence, $\dim(A_L(C(\mathcal{D}))) = N$ *and a basis of the approximation space is given by*

$$\{\mathcal{A}_L(\boldsymbol{v}_j) \mid 1 \leq j \leq N\},$$

where $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$ denotes an arbitrary basis of $\mathbb{R}^N$.

Proof. The representation of A_L and the surjectivity of T are clear. In order to prove the injectivity of $\mathcal{A}_L$ we use that $\mathcal{A}_L$ is linear, since A_L is linear. Therefore, it is sufficient to prove that $\boldsymbol{v} = \mathbf{0} \in \mathbb{R}^N$ is the only solution to $\mathcal{A}_L(\boldsymbol{v}) = 0$. We prove the statement by induction over L. For $L = 1$, we see that

$$0 = \mathcal{A}_L(\boldsymbol{v}) = \Phi_1(\cdot, X)\Phi(X, X)^{-1}\boldsymbol{v},$$

which can only be true if $\boldsymbol{v} = \boldsymbol{v}$, as $\Phi(X, X)$ is invertible. Now, assume the statement is true for $L - 1$. For convenience of notation, we choose a function $f \in C(\mathcal{D})$ such that $f(X) = \boldsymbol{v}$ and use the definition of A_L. Since $\mathcal{A}_L(\boldsymbol{v}) = 0$, we also obtain that $A_L(f) = \mathcal{A}_L(\boldsymbol{v}) = 0$. Using the recursive representation of the method yields

$$0 = A_L(f) = s_1 + \cdots + s_{L-1} + s_L = A_{L-1}(f) + s_L.$$

As $A_{L-1}(f) \in W_1 \oplus \cdots \oplus W_{L-1}$ and $s_L \in W_L$, this can only be true if both $A_{L-1}(f) = 0$ and $s_L = 0$. By induction hypothesis, the former implies that $f(x) = 0$ for all $x \in \bigcup_{j=1}^{L-1} X_j$, while the latter implies, similarly to the first step that $f(x) = 0$ for $x \in X_L$. Combining both statments implies that $\mathbf{0} = f(X) = \boldsymbol{v}$ and proving injectivity of $\mathcal{A}_L$.
The rest of thr statement follows by linear algebra: Since T is surjective, we can conclude

$$A_L(C(\mathcal{D})) = \mathcal{A}_L\left(T\left(C(\mathcal{D})\right)\right) = \mathcal{A}_L(\mathbb{R}^N).$$

With injectivity proven before, $\mathcal{A}_L \colon \mathbb{R}^N \to \mathcal{A}_L(\mathbb{R}^N)$ is bijective, which implies that $\dim(\mathcal{A}_L(\mathbb{R}^N)) = N$. As $\mathcal{A}_L$ is additionally linear, a basis of the space $\mathcal{A}_L(\mathbb{R}^N)$ is given by $\{\mathcal{A}_L(\boldsymbol{v}_j) \mid 1 \leq j \leq N\}$, where $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$ is a basis of $\mathbb{R}^N$. □

We mark down direct consequences of this result.

Remark 6.4.3.

1. Although the result is formulated for the domain of definition $C(\mathcal{D})$, it of course stays valid for all subsets $U \subseteq C(\mathcal{D})$, as long as we can guarantee that for all $\boldsymbol{v} \in \mathbb{R}^N$, there is a pointwise defined function $f \in U$ with $f(X) = \boldsymbol{v}$.

2. For a specific base $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_N)$ of $\mathbb{R}^N$, one can compute the multilevel approximation by calculating the basis representation of $f(X) \in \mathbb{R}^N$ in this basis, i.e., by writing
$$f(X) = \sum_{j=1}^{N} \alpha_j \boldsymbol{v}_j$$
and then using this representation and linearity to obtain
$$A_L(f) = \mathcal{A}_L(f(X)) = \mathcal{A}_L\left(\sum_{j=1}^{N} \alpha_j \boldsymbol{v}_j\right) = \sum_{j=1}^{N} \alpha_j \mathcal{A}_L(\boldsymbol{v}_j).$$
More specifically, when using the standard basis $\boldsymbol{e}_j, 1 \leq j \leq N$, together with $X = \{x_1, \dots, x_N\}$, the multilevel scheme is given by
$$A_L(f) = \sum_{j=1}^{N} f(x_j)\mathcal{A}_L(\boldsymbol{e}_j) = \sum_{j=1}^{N} f(x_j) b_j, \tag{6.9}$$
which provides a Lagrange-type representation of the multilevel scheme with basis $b_j := \mathcal{A}_L(\boldsymbol{e}_j)$, as we motivated in the beginning. Note that the representation also remains true if we drop the assumption $V_N = W_1 \oplus \dots \oplus W_L$. However, in this case, we can not ensure that $\{b_1, \dots, b_N\}$ is a basis of the approximation space. Moreover, such a representation can easily be generalized to all linear approximation operator that only use a finite number of point evaluation of the target function.

3. If the point sets $X_1 \subseteq X_2 \subseteq \dots \subseteq X_L$ are nested, then $\dim(A_L(C(\mathcal{D}))) = |X_L|$ and the basis $(\mathcal{A}_L(\boldsymbol{e}_j))_{1 \leq j \leq N}$ additionally consists of cardinal functions on X_L. We will discuss this case in greater detail in the next subsection.

4. If, on the other hand, all X_j are disjoint, the dimension of the approximation space fulfills $\dim(A_L(C(\mathcal{D}))) = \sum_{j=1}^{L} |X_j| = \dim(V_L)$, i.e., in this case we can conclude that the multilevel operator is actually bijective onto this space.

Next, we want to analyze the scheme itself in more detail and therefore express the dependence on the interpolation operators $\mathcal{I}_{X_j,\Phi_j}$ more explicitly. For this, we use the following lemma:

Lemma 6.4.4.
Let V be a linear space, and $\mathfrak{v} = \{v_1, \dots, v_{|\mathfrak{v}|}\} \subseteq \mathbb{N}$ with $v_1 < v_2 < \dots < v_{|\mathfrak{v}|}$ be an ordered index set. Assume, for every $j \in \mathfrak{v}$, there are operators $A_j, B_j \colon V \to V$. Additionally suppose that for all $j, k \in \mathfrak{v}$, the operator A_j and B_k are commutative, i.e., $A_j B_k = B_k A_j$. Then the following representation holds:

$$\prod_{j\in\mathfrak{v}}(A_j + B_j) = (A_{\mathfrak{v}_{|\mathfrak{v}|}} + B_{\mathfrak{v}_{|\mathfrak{v}|}}) \cdots (A_{\mathfrak{v}_1} + B_{\mathfrak{v}_1}) = \sum_{\mathfrak{u}\subseteq\mathfrak{v}} \prod_{j\in\mathfrak{u}} A_j \prod_{j\in\mathfrak{v}\setminus\mathfrak{u}} B_j,$$

where we assume that every subset $\emptyset \neq \mathfrak{u} = \{u_1, \dots, u_{|\mathfrak{u}|}\} \subseteq \mathfrak{v}$ is again increasingly ordered, i.e., $u_1 < u_2 < \dots u_{|\mathfrak{u}|}$, and the product

$$\prod_{j\in\mathfrak{u}} A_j = A_{u_{|\mathfrak{u}|}} A_{u_{|\mathfrak{u}|-1}} \dots A_{u_1}$$

denotes the decreasingly sorted product of the (maybe non-commutative) operators A_j for all $j \in \mathfrak{u}$. For $\mathfrak{u} = \emptyset$, the product is defined as the identity operator $Id_V \colon V \to V$.

Proof. Without loss of generality, we only prove the case $\mathfrak{v} = \{1, \dots, n\}$ for a natural number $n \in \mathbb{N}$ by induction. For $n = 1$, there is nothing to show. If the formula holds for $n \in \mathbb{N}$, then the product for $n+1$ can be rewritten as

$$\begin{aligned}
&\prod_{j=1}^{n+1}(A_j + B_j) = (A_{n+1} + B_{n+1}) \prod_{j=1}^{n}(A_j + B_j) \\
&= (A_{n+1} + B_{n+1}) \sum_{\mathfrak{u}\subseteq\{1,\dots,n\}} \prod_{j\in\mathfrak{u}} A_j \prod_{j\in\{1,\dots,n\}\setminus\mathfrak{u}} B_j \\
&= A_{n+1} \left(\sum_{\mathfrak{u}\subseteq\{1,\dots,n\}} \prod_{j\in\mathfrak{u}} A_j \prod_{j\in\{1,\dots,n\}\setminus\mathfrak{u}} B_j \right) + B_{n+1} \left(\sum_{\mathfrak{u}\subseteq\{1,\dots,n\}} \prod_{j\in\mathfrak{u}} A_j \prod_{j\in\{1,\dots,n\}\setminus\mathfrak{u}} B_j \right) \\
&= \sum_{\mathfrak{u}\subseteq\{1,\dots,n\}} \left(A_{n+1} \prod_{j\in\mathfrak{u}} A_j \right) \prod_{j\in\{1,\dots,n\}\setminus\mathfrak{u}} B_j + \sum_{\mathfrak{u}\subseteq\{1,\dots,n\}} \prod_{j\in\mathfrak{u}} A_j \left(B_{n+1} \prod_{j\in\{1,\dots,n\}\setminus\mathfrak{u}} B_j \right),
\end{aligned}$$

where we used the commutativity of B_{n+1} with $A_1, \dots, A_n$. We rewrite this as subsets $\mathfrak{u} \subseteq \{1, \dots, n+1\}$ by

$$\begin{aligned}
&\sum_{\mathfrak{u}\subseteq\{1,\dots,n\}} \left(A_{n+1} \prod_{j\in\mathfrak{u}} A_j \right) \prod_{j\in\{1,\dots,n\}\setminus\mathfrak{u}} B_j + \sum_{\mathfrak{u}\subseteq\{1,\dots,n\}} \prod_{j\in\mathfrak{u}} A_j \left(B_{n+1} \prod_{j\in\{1,\dots,n\}\setminus\mathfrak{u}} B_j \right) \\
&= \sum_{\substack{\mathfrak{u}\subseteq\{1,\dots,n+1\} \\ n+1\in\mathfrak{u}}} \prod_{j\in\mathfrak{u}} A_j \prod_{j\in\{1,\dots,n+1\}\setminus\mathfrak{u}} B_j + \sum_{\substack{\mathfrak{u}\subseteq\{1,\dots,n+1\} \\ n+1\notin\mathfrak{u}}} \prod_{j\in\mathfrak{u}} A_j \prod_{j\in\{1,\dots,n+1\}\setminus\mathfrak{u}} B_j \\
&= \sum_{\mathfrak{u}\subseteq\{1,\dots,n+1\}} \prod_{j\in\mathfrak{u}} A_j \prod_{j\in\{1,\dots,n+1\}\setminus\mathfrak{u}} B_j,
\end{aligned}$$

which yields the claim. □

We now use this result to obtain the following explicit representation of the multilevel scheme ([30, Theorem 3.8]).

Corollary 6.4.5.
In the setting of Theorem 6.3.1, we obtain the representations

1.
$$\begin{aligned} f - A_L(f) = e_L &= (Id_{\mathcal{D}} - \mathcal{I}_{X_L,\Phi_L}) \cdots (Id_{\mathcal{D}} - \mathcal{I}_{X_1,\Phi_1})(f) \\ &= \sum_{\mathfrak{u} \subseteq \{1,\dots,L\}} (-1)^{|\mathfrak{u}|} \prod_{j \in \mathfrak{u}} \mathcal{I}_{X_j,\Phi_j}(f), \end{aligned}$$

2.
$$A_L(f) = \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,\dots,L\}} (-1)^{|\mathfrak{u}|+1} \prod_{j \in \mathfrak{u}} \mathcal{I}_{X_j,\Phi_j}(f)$$

Proof. The first step is to observe that the error on the L-th level can inductively be written as

$$\begin{aligned} e_L &= e_{L-1} - \mathcal{I}_{X_L,\Phi_L} e_{L-1} = (Id_{\mathcal{D}} - \mathcal{I}_{X_L,\Phi_L}) e_{L-1} \\ &= \cdots = (Id_{\mathcal{D}} - \mathcal{I}_{X_L,\Phi_L}) \cdots (Id_{\mathcal{D}} - \mathcal{I}_{X_1,\Phi_1}) f, \end{aligned}$$

since $e_0 = f$. On this decreasing product we can apply Lemma 6.4.4, as $-Id_{\mathcal{D}}$ and $\mathcal{I}_{X_j,\Phi_j}$ commute, which leads to

$$\begin{aligned} e_L &= \prod_{j \in \{1,\dots,L\}} (Id - \mathcal{I}_{X_j,\Phi_j}) f = (-1)^L \prod_{j \in \{1,\dots,L\}} (\mathcal{I}_{X_j,\Phi_j} - Id) f \\ &= (-1)^L \sum_{\mathfrak{u} \subseteq \{1,\dots,L\}} \prod_{j \in \mathfrak{u}} \mathcal{I}_{X_j,\Phi_j} \prod_{j \in \{1,\dots,L\} \setminus \mathfrak{u}} (-Id) f \\ &= \sum_{\mathfrak{u} \subseteq \{1,\dots,L\}} (-1)^{2L-|\mathfrak{u}|} \prod_{j \in \mathfrak{u}} \mathcal{I}_{X_j,\Phi_j} f = \sum_{\mathfrak{u} \subseteq \{1,\dots,L\}} (-1)^{|\mathfrak{u}|} \prod_{j \in \mathfrak{u}} \mathcal{I}_{X_j,\Phi_j} f. \end{aligned}$$

For the second equation, we see that the term for $\mathfrak{u} = \emptyset$ in the representation of e_L just gives f, so we rewrite e_L as

$$f - A_L(f) = e_L = f + \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,\dots,L\}} (-1)^{|\mathfrak{u}|} \prod_{j \in \mathfrak{u}} \mathcal{I}_{X_j,\Phi_j} f.$$

By cancelling out f and multiplying by -1, the second statement follows. □

As already proposed earlier, we will now have a more detailed look at the basis functions of the scheme in the nested case.

6.4.2 Representation of the multilevel scheme in the nested case

Next, we combine the results from the last section to explicitly write down the basis in the nested case. For this, let $L \in \mathbb{N}$ and consider the point sets

$$X_1 \subseteq X_2 \subseteq \cdots \subseteq X_L \subseteq \mathcal{D}.$$

To simplify the results, we use the following notations:

- Every point set X_j is given as $X_j = \left\{x_1^{(j)}, x_2^{(j)}, \dots, x_{N_j}^{(j)}\right\}$ with $N_j \in \mathbb{N}$ for $1 \leq j \leq L$.
- Additionally, we assume that the first elements of X_j are given by the ones in X_{j-1}, i.e., we can write

$$X_j = \left\{x_1^{(j)}, x_2^{(j)}, \dots, x_{N_j}^{(j)}\right\} = X_{j-1} \dot\cup \left\{x_{N_{j-1}+1}^{(j)}, \dots, x_{N_j}^{(j)}\right\},$$

  which implies $x_k^{(j)} = x_k^{(\ell)}$ for every $1 \leq k \leq N_j$ and $\ell \geq j$. Hence, we suppress the upper index whenever there is no confusion, i.e., $X_L = \{x_1, \dots, x_{N_L}\}$. Additionally, for every $x \in X_L$, there exists a unique index $1 \leq k \leq L$ such that x is in $X_k \setminus X_{k-1}$, where $X_0 := \emptyset$.
- The cardinal functions in X_j using Φ_j are denoted by

$$\chi^{(j)} = \left(\chi_1^{(j)}, \dots, \chi_{N_j}^{(j)}\right),$$

  i.e., every component can be written as

$$\chi_k^{(j)} = \Phi_j(\cdot, X_j)\Phi_j(X_j, X_j)^{-1}\boldsymbol{e}_k,$$

  where $\boldsymbol{e}_k$ denotes the k-th unit vector in $\mathbb{R}^{N_j}$.

Motivated by the following observation, we introduce a different notation for the multilevel operator:

Proposition 6.4.6.
Consider the setting of Theorem 6.3.1, where the point sets $X_1 \subseteq \dots \subseteq X_L$ are additionally nested. Now assume there is an index $1 \leq j \leq L$ such that $f \in H^\sigma(\mathcal{D})$ vanishes on X_j. Then $s_1 = \dots = s_j = 0$, and the multilevel approximant to f is given by

$$A_L(f) = s_{j+1} + \dots + s_L.$$

Hence, we can interpret this approximation as the multilevel method applied to f, which starts at level $j+1$. To emphasize this property, we also write

$$A_{[j+1:L]}(f) := A_L(f)$$

in this case.

Proof. As $f(X_j) = 0$, one can conclude by the nestedness of the point sets that $f(X_k) = 0$ for $1 \leq k \leq j$ as well. When computing s_1, one can see that

$$s_1 = \mathcal{I}_{X_1,\Phi_1}(f) = \Phi_1(\cdot, X_1)\Phi_1(X_1, X_1)^{-1} f(X_1) = 0.$$

Inductively, if $s_{k-1} = 0$ for $2 \leq k \leq j$, then

$$s_k = \mathcal{I}_{X_k,\Phi_k}\left(f - A_{k-1}(f)\right) = \mathcal{I}_{X_k,\Phi_k}(f) = 0$$

holds as well, since $f(X_k) = 0$ by the same argument as before. □

Remark.

1. Comparing this to the matrix representation of the multilevel method given in (6.1), one can see that the setting of Proposition 6.4.6 leads to a *reduced* system of linear equations given as

$$\begin{pmatrix} \Phi_{j+1}(X_{j+1}, X_{j+1}) & & \\ \vdots & \ddots & \\ \Phi_L(X_L, X_{j+1}) & \dots & \Phi_L(X_L, X_L) \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_{j+1} \\ \vdots \\ \boldsymbol{\alpha}_L \end{pmatrix} = \begin{pmatrix} f(X_{j+1}) \\ \vdots \\ f(X_L) \end{pmatrix}.$$

2. Of course, one can easily extend the representation given in Corollary 6.4.5 to this case. The representation for $A_{[j:L]}(f)$ is for example given by

$$A_{[j:L]}(f) = \sum_{\emptyset \neq \mathfrak{u} \subseteq \{j,\dots,L\}} (-1)^{|\mathfrak{u}|+1} \prod_{j \in \mathfrak{u}} \mathcal{I}_{X_j, \Phi_j}(f). \tag{6.10}$$

We now want to analyze the basis functions given in (6.9) of A_L in more detail. For this, we want to give a slightly more illustrative motivation: As the operator $A_{[1:L]}(f)$ only depends on the function values of f at X_L, we can interchange f in the argument with every function $g \in C(\mathcal{D})$ that interpolates f in X_L, for example $g = \mathcal{I}_{X_L, \Phi_L}(f)$. Using the linearity of $A_{[1:L]}$ leads to the basis representation:

$$A_{[1:L]}(f) = A_{[1:L]}\left(\mathcal{I}_{X_L,\Phi_L}(f)\right) = A_{[1:L]}\left(\sum_{j=1}^{N_L} f(x_j)\chi_j^{(L)}\right) = \sum_{j=1}^{N_L} f(x_j) A_{[1:L]}\left(\chi_j^{(L)}\right).$$

In the following, we denote this basis by

$$b_j^{(L)} := A_{[1:L]}\left(\chi_j^{(L)}\right), \quad 1 \leq j \leq N_L. \tag{6.11}$$

Our goal is to prove different representations of the basis functions. For this, we have to investigate how kernel interpolants act on Lagrange functions as their argument.

Lemma 6.4.7.
Let $K_1, K_2 \colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ be two reproducing kernels. Assume two point sets

$$X_1 = \{x_1^{(1)}, \dots, x_{N_1}^{(1)}\} \subseteq X_2 = \{x_1^{(2)}, \dots, x_{N_2}^{(2)}\} \subseteq \mathcal{D}$$

are given, which are ordered as before, i.e., for all $1 \leq k \leq N_1$ we have $x_k^{(1)} = x_k^{(2)}$. Further denote the corresponding cardinal functions by

$$\chi^{(1)} = \left(\chi_1^{(1)}, \dots, \chi_{N_1}^{(1)}\right) = K_1(\cdot, X_1) K_1(X_1, X_1)^{-1},$$
$$\chi^{(2)} = \left(\chi_1^{(2)}, \dots, \chi_{N_2}^{(2)}\right) = K_2(\cdot, X_2) K_2(X_2, X_2)^{-1}.$$

Then the two following statements hold:

1. *The kernel interpolant of $\chi_k^{(2)}$ for $1 \leq k \leq N_2$ in X_1 using K_1 is given as*

$$\mathcal{I}_{X_1,K_1}\left(\chi_k^{(2)}\right) = \begin{cases} \chi_k^{(1)}, & 1 \leq k \leq N_1, \\ 0, & N_1 + 1 \leq k \leq N_2. \end{cases}$$

2. *The kernel interpolant of* $\chi_k^{(1)}$ *for* $1 \leq k \leq N_1$ *in* X_2 *using* K_2 *is given as*

$$\mathcal{I}_{X_2,K_2}\left(\chi_k^{(1)}\right) = \chi_k^{(2)} + \sum_{j=N_1+1}^{N_2} \chi_j^{(2)} \chi_k^{(1)}(x_j).$$

Proof. For the first case, let $1 \leq k \leq N_2$. Then, the interpolation can be written as

$$\mathcal{I}_{X_1,K_1}(\chi_k^{(2)}) = K_1(\cdot, X_1) K_1(X_1, X_1)^{-1} \chi_k^{(2)}(X_1) = \chi^{(1)} \chi_k^{(2)}(X_1).$$

If $1 \leq k \leq N_1$, the vector $\chi_k^{(2)}(X_1) = \boldsymbol{e}_k \in \mathbb{R}^{N_1}$ is the k-th unit vector. Thus, the above equation shows immediately that $\mathcal{I}_{X_1,K_1}(\chi_k^{(2)}) = \chi_k^{(1)}$. If otherwise $N_1 + 1 \leq k \leq N_2$, then $\chi_k^{(2)}(X_1) = 0 \in \mathbb{R}^{N_1}$, which proves the first claim.
For the second statement, let $1 \leq k \leq N_1$. Using the Lagrange property of $\chi_k^{(1)}$ on X_1, we can write $\chi_k^{(1)}(X_2)$ as

$$\chi_k^{(1)}(X_2) = (0, \ldots, 0, 1, 0, \ldots, 0, \chi_k^{(1)}(x_{N_1+1}), \ldots, \chi_k^{(1)}(x_{N_2}))^T,$$

where the 1 in the first part is at the k-th position. Inserting this representation into the interpolation formula yields

$$\mathcal{I}_{X_2,K_2}(\chi_k^{(1)}) = \chi^{(2)} \chi_k^{(1)}(X_2) = \chi_k^{(2)} + \sum_{j=N_1+1}^{N_2} \chi_j^{(2)} \chi_k^{(1)}(x_j),$$

which proves the second claim. □

This allows us to prove an alternative representation of $b_j^{(L)}$ for $1 \leq j \leq N_L$.

Theorem 6.4.8.
Assume Setting 6.2.1 with nested point sets $X_1 \subseteq X_2 \subseteq \cdots \subseteq X_L$ *holds. Define* $X_0 := \emptyset$ *and* $N_0 := 0$. *Let* $j \in \{1, \ldots, N_L\}$ *be arbitrary and denote with* $k \in \mathbb{N}$ *the unique index such that* $j \in \{N_{k-1} + 1, \ldots, N_k\}$ *(i.e.,* $x_j \in X_k \setminus X_{k-1}$*). Then the corresponding basis function* $b_j^{(L)}$ *can be written as*

$$b_j^{(L)} = A_{[k:L]}\left(\chi_j^{(L)}\right) = \chi_j^{(L)} + \sum_{\ell=k}^{L-1} \left(\chi_j^{(\ell)} - A_{[\ell+1:L]}\left(\chi_j^{(\ell)}\right)\right). \tag{6.12}$$

In particular, the basis functions $b_j^{(L)}$ *coincide with the corresponding Lagrange functions on level* L *when* $N_{L-1} \leq j \leq N_L$. *Moreover, the basis functions can be calculated recursively as*

$$\begin{aligned} b_j^{(L)} &= \begin{cases} b_j^{(L-1)} - \sum_{\ell=N_{L-1}+1}^{N_L} b_j^{(L-1)}(x_\ell)\, \chi_\ell^{(L)}, & 1 \leq j \leq N_{L-1}, \\ \chi_j^{(L)}, & N_{L-1}+1 \leq j \leq N_L, \end{cases} \\ &= \begin{cases} b_j^{(L-1)} - \mathcal{I}_{X_L,\Phi_L}\left(b_j^{(L-1)}\right) + \chi_j^{(L)}, & 1 \leq j \leq N_{L-1}, \\ \chi_j^{(L)}, & N_{L-1}+1 \leq j \leq N_L, \end{cases} \end{aligned} \tag{6.13}$$

where $b_j^{(L-1)} = A_{[1:L-1]}\left(\chi_j^{(L-1)}\right)$ *for* $1 \leq j \leq N_{L-1}$.

Proof. Let $1 \leq j \leq N_L$ be arbitrary, and choose $1 \leq k \leq L$ accordingly, as described in the statement. Using the representation (6.11) of $b_j^{(L)} = A_{[1:L]}\left(\chi_j^{(L)}\right)$ and $\chi_j^{(L)}(X_{k-1}) = 0$ implies by Proposition 6.4.6 that

$$A_{[1:L]}\left(\chi_j^{(L)}\right) = A_{[k:L]}\left(\chi_j^{(L)}\right). \tag{6.14}$$

We now plug in the explicit representation of $A_{[k:L]}\left(\chi_j^{(L)}\right)$ given in (6.10) to obtain

$$b_j^{(L)} = A_{[k:L]}\left(\chi_j^{(L)}\right) = \sum_{\emptyset \neq \mathfrak{u} \subseteq \{k,\dots,L\}} (-1)^{|\mathfrak{u}|+1} \prod_{i \in \mathfrak{u}} \mathcal{I}_{X_i,\Phi_i}\left(\chi_j^{(L)}\right).$$

The main idea in the following is to order the subsets $\emptyset \neq \mathfrak{u} \subseteq \{k,\dots,L\}$ by their smallest element $\ell \in \{k,\dots,L\}$, i.e., we write every subset $\emptyset \neq \mathfrak{u} \subseteq \{k,\dots,L\}$ as

$$\mathfrak{u} = \{\ell\} \dot\cup \widetilde{\mathfrak{u}}, \quad \text{where } \widetilde{\mathfrak{u}} \subseteq \{\ell+1,\dots,L\}.$$

Obviously, we then have $|\mathfrak{u}| = |\widetilde{\mathfrak{u}}| + 1$. For $\ell = L$, the only decomposition admissible is $\{L\} = \{L\} \cup \emptyset$. Separating this one from the others, we obtain

$$\begin{aligned} b_j^{(L)} &= \mathcal{I}_{X_L,\Phi_L}\left(\chi_j^{(L)}\right) + \sum_{\ell=k}^{L-1} \sum_{\widetilde{\mathfrak{u}} \subseteq \{\ell+1,\dots,L\}} (-1)^{|\widetilde{\mathfrak{u}}|} \left(\prod_{i \in \widetilde{\mathfrak{u}}} \mathcal{I}_{X_i,\Phi_i}\right) \mathcal{I}_{X_\ell,\Phi_\ell}\left(\chi_j^{(L)}\right) \\ &= \chi_j^{(L)} + \sum_{\ell=k}^{L-1} \sum_{\widetilde{\mathfrak{u}} \subseteq \{\ell+1,\dots,L\}} (-1)^{|\widetilde{\mathfrak{u}}|} \prod_{i \in \widetilde{\mathfrak{u}}} \mathcal{I}_{X_i,\Phi_i}\left(\chi_j^{(\ell)}\right), \end{aligned}$$

where, in the last step, we applied the first equation of Lemma 6.4.7. If we now fix $\ell \in \{k,\dots,L-1\}$, the inner expression can be represented using (6.10) as

$$\begin{aligned} &\sum_{\widetilde{\mathfrak{u}} \subseteq \{\ell+1,\dots,L\}} (-1)^{|\widetilde{\mathfrak{u}}|} \prod_{i \in \widetilde{\mathfrak{u}}} \mathcal{I}_{X_i,\Phi_i}\left(\chi_j^{(\ell)}\right) \\ &= \chi_j^{(\ell)} - \sum_{\emptyset \neq \widetilde{\mathfrak{u}} \subseteq \{\ell+1,\dots,L\}} (-1)^{|\widetilde{\mathfrak{u}}|+1} \prod_{i \in \widetilde{\mathfrak{u}}} \mathcal{I}_{X_i,\Phi_i}\left(\chi_j^{(\ell)}\right) \\ &= \chi_j^{(\ell)} - A_{[\ell+1:L]}\left(\chi_j^{(\ell)}\right), \end{aligned}$$

which all together yields the proclaimed expression. As $A_{[\ell+1:L]}\left(\chi_j^{(\ell)}\right)$ interpolates $\chi_j^{(\ell)}$ in X_L, the cardinality of $b_j^{(L)}$ follows immediately.
Next, we want to prove the recursive representations. This can be achieved using the recursion formula of A_L, which, for $1 \leq j \leq N_{L-1}$, leads to

$$\begin{aligned} b_j^{(L)} &= A_L\left(\chi_j^{(L)}\right) = A_{L-1}(\chi_j^{(L)}) + \mathcal{I}_{X_L,\Phi_L}\left(\chi_j^{(L)} - A_{L-1}\left(\chi_j^{(L)}\right)\right) \\ &= \sum_{\ell=1}^{N_{L-1}} \chi_j^{(L)}(x_\ell) b_\ell^{(L-1)} + \sum_{\ell=1}^{N_L} \chi_\ell^{(L)}\left(\chi_j^{(L)}(x_\ell) - A_{L-1}\left(\chi_j^{(L)}\right)(x_\ell)\right) \\ &= b_j^{(L-1)} - \sum_{\ell=N_{L-1}+1}^{N_L} \chi_\ell^{(L)} \sum_{m=1}^{N_{L-1}} \chi_j^{(L)}(x_m) b_m^{(L-1)}(x_\ell) \\ &= b_j^{(L-1)} - \sum_{\ell=N_{L-1}+1}^{N_L} \chi_\ell^{(L)} b_j^{(L-1)}(x_\ell). \end{aligned}$$

The representation for $N_{L-1} \leq j \leq N_L$ has already been shown beforehand.
Finally, using the cardinality of the basis function $b_j^{(L-1)}(x_\ell) = \delta_{j,\ell}$, the latter term can be reformulated to obtain

$$\sum_{\ell=N_{L-1}+1}^{N_L} b_j^{(L-1)}(x_\ell)\chi_\ell^{(L)} = \sum_{\ell=1}^{N_L} b_j^{(L-1)}(x_\ell)\chi_\ell^{(L)} - \chi_j^{(L)} = \mathcal{I}_{X_L,\Phi_L}\left(b_j^{(L-1)}\right) - \chi_j^{(L)},$$

proving the last statement. □

The first representation (6.12) allows us to interpret the multilevel scheme as a 'correction' algorithm for the highest level kernel interpolant, while we respect the interpolation property on X_L.
From a numerical point of view, we can use the representation (6.14) to see that

$$b_j^{(L)} = A_{[k:L]}\left(\chi_j^{(L)}\right), \quad N_{k-1}+1 \leq j \leq N_k,$$

i.e., if we additionally track the first level k in which x_j occurs we can speed up the calculation. Moreover, as seen in Theorem 6.4.2, we do not need to actually compute the cardinal functions $\chi_j^{(L)}$, but rather use the fact that $\chi_j^{(L)}(X_L) = \boldsymbol{e}_j \in \mathbb{R}^{N_L}$, which again speeds up the computation time.

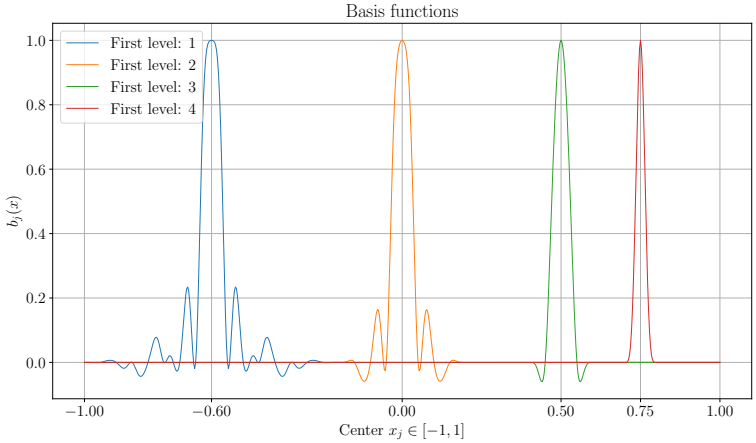


Figure 6.1: Basis functions of a multilevel method using $L = 4$ levels, starting with 5 equidistant points in $[-1, 1]$ and adding the middle point of each interval $[x_{j-1}, x_j]$ per level, leading to total of 41 equidistant points in level 4. The basis functions are ordered from left to right by the first appearance of the center $x_j \in X_k \setminus X_{k-1}$ of the basis function $b_j^{(4)}$ for $1 \leq j, k \leq 4$.

As Figure 6.1 already suggests, we can prove that the basis functions $b_j^{(L)}$ decay exponentially with rate $\|x - x_j\|/q_k$, where $1 \leq k \leq L$ is chosen such that $x_j \in X_k \setminus X_{k-1}$.

Theorem 6.4.9.
Assume the setting of Theorem 6.4.8 holds, where additionally the point sets $(X_j)_{j\in\mathbb{N}}$ are quasi-uniform.
Then there exists a constant $\eta > 0$ and another constant $C_L > 0$ depending on the current level L such that for all $1 \le k \le L$ and $N_{k-1} + 1 \le j \le N_k$, i.e., $x_j \in X_k \setminus X_{k-1}$, the following estimate holds:

$$\left|b_j^{(L)}(x)\right| \le C_L e^{-\frac{\eta}{q_k}\|x-x_j\|_2}, \quad x \in \mathcal{D}.$$

This means that the basis functions decay exponentially from their center point x_j.

Proof. The proof mostly depends on the exponential decay of the Lagrange functions shown in [13, Theorem 2.3], which states that there exist two constants $C_1 > 0$ and $\eta > 0$ such that, for all $1 \le k \le L$, the estimate

$$\left|\chi_j^{(k)}(x)\right| \le C_1 e^{-\frac{\eta}{q_k}\|x-x_j\|_2}, \quad 1 \le j \le N_k, \quad x \in \mathcal{D} \tag{6.15}$$

holds. We note that the constant C_1 can be chosen the same for all different levels $1 \le k \le L$. The core idea in the following is to combine this estimate (6.15) with the recursive representation (6.13) of the basis functions given in Theorem 6.4.8, by induction. For $L = 1$, we immediately get the claim, as the basis functions are given as Lagrange functions. If the statement for $L - 1$ holds, we proceed as follows: Let $1 \le k \le L$ and $N_{k-1} + 1 \le j \le N_k$ be arbitrary. For $k = L$, i.e., $x_j \in X_L \setminus X_{L-1}$, the statement also follows directly by (6.15). Hence, we consider $k < L$. Here, we can use the recursive representation

$$b_j^{(L)}(x) = b_j^{(L-1)}(x) - \sum_{\ell=N_{L-1}+1}^{N_L} b_j^{(L-1)}(x_\ell)\chi_\ell^{(L)}(x),$$

where for fixed $\ell \in \{N_{L-1} + 1, \dots, N_L\}$ and $x \in \mathbb{R}^d$ we can use the induction hypothesis and (6.15) to obtain the bound

$$\begin{aligned}
\left|b_j^{(L-1)}(x_\ell)\chi_\ell^{(L)}(x)\right| &\le C_{L-1} e^{-\frac{\eta}{q_k}\|x_j-x_\ell\|_2} C_1 e^{-\frac{\eta}{q_L}\|x_\ell-x\|_2} \\
&= C_{L-1}C_1 e^{-\frac{\eta}{q_k}(\|x_j-x_\ell\|_2+\|x_\ell-x\|_2)} e^{-\eta\|x_\ell-x\|_2\left(\frac{1}{q_L}-\frac{1}{q_k}\right)} \\
&\le C_{L-1}C_1 e^{-\frac{\eta}{q_k}\|x-x_j\|_2} e^{-\eta\|x_\ell-x\|_2\left(\frac{1}{q_L}-\frac{1}{q_k}\right)}.
\end{aligned}$$

With this estimate, we can bound the sum by

$$\sum_{\ell=N_{L-1}+1}^{N_L} \left|b_j^{(L-1)}(x_\ell)\chi_\ell^{(L)}(x)\right| \le C_1 C_{L-1} e^{-\frac{\eta}{q_k}\|x-x_j\|_2} \sum_{\ell=N_{L-1}+1}^{N_L} e^{-\eta\|x_\ell-x\|_2\left(\frac{1}{q_L}-\frac{1}{q_k}\right)}.$$

Next, we want to prove that the latter sum is bounded independently of x. Therefore, we proceed as in the proof of [13, Theorem 12.3] and define

$$E_n := \{y \in \mathbb{R}^d \mid nq_L \le \|x - y\|_2 < (n + 1)q_L\}, \quad n \in \mathbb{N}_0.$$

Thus, any $x_\ell \in X_L$ is contained in exactly one E_n. Additionally, we can estimate the elements in $X_L \cap E_n$ by $|X_L \cap E_0| \leq 2^d$ and $|X_L \cap E_n| \leq 3^d n^{d-1}$ for $n \geq 1$. Therefore, the sum is bounded by

$$\begin{aligned}
&\sum_{\ell=N_{L-1}+1}^{N_L} e^{-\eta\|x_\ell - x\|_2\left(\frac{1}{q_L}-\frac{1}{q_k}\right)} \leq \sum_{n=0}^{\infty} \sum_{x_\ell \in X_L \cap E_n} e^{-\eta\|x_\ell - x\|_2\left(\frac{1}{q_L}-\frac{1}{q_k}\right)} \\
&\leq \sum_{n=0}^{\infty} \sum_{x_\ell \in X_L \cap E_n} e^{-\eta\, n q_L\left(\frac{1}{q_L}-\frac{1}{q_k}\right)} \leq 2^d + 3^d \sum_{n=1}^{\infty} n^{d-1} e^{-n\eta\left(1-\frac{q_L}{q_k}\right)} \\
&\leq 2^d + 3^d \sum_{n=1}^{\infty} n^{d-1} e^{-n\eta\left(1-\frac{q_L}{q_{L-1}}\right)} =: \widetilde{C}_L < \infty
\end{aligned}$$

as our assumptions imply that $q_{L-1} \geq q_L$.
Altogether, we can use this estimate in the recursive formula and obtain, with the induction hypothesis,

$$\begin{aligned}
\left|b_j^{(L)}(x)\right| &\leq \left|b_j^{(L-1)}(x)\right| + \sum_{\ell=N_{L-1}+1}^{N_L} \left|b_j^{(L-1)}(x_\ell)\chi_\ell^{(L)}(x)\right| \\
&\leq C_{L-1} e^{-\frac{\eta}{q_k}\|x-x_j\|_2} + C_1 C_{L-1} \widetilde{C}_L e^{-\frac{\eta}{q_k}\|x-x_j\|_2} \\
&= C_L e^{-\frac{\eta}{q_k}\|x-x_j\|_2},
\end{aligned}$$

where $C_L := C_{L-1}\left(1 + C_1 \widetilde{C}_L\right)$.

□

Motivated by the recursive representation (6.13), we next want to analyze how the approximation spaces behave when we increase the number of levels.

Corollary 6.4.10.
For all $1 \leq k \leq L$ and $f \in C(\mathcal{D})$, we have

$$A_{[1:L]}\left(A_{[1:k]}(f)\right) = A_{[1:k]}(f).$$

This implies that the multilevel operator is idempotent and that the approximation spaces $A_k(C(\mathcal{D}))$ form an increasing sequence of linear subspaces, i.e.,

$$A_1(C(\mathcal{D})) \subseteq A_2(C(\mathcal{D})) \subseteq \cdots \subseteq A_L(C(\mathcal{D})).$$

Proof. The proof of the first statement follows by induction over $j \in \mathbb{N}_0$, which describes the difference $j = L - k$ for all $k, L \in \mathbb{N}$ with $k \leq L$. Let $f \in C(\mathcal{D})$ be arbitrary but fixed. The induction starts with $j = 0$, which directly implies $k = L$. As A_L interpolates f in X_L, we have $A_L(f)(X_L) = f(X_L)$. Substituting the expression in the multilevel scheme, we directly obtain

$$A_L(A_L(f)) = A_L(f).$$

Now, assume the statement is true for $j \in \mathbb{N}$. Then, for all $k, L \in \mathbb{N}$ with $k \leq L$ and $L - k = j + 1$, we can use the recursive formulation of the multilevel method to obtain

$$A_L(A_k(f)) = A_{L-1}(A_k(f)) + \mathcal{I}_{X_L,\Phi_L}\left(A_k(f) - A_{L-1}(A_k(f))\right).$$

Here, the induction hypothesis is applicable on $A_{L-1}(A_j(f))$, since $L-1-k=j$, which leads to

$$\begin{aligned} A_L(A_k(f)) &= A_{L-1}(A_k(f)) + \mathcal{I}_{X_L,\Phi_L}\left(A_k(f) - A_{L-1}(A_k(f))\right) \\ &= A_k(f) + \mathcal{I}_{X_L,\Phi_L}\left(A_k(f) - A_k(f)\right) \\ &= A_k(f), \end{aligned}$$

and thus proves the claim. □

Following the preceding statement, we are able to decompose the image of $A_L(C(\mathcal{D}))$ as

$$A_L(C(\mathcal{D})) = A_{L-1}(C(\mathcal{D})) + \widetilde{V}_L.$$

This again leads to the question whether this sum is direct and, moreover, if we can construct a basis $\widetilde{b}_1^{(L)}, \dots, \widetilde{b}_{N_L}^{(L)}$ of $A_L(C(\mathcal{D}))$ such that $\widetilde{b}_1^{(L)}, \dots, \widetilde{b}_{N_k}^{(L)}$ is a basis of $A_k(C(\mathcal{D}))$ for all $1 \le k \le L$. Fortunately, we can use the recursive definition of A_L to tackle these questions.

Corollary 6.4.11.
Assume Setting 6.2.1 holds, where the point sets $X_1 \subseteq \cdots \subseteq X_L$ are nested. Then, for all $L \in \mathbb{N}$, we can write the image of A_L recursively as

$$A_L(C(\mathcal{D})) = \begin{cases} \operatorname{span}\left\{\chi_j^{(1)} \mid 1 \le j \le N_1\right\}, & L = 1, \\ A_{L-1}(C(\mathcal{D})) \oplus \operatorname{span}\left\{\chi_j^{(L)} \mid N_{L-1}+1 \le j \le N_L\right\}, & L \ge 2, \end{cases}$$

where $N_0 := 0$. This implies that only the cardinal functions $\chi_j^{(L)}$ with $x_j \in X_L \setminus X_{L-1}$ provide a new contribution to the approximation space. Additionally, the multilevel approximation on the L-th level can be written as

$$\begin{aligned} A_L(f) = &\sum_{j=1}^{N_1} f(x_j)\chi_j^{(1)} + \sum_{j=N_1+1}^{N_2} \left(f(x_j) - A_1(f)(x_j)\right)\chi_j^{(2)} \\ &+ \cdots + \sum_{j=N_{L-1}+1}^{N_L} \left(f(x_j) - A_{L-1}(f)(x_j)\right)\chi_j^{(L)} \end{aligned}$$

for all $f \in C(\mathcal{D})$. In particular, if we define for all $1 \le k \le L$ and $j \in \{N_{k-1}+1, \dots, N_k\}$ the functions

$$\widetilde{b}_j^{(L)} := \chi_j^{(k)},$$

we obtain a basis of $A_L(C(\mathcal{D}))$ such that $\widetilde{b}_1^{(L)}, \dots, \widetilde{b}_{N_k}^{(L)}$ is a basis of $A_k(C(\mathcal{D}))$ for all $1 \le k \le L$.

Proof. We again use induction and start with $L = 1$: Here, the multilevel approximation A_1 is given by

$$A_1(f) = \mathcal{I}_{X_1,\Phi_1}(f) = \sum_{j=1}^{N_1} f(x_j)\chi_j^{(1)}, \quad f \in C(\mathcal{D}),$$

where we used the Lagrange representation of the kernel interpolation. The Lagrange functions are independent as cardinal functions on X_1 and $\dim(A_1(C(\mathcal{D}))) = N_1 = |X_1|$, thus, the base step follows.
Now assume the corollary is true for $L-1$ with $L \geq 2$. The recursive definition of A_L yields

$$\begin{aligned} A_L(f) &= A_{L-1}(f) + \mathcal{I}_{X_L,\Phi_L}(f - A_{L-1}(f)) \\ &= A_{L-1}(f) + \sum_{j=1}^{N_L} \left(f\left(x_j\right) - A_{L-1}(f)\left(x_j\right)\right) \chi_j^{(L)}, \end{aligned}$$

for all $f \in C(\mathcal{D})$, where we again used the Lagrange representation of $\mathcal{I}_{X_L,\Phi_L}$. If we now take a closer look at the last expression, we can use that $A_{L-1}(f)$ interpolates f in X_{L-1} and the sum degenerates to

$$A_L(f) = A_{L-1}(f) + \sum_{j=N_{L-1}+1}^{N_L} \left(f\left(x_j\right) - A_{L-1}(f)\left(x_j\right)\right) \chi_k^{(L)}.$$

With the induction hypothesis applied to $A_{L-1}(f)$, we observe that the Lagrange functions in $\{\chi_j^{(L)} \mid N_{L-1}+1 \leq j \leq N_L\}$ are independent, as they are cardinal, and that all of these functions vanish on the points of $X_1 \subseteq X_2 \subseteq \cdots \subseteq X_{L-1}$, thus they are also independent of the basis functions of $A_{L-1}(C(\mathcal{D}))$ given as cardinal functions by the induction hypothesis. Together with these observations, the claim follows by induction. □

Remark 6.4.12.
Using the interpolation property $f(x_j) = A_L(f)(x_j)$ for all $1 \leq j \leq N_L$, allows us to determine the coefficients $\alpha_1, \dots \alpha_{N_L} \in \mathbb{R}$ of the representation

$$A_L(f) = \sum_{k=1}^{L} \sum_{j=N_{k-1}+1}^{N_k} \alpha_j \chi_j^{(k)}$$

as follows: If $N_{k-1}+1 \leq j \leq N_k$, i.e., $x_j \in X_k \setminus X_{k-1}$, all basis functions $\widetilde{b}_\ell^{(L)}$ with $\ell > N_k$ evaluated in x_j vanish, and, for $N_{k-1}+1 \leq \ell \leq N_k$ the basis functions satisfy $\widetilde{b}_\ell^{(L)}(x_k) = \delta_{k,\ell}$. This leads to a lower triangular block matrix in $\mathbb{R}^{N_L \times N_L}$ given by

$$M := \begin{pmatrix} I_{N_1} & 0_{N_1 \times (N_2 - N_1)} & \cdots & 0_{N_1 \times (N_L - N_{L-1})} \\ \chi_{[1:N_1]}^{(1)}(X_2 \setminus X_1) & I_{N_2 - N_1} & \cdots & 0_{(N_2 - N1) \times (N_L - N_{L-1})} \\ \vdots & \vdots & \ddots & \vdots \\ \chi_{[1:N_1]}^{(1)}(X_L \setminus X_{L-1}) & \chi_{[N_1+1:N_2]}^{(2)}(X_L \setminus X_{L-1}) & \cdots & I_{N_L - N_{L-1}}, \end{pmatrix}$$

where

$$\chi_{[N_{k-1}+1:N_k]}^{(j)}(X_\ell \setminus X_{\ell-1}) = \left(\chi_{N_{k-1}+1}^{(j)}(X_\ell \setminus X_{\ell-1}) \dots \chi_{N_k}^{(j)}(X_\ell \setminus X_{\ell-1})\right)$$

denotes a matrix in $\mathbb{R}^{(N_\ell - N_{\ell-1}) \times (N_j - N_{j-1})}$ consisting of Lagrange functions evaluated at specific point sets. The corresponding solution of

$$M \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_L \end{pmatrix} = \begin{pmatrix} f(X_1) \\ \vdots \\ f(X_L \setminus X_{L-1}) \end{pmatrix}$$

with $\boldsymbol{\alpha}_j \in \mathbb{R}^{N_j - N_{j-1}}$ can now be calculated by blockwise backwards substitution. We note that the diagonal blocks consist of identity matrices, which leads to the recursive formula

$$\begin{aligned}\boldsymbol{\alpha}_1 &= f(X_1),\\ \boldsymbol{\alpha}_\ell &= f(X_\ell \setminus X_{\ell-1}) - \sum_{k=1}^{j-1} \chi_{[N_{k-1}+1:N_k]}(X_\ell \setminus X_{\ell-1})\boldsymbol{\alpha}_k, \quad \ell > 1.\end{aligned}$$

6.5 Application to random fields

As in Section 5.2, we apply the deterministic multilevel method pathwise to a random field $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ for $\sigma > \frac{d}{2}$ and provide error bounds.

Corollary 6.5.1.
Consider the setting of Theorem 6.3.1 with $\sigma > \frac{d}{2}$ and point sets $X_1, \dots, X_L \subseteq \mathcal{D}$, and let $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ be a second order random field with covariance function Cov*. Denote the covariance function of $A_L(a)$ by* Cov**. Then the following estimates hold:*

1. *The mean integrated square error of a to the approximant $A_L(a)$ can be bounded by*

$$\mathbb{E}\left[\|a - A_L(a)\|^2_{L^2(\mathcal{D})}\right]^{\frac{1}{2}} \leq C\alpha^L \mathbb{E}\left[\|a\|_{H^\sigma(\mathcal{D})}\|^2\right]^{\frac{1}{2}}.$$

   For the error of the covariance functions we obtain

$$\|\operatorname{Cov} - \operatorname{Cov}^*\|_{L^2(\mathcal{D}\times\mathcal{D})} \leq C\alpha^L \|\operatorname{Cov}\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})},$$

   where in both cases $C = C(\mathcal{D}, \sigma, \Phi) > 0$ is a constant independent of the random field a and the point sets $X_1, \dots, X_L$. Moreover, α given as in Theorem 6.3.1 is assumed to be less than 1*.*

2. *If the coupling of μ and ν in Corollary 6.3.4 holds, the uniform mean square error can be bounded by*

$$\mathbb{E}\left[\|a - A_L(a)\|^2_{C(\mathcal{D})}\right]^{\frac{1}{2}} \leq C\beta^L \|a\|_{L^2(\Omega; H^\sigma(\mathcal{D}))},$$

   and the error of the covariance functions fulfills

$$\|\operatorname{Cov} - \operatorname{Cov}^*\|_{C(\mathcal{D}\times\mathcal{D})} \leq C\beta^L \|\operatorname{Cov}\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})},$$

   where in both cases $C = C(\mathcal{D}, \sigma, \gamma, \Phi) > 0$ is a constant independent of the random field a and the point sets $X_1, \dots, X_L$. Moreover, β given as in Corollary 6.3.4 is assumed to be less than 1*.*

Proof. As before, we apply the error estimate of Theorem 6.3.1 to every path of a and obtain

$$\begin{aligned}\mathbb{E}\left[\|a - A_L(a)\|^2_{L^2(\mathcal{D})}\right] &= \int_\Omega \|a(\omega,\cdot) - A_L(a)(\omega,\cdot)\|^2_{L^2(\mathcal{D})}\, d\mathbb{P}(\omega)\\ &\leq C^2\alpha^{2L} \int_\Omega \|a(\omega,\cdot)\|^2_{H^\sigma(\mathcal{D})}\, d\mathbb{P}(\omega)\\ &= C^2\alpha^{2L}\|a\|^2_{L^2(\Omega; H^\sigma(\mathcal{D}))}.\end{aligned}$$

In the same way, we use Corollary 6.3.4 to obtain

$$\begin{aligned}&\mathbb{E}\left[\|a - A_L(a)\|^2_{C(\mathcal{D})}\right]\\ &\leq C^2\beta^{2L}\int_\Omega \|a(\omega,\cdot)\|^2_{H^\sigma(\mathcal{D})}\,d\mathbb{P}(\omega)\\ &= C^2\beta^{2L}\|a\|^2_{L^2(\Omega;H^\sigma(\mathcal{D}))}.\end{aligned}$$

For the covariance function, we use Lemma 5.1.2 to see that $\mathrm{Cov}^* = A_L\overline{\otimes}A_L(\mathrm{Cov})$ and then apply Lemma 3.5.5 to rewrite

$$Id_\mathcal{D}\overline{\otimes}Id_\mathcal{D} - A_L\overline{\otimes}A_L = Id_\mathcal{D}\overline{\otimes}(Id_\mathcal{D} - A_L) + (Id_\mathcal{D} - A_L)\overline{\otimes}Id_\mathcal{D} - (Id_\mathcal{D} - A_L)\overline{\otimes}(Id_\mathcal{D} - A_L).$$

Next, we know by Proposition 2.4.23 that both the norms $\|\cdot\|_{L^2(\mathcal{D}\times\mathcal{D})}$ and $\|\cdot\|_{C(\mathcal{D}\times\mathcal{D})}$ are compatible with $\|\cdot\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})}$, thus we can proceed analogously to Theorem 3.5.6 and obtain

$$\begin{aligned}\|\operatorname{Cov} - \operatorname{Cov}^*\|_{L^2(\mathcal{D}\times\mathcal{D})} &\leq 2\|Id_\mathcal{D}\|_{H^\sigma(\mathcal{D})\to L^2(\mathcal{D})}\|Id_\mathcal{D} - A_L\|_{H^\sigma(\mathcal{D})\to L^2(\mathcal{D})}\|\operatorname{Cov}\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})}\\ &\quad + \|Id_\mathcal{D} - A_L\|^2_{H^\sigma(\mathcal{D})\to L^2(\mathcal{D})}\|\operatorname{Cov}\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})}\\ &\leq C\left(2\alpha^L + \alpha^{2L}\right)\|\operatorname{Cov}\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})} \leq C\alpha^L\|\operatorname{Cov}\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})},\end{aligned}$$

where we used the estimate given in Theorem 6.3.1 to bound the operator norm

$$\|Id_\mathcal{D} - A_L\|_{H^\sigma(\mathcal{D})\to L^2(\mathcal{D})} \leq C\alpha^L$$

and additionally used $\|Id_\mathcal{D}\|_{H^\sigma(\mathcal{D})\to L^2(\mathcal{D})} \leq 1$. For the error estimate in $C(\mathcal{D}\times\mathcal{D})$, we proceed analogously using Corollary 6.3.4 and Theorem 2.2.10 to bound the norm of the inclusion by $\|Id_\mathcal{D}\|_{H^\sigma(\mathcal{D})\to C(\mathcal{D})} \leq C_E$. □

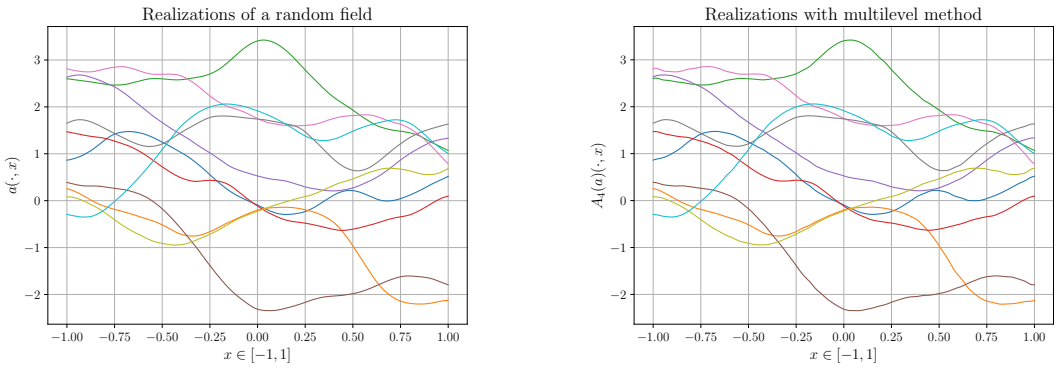


Figure 6.2: A total of 10 realizations of a Gaussian random field with mean $m(x) := 1$ and covariance function using the Matérn kernel $\Phi_{2.5}$. The left-hand side shows the true realizations, while the right-hand side shows the multilevel approximations of the paths using the Wendland kernel $\Phi = \Phi_{1,1}$ and a total of 41 points with $L = 4$ levels and parameters $\mu = \frac{1}{2}$ and $\nu = 2.5$.

As in the proof of Corollary 6.5.1, we can use Lemma 5.1.2 to obtain a formula for the mean covariance function of $A_L(a)$. Using the Lagrange type expression given in (6.9) of Remark 6.4.3, we can additionally give an explicit representation of this covariance function. However, we postpone this result to Corollary 7.2.3, where we prove the representation in a more general setting.

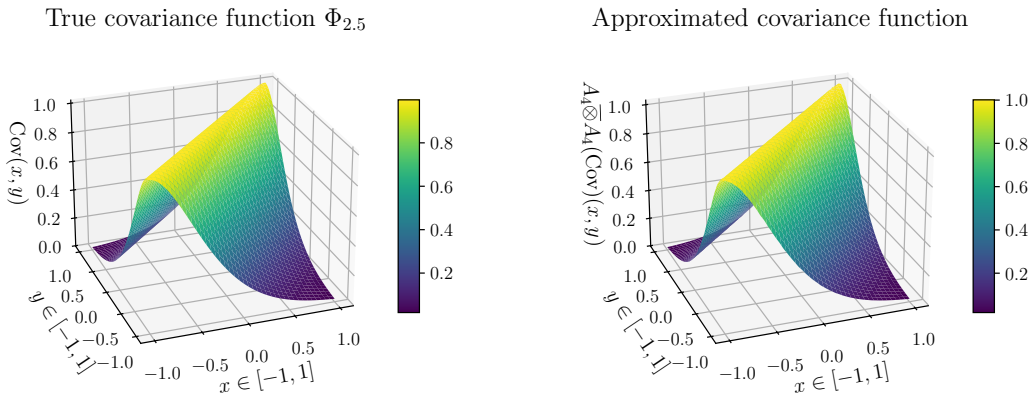


Figure 6.3: Covariance functions to the processes considered in Figure 6.2. Similar to Figure 5.1 and Figure 5.2, the kernel $\Phi_{2.5} \in H^{3,3}([-1,1] \times [-1,1])$, hence, by Theorem 4.3.5, the process fulfills $a \in L^2(\Omega; H^\tau([-1,1]))$ for $0 \leq \tau < 2.5$. The Wendland kernel $\Phi_{1,1}$ reproduces $H^2(\mathbb{R})$. Thus, the error estimate of Corollary 6.5.1 holds.

7 Multilevel methods on mixed regularity spaces

In the last chapter, we discussed the multilevel method $A_L \colon H^\sigma(\mathcal{D}) \to L^2(\mathcal{D})$ in great detail and provided an error estimate as well as different representations of the method with various basis functions. Afterwards, we applied the method pathwise to a random field $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ with covariance function $\mathrm{Cov} \in H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})$. By Lemma 5.1.2, we additionally obtained an explicit formula for the emerging covariance function of $A_L(a)$, which is given by

$$(A_L \overline{\otimes} A_L)\,\mathrm{Cov}\,.$$

This tensor product operator can be interpreted as an approximation operator on the mixed regularity space $H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})$, which motivates us to study this method in a slightly more general setting on $H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})$.
However, this is not the only way to generalize the multilevel method to mixed regularity spaces. As we saw in Corollary 3.1.12, we can construct a reproducing kernel K on $\mathcal{D}\times\mathcal{T}$ with $H_K(\mathcal{D}\times\mathcal{T}) \cong H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})$ using two positive definite functions $\Phi \in L^1(\mathbb{R}^d)\cap C(\mathbb{R}^d)$ and $\Psi \in L^1(\mathbb{R}^n)\cap C(\mathbb{R}^n)$ with Hilbert spaces $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$ and $H_\Psi(\mathbb{R}^n) \cong H^\tau(\mathbb{R}^n)$, respectively, setting $K := \Phi \otimes \Psi$, and restricting the result to the proper domain of definition. This allows us to mimic the multilevel method presented in Algorithm 1, by replacing the occurring interpolation steps with the ones using scaled versions of the tensor kernel $\Phi \otimes \Psi$. As it turns out, this idea leads to a different way of generalizing the multilevel method to mixed regularity spaces. In the following, we present both methods, beginning with the one using tensorized kernels, and compare them afterwards.

7.1 Multilevel method with tensorized kernels

As motivated before, we first investigate the multilevel method that is based on Algorithm 1 but uses tensorized kernels for the interpolation steps. To be more precise, we consider the following setting:

Setting 7.1.1.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ and $\mathcal{T} \subseteq \mathbb{R}^n$ be two bounded domains with Lipschitz boundary, and let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ and $\Psi \in L^1(\mathbb{R}^n) \cap C(\mathbb{R}^n)$ be two positive definite functions with Hilbert spaces $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$, and $H_\Psi(\mathbb{R}^n) \cong H^\tau(\mathbb{R}^n)$, where $\tau > \frac{n}{2}$, respectively. We denote by $0 < c_{1,\Phi} \le c_{2,\Phi}$ and $0 < c_{1,\Psi} \le c_{2,\Psi}$ the equivalence constants of (3.1) for Φ and Ψ, i.e.,

$$c_{1,\Phi}(1+\|\omega_1\|_2^2)^{-\sigma} \le \widehat{\Phi}(\omega_1) \le c_{2,\Phi}\left(1+\|\omega_1\|_2^2\right)^{-\sigma}, \quad \omega_1 \in \mathbb{R}^d,$$
$$c_{1,\Psi}(1+\|\omega_2\|_2^2)^{-\tau} \le \widehat{\Psi}(\omega_2) \le c_{2,\Psi}\left(1+\|\omega_2\|_2^2\right)^{-\tau}, \quad \omega_2 \in \mathbb{R}^n.$$

Furthermore, let $(X_j)_{j\in\mathbb{N}} \subseteq \mathcal{D}$ and $(Y_j)_{j\in\mathbb{N}} \subseteq \mathcal{T}$ be sequences of point sets with mesh norms $h_{X_j} := h_{X_j,\mathcal{D}}$ and $h_{Y_j} := h_{Y_j,\mathcal{D}}$, for $j \in \mathbb{N}$, respectively. For all $j \in \mathbb{N}$, we set $\delta_j := \nu_{\mathcal{D}} h_{X_j,\mathcal{D}}$ and $\kappa_j := \nu_{\mathcal{T}} h_{Y_j,\mathcal{T}}$, with $\nu_{\mathcal{D}}, \nu_{\mathcal{T}} \geq 1$, and define

$$\Phi_j := \delta_j^{-d} \Phi\left(\frac{\cdot}{\delta_j}\right), \quad \Psi_j := \kappa_j^{-n} \Psi\left(\frac{\cdot}{\kappa_j}\right)$$

as in Lemma 6.1.1.

In this setting, the kernel $\Phi_j \otimes \Psi_j$ is a positive definite function on $\mathbb{R}^d \times \mathbb{R}^n$ with native space $H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d \times \mathbb{R}^n) \cong H^{\sigma,\tau}_{\mathrm{mix}}(\mathbb{R}^d \times \mathbb{R}^n)$, as we will see in Lemma 7.1.3 below. We then consider the following algorithm:

Algorithm 2 Multilevel method with tensorized kernels

Input: $f_0 := 0,\ e_0 := f$
for $j = 1, 2, \ldots$ **do**
  $s_j \leftarrow \mathcal{I}_{X_j\times Y_j, \Phi_j\otimes\Psi_j} e_{j-1}$
  $f_j \leftarrow f_{j-1} + s_j$
  $e_j \leftarrow e_{j-1} - s_j$
end for

Similarly to before, we can define the *multilevel operator with tensorized kernels* for a fixed level $L \in \mathbb{N}$.

Definition 7.1.2.
Assume Setting 7.1.1 holds, and let $L \in \mathbb{N}$. The *multilevel operator with tensorized kernels* on level L using Algorithm 2 is defined as

$$\boldsymbol{A}_L \colon C(\mathcal{D}\times\mathcal{T}) \to L^2(\mathcal{D}\times\mathcal{T}), \quad f \mapsto \boldsymbol{A}_L(f) := f_L.$$

As already claimed, we can give a norm equivalence similar to Lemma 6.1.1.

Lemma 7.1.3.
Let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ and $\Psi \in L^1(\mathbb{R}^n) \cap C(\mathbb{R}^n)$ be two positive definite functions with $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$, and $H_\Psi(\mathbb{R}^n) \cong H^\tau(\mathbb{R}^n)$, where $\tau > \frac{n}{2}$, respectively. Then, for all $0 < \delta, \kappa \leq 1$, the kernel

$$\Phi_\delta \otimes \Psi_\gamma \colon \mathbb{R}^d \times \mathbb{R}^n, \quad (x,y) \mapsto \delta^{-d}\kappa^{-n}\Phi\left(\frac{x}{\delta}\right)\Psi\left(\frac{y}{\kappa}\right)$$

fulfills

$$\begin{aligned} &c_{1,\Phi}c_{1,\Psi}\left(1+\delta^2\|\omega_1\|_2^2\right)^{-\sigma}\left(1+\kappa^2\|\omega_2\|_2^2\right)^{-\tau} \leq \widehat{\Phi_\delta\otimes\Psi}_\kappa(\omega_1,\omega_2) \\ &\leq c_{2,\Phi}c_{2,\Psi}\left(1+\delta^2\|\omega_1\|_2^2\right)^{-\sigma}\left(1+\kappa^2\|\omega_2\|_2^2\right)^{-\tau}. \end{aligned}$$

In particular, this estimate proves that

$$\begin{aligned} &c_{1,\Phi}c_{1,\Psi}\left(1+\|\omega_1\|_2^2\right)^{-\sigma}\left(1+\|\omega_2\|_2^2\right)^{-\tau} \leq \widehat{\Phi_\delta \otimes \Psi}_\kappa(\omega_1,\omega_2) \\ &\leq \delta^{-\sigma}\kappa^{-\tau}c_{2,\Phi}c_{2,\Psi}\left(1+\|\omega_1\|_2^2\right)^{-\sigma}\left(1+\|\omega_2\|_2^2\right)^{-\tau}. \end{aligned}$$

Thus, $\Phi_\delta \otimes \Psi_\kappa$ is a reproducing kernel with $H_{\Phi_\delta\otimes\Psi_\kappa}(\mathbb{R}^d\times\mathbb{R}^n) \cong H^{\sigma,\tau}_{\mathrm{mix}}(\mathbb{R}^d\times\mathbb{R}^n)$, where the equivalence constants are given by

$$\begin{aligned} &\sqrt{c_{1,\Phi}c_{1,\Psi}}(2\pi)^{\frac{d+n}{4}}\|f\|_{H_{\Phi_\delta\otimes\Psi_\kappa}(\mathbb{R}^d\times\mathbb{R}^n)} \leq \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathbb{R}^d\times\mathbb{R}^n)} \\ &\leq \sqrt{c_{2,\Phi}c_{2,\Psi}}(2\pi)^{\frac{d+n}{4}}\delta^{-\sigma}\kappa^{-\tau}\|f\|_{H_{\Phi_\delta\otimes\Psi_\kappa}(\mathbb{R}^d\times\mathbb{R}^n)} \end{aligned}$$

for all $f \in H^{\sigma,\tau}_{\mathrm{mix}}(\mathbb{R}^d\times\mathbb{R}^n)$. Moreover, the Fourier transform of $\Phi_\delta\otimes\Psi_\kappa$ satisfies

$$\begin{aligned} \widehat{\Phi_\delta \otimes \Psi}_\kappa(\omega_1,\omega_2) &= (2\pi)^{-\frac{n+d}{2}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^n} e^{-(x_1,x_2)^T(\omega_1,\omega_2)}\Phi_\delta(\omega_1)\Psi_\kappa(\omega_2)\,d\omega_2\,d\omega_1 \\ &= \widehat{\Phi_\delta}(\omega_1)\widehat{\Psi_\kappa}(\omega_2). \end{aligned}$$

Proof. By Corollary 3.1.12, we already know that $\Phi_\delta\otimes\Psi_\kappa$ is a positive definite function and that its Fourier transform is given by

$$\widehat{\Phi_\delta \otimes \Psi}_\kappa(\omega_1,\omega_2) = \widehat{\Phi_\delta}(\omega_1)\widehat{\Psi_\kappa}(\omega_2), \quad \omega_1\in\mathbb{R}^d, \quad \omega_2\in\mathbb{R}^n.$$

We can therefore apply Lemma 6.1.1 on both kernels to obtain

$$\begin{aligned} &c_{1,\Phi}c_{1,\Psi}(1+\|\omega_1\|_2^2)^{-\sigma}(1+\|\omega_2\|_2^2)^{-\tau} \\ &\leq c_{1,\Phi}c_{1,\Psi}(1+\delta^2\|\omega_1\|_2^2)^{-\sigma}(1+\kappa^2\|\omega_2\|_2^2)^{-\tau} \\ &\leq \widehat{\Phi_\delta \otimes \Psi}_\kappa(\omega_1,\omega_2) \\ &\leq c_{2,\Phi}c_{2,\Psi}(1+\delta^2\|\omega_1\|_2^2)^{-\sigma}(1+\kappa^2\|\omega_2\|_2^2)^{-\tau} \\ &\leq c_{2,\Phi}c_{2,\Psi}\delta^{-\sigma}\kappa^{-\tau}(1+\|\omega_1\|_2^2)^{-\sigma}(1+\|\omega_2\|_2^2)^{-\tau}. \end{aligned}$$

Now, using the representation of the native space norm in Corollary 3.1.12 given by

$$\|f\|^2_{\Phi_\delta\otimes\Psi_\kappa} = (2\pi)^{-\frac{d+n}{2}}\int_{\mathbb{R}^d}\int_{\mathbb{R}^n}\frac{\left|\widehat{f}(\omega_1,\omega_2)\right|^2}{\widehat{\Phi_\delta}(\omega_1)\widehat{\Psi_\kappa}(\omega_2)}\,d\omega_2\,d\omega_1,$$

we can apply these estimates on $\widehat{\Phi_\delta\otimes\Psi}_\kappa = \widehat{\Phi_\delta}\widehat{\Psi_\kappa}$ to get the claim. □

Next, we want to give a convergence result similar to Theorem 6.3.1. As it turns out, the proof of this result can also be established using the ideas from the aforementioned theorem. It reads as follows:

Theorem 7.1.4 (Multilevel method with tensorized kernels)**.**
Assume Setting 7.1.1 holds, where the mesh norms of $(X_j)_{j\in\mathbb{N}} \subseteq \mathcal{D}$ and $(Y_j)_{j\in\mathbb{N}} \subseteq \mathcal{T}$ fulfill

$$h_{X_1,\mathcal{D}} \geq h_{X_2,\mathcal{D}} \geq \cdots > 0,$$
$$h_{Y_1,\mathcal{T}} \geq h_{Y_2,\mathcal{T}} \geq \cdots > 0.$$

Additionally, let there be constants $0 < \mu_{\mathcal{D}}, \mu_{\mathcal{T}} < 1$ and $c_{\mathcal{D}}, c_{\mathcal{T}} > 0$ such that

$$c_{\mathcal{D}}\mu_{\mathcal{D}} h_{X_j,\mathcal{D}} \leq h_{X_{j+1},\mathcal{D}} \leq \mu_{\mathcal{D}} h_{X_j,\mathcal{D}},$$
$$c_{\mathcal{T}}\mu_{\mathcal{T}} h_{Y_j,\mathcal{T}} \leq h_{Y_{j+1},\mathcal{T}} \leq \mu_{\mathcal{T}} h_{Y_j,\mathcal{T}}$$

holds.
Now, choose $\nu_{\mathcal{D}}, \nu_{\mathcal{T}} \geq 1$ such that $\nu_{\mathcal{D}} h_{X_1,\mathcal{D}} \leq 1$ and $\nu_{\mathcal{T}} h_{Y_1,\mathcal{T}} \leq 1$. We set the scaling parameter as

$$\delta_j := \nu_{\mathcal{D}} h_{X_j,\mathcal{D}}, \quad \kappa_j := \nu_{\mathcal{T}} h_{Y_j,\mathcal{T}},$$

and the sequence of kernels

$$\Phi_j \otimes \Psi_j := \Phi_{\delta_j} \otimes \Psi_{\kappa_j}$$

for all $j \in \mathbb{N}$.
Then there exist two constants $C, C_1 > 0$ with

$$\alpha := C_1 \left[\left(\mu_{\mathcal{D}}^{\sigma} + \nu_{\mathcal{D}}^{-\sigma} \right) \left(\mu_{\mathcal{T}}^{\tau} + \nu_{\mathcal{T}}^{-\tau} \right) + \mu_{\mathcal{D}}^{\sigma} + \nu_{\mathcal{D}}^{-\sigma} + \mu_{\mathcal{T}}^{\tau} + \nu_{\mathcal{T}}^{-\tau} \right] \tag{7.1}$$

such that for all $f \in H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})$ the estimate

$$\|f - \boldsymbol{A}_L(f)\|_{L^2(\mathcal{D}\times\mathcal{T})} \leq C\alpha^L \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}$$

holds. Hence, the multilevel method converges linearly to f in $L^2(\mathcal{D}\times\mathcal{T})$ if $\alpha < 1$.

As in Theorem 6.3.1, we first provide a recursive estimate similar to Lemma 6.3.2. For this, it is useful to state several bounds of occurring operator norms separately.

Lemma 7.1.5.
Consider the setting of Theorem 7.1.4. Then, for all $j \in \mathbb{N}$, the following bounds hold:

1. $\|Id_{\mathcal{D}} - \mathcal{I}_{X_j,\Phi_j}\|_{H_{\Phi_j}(\mathbb{R}^d)\to H^{\sigma}(\mathcal{D})} \leq \sqrt{c_{2,\Phi}}(2\pi)^{\frac{d}{4}} \delta_j^{-\sigma}$,
2. $\|Id_{\mathcal{D}} - \mathcal{I}_{X_j,\Phi_j}\|_{H_{\Phi_j}(\mathbb{R}^d)\to L^2(\mathcal{D})} \leq C_{S,\mathcal{D}} \sqrt{c_{2,\Phi}}(2\pi)^{\frac{d}{4}} \left(\frac{h_{X_j,\mathcal{D}}}{\delta_j} \right)^{\sigma}$,
3. $\|Id_{\mathcal{D}}\|_{H_{\Phi_j}(\mathbb{R}^d)\to L^2(\mathcal{D})} \leq \sqrt{c_{2,\Phi}}(2\pi)^{\frac{d}{4}}$,
4. $\|Id_{\mathcal{D}}\|_{H_{\Phi_j}(\mathbb{R}^d)\to H^{\sigma}(\mathcal{D})} \leq \sqrt{c_{2,\Phi}}(2\pi)^{\frac{d}{4}} \delta_j^{-\sigma}$.

Analogous bounds also hold when replacing $Id_{\mathcal{D}}$ with $Id_{\mathcal{T}}$ and $\mathcal{I}_{X_j,\Phi_j}$ with $\mathcal{I}_{Y_j,\Psi_j}$ in the corresponding spaces $H_{\Psi_j}(\mathbb{R}^n)$ and $H^{\tau}(\mathcal{T})$. Here and in the following, $C_{S,\mathcal{D}}, C_{S,\mathcal{T}} > 0$ denote the constants from the sampling inequalities, given in Corollary 3.3.3, on $\mathcal{D}$ and $\mathcal{T}$, respectively.

Proof. In order to bound the occurring operator norms, let $f \in H_{\Phi_j}(\mathbb{R}^d)$ be arbitrary, but fixed. We begin with the first bound. With the norm equivalence given in Lemma 6.1.1 and the best approximation property, we obtain

$$\begin{aligned}\|f - \mathcal{I}_{X_j,\Phi_j}(f)\|_{H^\sigma(\mathcal{D})} &\leq \|f - \mathcal{I}_{X_j,\Phi_j}(f)\|_{H^\sigma(\mathbb{R}^d)} \\ &\leq \sqrt{c_{2,\Phi}}(2\pi)^{\frac{d}{4}}\delta_j^{-\sigma}\|f - \mathcal{I}_{X_j,\Phi_j}(f)\|_{H_{\Phi_j}(\mathbb{R}^d)} \\ &\leq \sqrt{c_{2,\Phi}}(2\pi)^{\frac{d}{4}}\delta_j^{-\sigma}\|f\|_{H_{\Phi_j}(\mathbb{R}^d)},\end{aligned}$$

which yields the estimate.
The second one is an application of the first statement and Theorem 3.3.4:

$$\begin{aligned}\|f - \mathcal{I}_{X_j,\Phi_j}(f)\|_{L^2(\mathcal{D})} &\leq C_{S,\mathcal{D}}h_{X_j,\mathcal{D}}^\sigma\|f - \mathcal{I}_{X_j,\Phi_j}(f)\|_{H^\sigma(\mathcal{D})} \\ &\leq C_{S,\mathcal{D}}h_{X_j,\mathcal{D}}^\sigma\|Id_{\mathcal{D}} - \mathcal{I}_{X_j,\Phi_j}\|_{H_{\Phi_j}(\mathbb{R}^d)\to H^\sigma(\mathcal{D})}\|f\|_{H_{\Phi_j}(\mathbb{R}^d)} \\ &\leq C_{S,\mathcal{D}}\sqrt{c_{2,\Phi}}(2\pi)^{\frac{d}{4}}\left(\frac{h_{X_j,\mathcal{D}}}{\delta_j}\right)^\sigma\|f\|_{H_{\Phi_j}(\mathbb{R}^d)}.\end{aligned}$$

For the third bound, we use the Plancherel formula to obtain

$$\begin{aligned}\|f\|_{L^2(\mathcal{D})}^2 \leq \|f\|_{L^2(\mathbb{R}^d)}^2 &= \int_{\mathbb{R}^d}\left|\widehat{f}(\omega)\right|^2 d\omega \\ &\leq \int_{\mathbb{R}^d}\left|\widehat{f}(\omega)\right|^2\left(1+\delta_j^2\|\omega\|_2^2\right)^\sigma d\omega \leq c_{2,\Phi}(2\pi)^{\frac{d}{2}}\|f\|_{H_{\Phi_j}(\mathbb{R}^d)}^2.\end{aligned}$$

The fourth estimate follows directly by the norm equivalence given in Lemma 6.1.1. □

With these bounds, we can now state and prove a recursive estimate analogous to Lemma 6.3.2.

Lemma 7.1.6.
Assume the setting of Theorem 7.1.4 holds. Then, for all $j \in \mathbb{N}$, the estimate

$$\|Ee_j\|_{H_{\Phi_{j+1}\otimes\Psi_{j+1}}(\mathbb{R}^d\times\mathbb{R}^n)} \leq \alpha\|Ee_{j-1}\|_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)}$$

holds, where α is given as in (7.1)*. Here, E denotes the extension operator established in Theorem 2.4.24.*

Proof. First, we introduce abbreviations used throughout this proof: For $j \in \mathbb{N}$, we denote by

$$R_{j,\mathcal{D}} := Id_{\mathcal{D}} - I_{X_j,\Phi_j}, \quad R_{j,\mathcal{T}} := Id_{\mathcal{T}} - I_{Y_j,\Psi_j}$$

the residual operator on the j-th level on $\mathcal{D}$ and $\mathcal{T}$, respectively. In the following, we suppress the dependency on $\mathcal{D}$ and $\mathcal{T}$ in the mesh norms, i.e., we write $h_{X_j} := h_{X_j,\mathcal{D}}$ and $h_{Y_j} := h_{Y_j,\mathcal{T}}$, for $j \in \mathbb{N}$. As it turns out, it is also useful to define additional constants $c_1 := (2\pi)^{\frac{d+n}{2}}c_{1,\Phi}c_{1,\Psi}$ and $c_2 := (2\pi)^{\frac{d+n}{2}}c_{2,\Phi}c_{2,\Psi}$.
To begin with, we seek to decompose the interpolation error in the j-th step. For this, we recapitulate Lemma 3.5.1, which states that the interpolation operator fulfills $\mathcal{I}_{X_j\times Y_j,\Phi_j\otimes\Psi_j} = \mathcal{I}_{X_j,\Phi_j}\overline{\otimes}\mathcal{I}_{Y_j,\Psi_j}$. Together with Lemma 3.5.5, this implies:

$$Id_{\mathcal{D}\times\mathcal{T}} - \mathcal{I}_{X_j\times Y_j,\Phi_j\otimes\Psi_j} = R_{j,\mathcal{D}}\overline{\otimes}Id_{\mathcal{T}} + Id_{\mathcal{D}}\overline{\otimes}R_{j,\mathcal{T}} - R_{j,\mathcal{D}}\overline{\otimes}R_{j,\mathcal{T}}. \tag{7.2}$$

In order to prove the recursive estimate, we observe that

$$\delta_{j+1} = \nu_{\mathcal{D}} h_{X_{j+1}} \leq \nu_{\mathcal{D}} \mu_{\mathcal{D}} h_{X_j} = \mu_{\mathcal{D}} \delta_j \leq \delta_j \leq \cdots \leq \delta_1 \leq 1,$$

and, analogously, $\kappa_{j+1} \leq \kappa_j \leq \cdots \leq \kappa_1 \leq 1$. Thus, Lemma 7.1.3 is applicable in both arguments. Similarly as in the proof of Lemma 6.3.2, we use Lemma 7.1.3 but divide the emerging integral into four parts:

$$\begin{aligned}
&\|Ee_j\|^2_{\Phi_{j+1}\otimes\Psi_{j+1}} \\
&\leq \frac{1}{c_1} \int_{\mathbb{R}^d} \int_{\mathbb{R}^n} \left|\widehat{Ee_j}(\omega_1,\omega_2)\right|^2 \left(1+\delta_{j+1}^2\|\omega_1\|_2^2\right)^{\sigma} \left(1+\kappa_{j+1}^2\|\omega\|_2^2\right)^{\tau} d\omega_2\, d\omega_1 \\
&= \frac{1}{c_1} \\
&\Bigg[\int_{\|\omega_1\|_2\leq\frac{1}{\delta_{j+1}}} \int_{\|\omega_2\|_2\leq\frac{1}{\kappa_{j+1}}} \left|\widehat{Ee_j}(\omega_1,\omega_2)\right|^2 \left(1+\delta_{j+1}^2\|\omega_1\|_2^2\right)^{\sigma} \left(1+\kappa_{j+1}^2\|\omega\|_2^2\right)^{\tau} d\omega_2\, d\omega_1 \\
&+ \int_{\|\omega_1\|_2\leq\frac{1}{\delta_{j+1}}} \int_{\|\omega_2\|_2>\frac{1}{\kappa_{j+1}}} \left|\widehat{Ee_j}(\omega_1,\omega_2)\right|^2 \left(1+\delta_{j+1}^2\|\omega_1\|_2^2\right)^{\sigma} \left(1+\kappa_{j+1}^2\|\omega\|_2^2\right)^{\tau} d\omega_2\, d\omega_1 \\
&+ \int_{\|\omega_1\|_2>\frac{1}{\delta_{j+1}}} \int_{\|\omega_2\|_2\leq\frac{1}{\kappa_{j+1}}} \left|\widehat{Ee_j}(\omega_1,\omega_2)\right|^2 \left(1+\delta_{j+1}^2\|\omega_1\|_2^2\right)^{\sigma} \left(1+\kappa_{j+1}^2\|\omega\|_2^2\right)^{\tau} d\omega_2\, d\omega_1 \\
&+ \int_{\|\omega_1\|_2>\frac{1}{\delta_{j+1}}} \int_{\|\omega_2\|_2>\frac{1}{\kappa_{j+1}}} \left|\widehat{Ee_j}(\omega_1,\omega_2)\right|^2 \left(1+\delta_{j+1}^2\|\omega_1\|_2^2\right)^{\sigma} \left(1+\kappa_{j+1}^2\|\omega\|_2^2\right)^{\tau} d\omega_2\, d\omega_1\Bigg] \\
&= \frac{1}{c_1}\left[I_1+I_2+I_3+I_4\right].
\end{aligned}$$

We want to estimate all terms separately, starting with I_1. Here, we use $\|\omega_1\|_2 \leq \frac{1}{\delta_{j+1}}$ and $\|\omega_2\|_2 \leq \frac{1}{\kappa_{j+1}}$ to obtain

$$(1+\delta_{j+1}^2\|\omega_1\|_2^2)^{\sigma} \leq 2^{\sigma}, \quad (1+\kappa_{j+1}^2\|\omega_2\|_2^2)^{\tau} \leq 2^{\tau}.$$

This leads to an upper bound of I_1:

$$\begin{aligned}
I_1 &\leq 2^{\sigma+\tau} \int_{\|\omega_1\|_2\leq\frac{1}{\delta_{j+1}}} \int_{\|\omega_2\|_2\leq\frac{1}{\kappa_{j+1}}} \left|\widehat{Ee_j}(\omega_1,\omega_2)\right|^2 d\omega_2\, d\omega_1 \\
&\leq 2^{\sigma+\tau} \int_{\mathbb{R}^d} \int_{\mathbb{R}^n} \left|\widehat{Ee_j}(\omega_1,\omega_2)\right|^2 d\omega_2\, d\omega_1 \\
&= 2^{\sigma+\tau} \|Ee_j\|^2_{L^2(\mathbb{R}^d\times\mathbb{R}^n)} \\
&\leq 2^{\sigma+\tau} C_E^2 \|e_j\|^2_{L^2(\mathcal{D}\times\mathcal{T})},
\end{aligned}$$

where we used Plancherels identity and the boundedness of the extension operator in the last step. We now insert the recursive formulation for e_j, which leads to

$$e_j = e_{j-1} - \mathcal{I}_{X_j\times Y_j,\Phi_j\otimes\Psi_j}(e_{j-1}) = Ee_{j-1} - \mathcal{I}_{X_j\times Y_j,\Phi_j\otimes\Psi_j}(Ee_{j-1}),$$

as E restricted to $\mathcal{D}\times\mathcal{T}$ is the identity. Next, we apply the tensor decomposition given in (7.2) to this expression and estimate further:

$$\begin{aligned}
&\|e_j\|_{L^2(\mathcal{D}\times\mathcal{T})}\\
&\leq \|R_{j,\mathcal{D}}\overline{\otimes}Id_{\mathcal{T}}(Ee_{j-1})\|_{L^2(\mathcal{D}\times\mathcal{T})} + \|Id_{\mathcal{D}}\overline{\otimes}R_{j,\mathcal{T}}(Ee_{j-1})\|_{L^2(\mathcal{D}\times\mathcal{T})}\\
&\quad + \|R_{j,\mathcal{D}}\overline{\otimes}R_{j,\mathcal{T}}(Ee_{j-1})\|_{L^2(\mathcal{D}\times\mathcal{T})}\\
&\leq \Big[\|R_{j,\mathcal{D}}\overline{\otimes}Id_{\mathcal{T}}\|_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)\to L^2(\mathcal{D}\times\mathcal{T})} + \|Id_{\mathcal{D}}\overline{\otimes}R_{j,\mathcal{T}}\|_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)\to L^2(\mathcal{D}\times\mathcal{T})}\\
&\qquad + \|R_{j\mathcal{D}}\overline{\otimes}R_{j,\mathcal{T}}\|_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)\to L^2(\mathcal{D}\times\mathcal{T})}\Big]\,\|Ee_{j-1}\|_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)}.
\end{aligned}$$

The emerging operator norms can now be further bounded using the compatibility of $H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n) = H_{\Phi_j}(\mathbb{R}^d)\overline{\otimes}H_{\Psi_j}(\mathbb{R}^n)$ with $L^2(\mathcal{D}\times\mathcal{T}) = L^2(\mathcal{D})\overline{\otimes}L^2(\mathcal{T})$, as shown in Proposition 2.4.23, which, for example, allows us to make the following estimate:

$$\begin{aligned}
&\|R_{j,\mathcal{D}}\overline{\otimes}Id_{\mathcal{T}}\|_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)\to L^2(\mathcal{D}\times\mathcal{T})}\\
&\leq \|R_{j,\mathcal{D}}\|_{H_{\Phi_j}(\mathbb{R}^d)\to L^2(\mathcal{D})}\|Id_{\mathcal{T}}\|_{H_{\Psi_j}(\mathbb{R}^n)\to L^2(\mathcal{T})}\\
&\leq C_{S,\mathcal{D}}\sqrt{c_{2,\Phi}c_{2,\Psi}}(2\pi)^{\frac{d+n}{4}}\left(\frac{h_{X_j}}{\delta_j}\right)^{\sigma} = C_{S,\mathcal{D}}\sqrt{c_2}\left(\frac{1}{\nu_{\mathcal{D}}}\right)^{\sigma}.
\end{aligned}$$

Here, we applied Lemma 7.1.5 in the penultimate step. Proceeding analogously with the other occurring terms ultimately yields the estimate

$$\begin{aligned}
I_1 &\leq 2^{\sigma+\tau}C_E^2\|e_j\|^2_{L^2(\mathcal{D}\times\mathcal{T})}\\
&\leq 2^{\sigma+\tau}C_E^2\Big[C_{S,\mathcal{D}}\sqrt{c_2}\nu_{\mathcal{D}}^{-\sigma} + C_{S,\mathcal{T}}\sqrt{c_2}\nu_{\mathcal{T}}^{-\tau}\\
&\qquad + C_{S,\mathcal{D}}C_{S,\mathcal{T}}\sqrt{c_2}\nu_{\mathcal{D}}^{-\sigma}\nu_{\mathcal{T}}^{-\tau}\Big]^2\|Ee_{j-1}\|^2_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)} \qquad (7.3)\\
&\leq 2^{\sigma+\tau}C_E^2c_2\max\left\{C_{S,\mathcal{D}}, C_{S,\mathcal{T}}, (C_{S,\mathcal{D}}C_{S,\mathcal{T}})\right\}^2\\
&\quad\cdot\left[\nu_{\mathcal{D}}^{-\sigma} + \nu_{\mathcal{T}}^{-\tau} + \nu_{\mathcal{D}}^{-\sigma}\nu_{\mathcal{T}}^{-\tau}\right]^2\|Ee_{j-1}\|^2_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)}.
\end{aligned}$$

We now proceed with a bound for I_2. In this case, we have $\|\omega_2\|_2 > \frac{1}{\kappa_{j+1}}$, which leads to

$$\left(1+\kappa_{j+1}^2\|\omega_2\|_2^2\right)^{\tau} = \kappa_{j+1}^{2\tau}\left(\kappa_{j+1}^{-2}+\|\omega_2\|_2^2\right)^{\tau} \leq \kappa_{j+1}^{2\tau}2^{\tau}\|\omega_2\|_2^{2\tau} \leq \kappa_{j+1}^{2\tau}2^{\tau}(1+\|\omega_2\|_2^2)^{\tau}. \quad (7.4)$$

Using this estimate in I_2 yields

$$\begin{aligned}
I_2 &= \int_{\|\omega_1\|_2\leq\frac{1}{\delta_{j+1}}}\int_{\|\omega_2\|_2>\frac{1}{\kappa_{j+1}}}\left|\widehat{Ee_j}(\omega_1,\omega_2)\right|^2\left(1+\delta_{j+1}^2\|\omega_1\|_2^2\right)^{\sigma}\left(1+\kappa_{j+1}^2\|\omega\|_2^2\right)^{\tau}\,d\omega_2\,d\omega_1\\
&\leq 2^{\sigma+\tau}\kappa_{j+1}^{2\tau}\int_{\mathbb{R}^d}\int_{\mathbb{R}^n}\left|\widehat{Ee_j}(\omega_1,\omega_2)\right|^2\left(1+\|\omega_2\|_2^2\right)^{\tau}\,d\omega_2\,d\omega_1\\
&\leq 2^{\sigma+\tau}C_E^2\kappa_{j+1}^{2\tau}\|e_j\|^2_{H^{0,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}.
\end{aligned}$$

As before, we can rewrite $e_j = Ee_{j-1} - \mathcal{I}_{X_j\times Y_j,\Phi_j\otimes\Psi_j}(Ee_{j-1})$ on $\mathcal{D}\times\mathcal{T}$ using (7.2), apply the triangle inequality, and bound the emerging terms using the corresponding operator

norms and compatibility.

$$
\begin{aligned}
&\|e_j\|_{H^{0,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}\\
&\leq \|R_{j,\mathcal{D}}\overline{\otimes} Id_{\mathcal{T}}(Ee_{j-1})\|_{H^{0,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} + \|Id_{\mathcal{D}}\overline{\otimes} R_{j,\mathcal{T}}(Ee_{j-1})\|_{H^{0,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}\\
&\quad + \|R_{j,\mathcal{D}}\overline{\otimes} R_{j,\mathcal{T}}(Ee_{j-1})\|_{H^{0,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}\\
&\leq \Big[\|R_{j,\mathcal{D}}\|_{H_{\Phi_j}(\mathbb{R}^d)\to L^2(\mathcal{D})}\|Id_{\mathcal{T}}\|_{H_{\Psi_j}(\mathbb{R}^n)\to H^\tau(\mathcal{T})} + \|Id_{\mathcal{D}}\|_{H_{\Phi_j}(\mathbb{R}^d)\to L^2(\mathcal{D})}\|R_{j,\mathcal{T}}\|_{H_{\Psi_j}(\mathbb{R}^n)\to H^\tau(\mathcal{T})}\\
&\qquad + \|R_{j,\mathcal{D}}\|_{H_{\Phi_j}(\mathbb{R}^d)\to L^2(\mathcal{D})}\|R_{j,\mathcal{T}}\|_{H_{\Psi_j}(\mathbb{R}^n)\to H^\tau(\mathcal{T})}\Big]\,\|Ee_{j-1}\|_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)}.
\end{aligned}
$$

Again, with Lemma 7.1.5, we obtain an upper bound for I_2:

$$
\begin{aligned}
I_2 &\leq 2^{\sigma+\tau}C_E^2\,\kappa_{j+1}^{2\tau}\Big[C_{S,\mathcal{D}}\sqrt{c_2}\nu_{\mathcal{D}}^{-\sigma}\kappa_j^{-\tau} + \sqrt{c_2}\kappa_j^{-\tau}\\
&\qquad + C_{S,\mathcal{D}}\sqrt{c_2}\nu_{\mathcal{D}}^{-\sigma}\kappa_j^{-\tau}\Big]^2\,\|Ee_{j-1}\|^2_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)} \qquad (7.5)\\
&\leq 2^{\sigma+\tau}C_E^2c_2\max\{1,2C_{S,\mathcal{D}}\}^2\Big[\nu_{\mathcal{D}}^{-\sigma}\mu_{\mathcal{T}}^{\tau} + \mu_{\mathcal{T}}^{\tau}\Big]^2\,\|Ee_{j-1}\|^2_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)}.
\end{aligned}
$$

By switching the roles of $\mathcal{D}$ and $\mathcal{T}$ in I_2, an analogous proof also leads to a bound for I_3:

$$
I_3 \leq 2^{\sigma+\tau}C_E^2c_2\max\{1,2C_{S,\mathcal{T}}\}^2\Big[\mu_{\mathcal{D}}^{\sigma}\nu_{\mathcal{T}}^{-\tau} + \mu_{\mathcal{D}}^{\sigma}\Big]^2\,\|Ee_{j-1}\|^2_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)}. \qquad (7.6)
$$

Finally, for I_4, we can use (7.4) in both arguments, as $\|\omega_1\|_2 > \frac{1}{\delta_{j+1}}$ and $\|\omega_2\|_2 > \frac{1}{\kappa_{j+1}}$, to obtain

$$
\begin{aligned}
I_4 &\leq 2^{\sigma+\tau}\delta_{j+1}^{2\sigma}\kappa_{j+1}^{2\tau}\int_{\mathbb{R}^d}\int_{\mathbb{R}^n}\Big|\widehat{Ee_j}(\omega_1,\omega_2)\Big|^2\Big(1+\|\omega_1\|_2^2\Big)^{\sigma}\Big(1+\|\omega_2\|_2^2\Big)^{\tau}\,d\omega_2\,d\omega_1\\
&= 2^{\sigma+\tau}\delta_{j+1}^{2\sigma}\kappa_{j+1}^{2\tau}\|Ee_j\|^2_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathbb{R}^d\times\mathbb{R}^n)}\\
&\leq 2^{\sigma+\tau}C_E^2\delta_{j+1}^{2\sigma}\kappa_{j+1}^{2\tau}\|e_j\|^2_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}.
\end{aligned}
$$

Together with the decomposition $e_j = Ee_{j-1} - \mathcal{I}_{X_j\times Y_j,\Phi_j\otimes\Psi_j}(Ee_{j-1})$ on $\mathcal{D}\times\mathcal{T}$ and the norm equivalence given in Lemma 7.1.3, we conclude

$$
\begin{aligned}
&\delta_{j+1}^{\sigma}\kappa_{j+1}^{\tau}\|e_j\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}\\
&\leq \sqrt{c_2}\left(\frac{\delta_{j+1}}{\delta_j}\right)^{\sigma}\left(\frac{\kappa_{j+1}}{\kappa_j}\right)^{\tau}\|Ee_{j-1} - \mathcal{I}_{X_j\times Y_j,\Phi_j\otimes\Psi_j}(Ee_{j-1})\|_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)}\\
&\leq \sqrt{c_2}\mu_{\mathcal{D}}^{\sigma}\mu_{\mathcal{T}}^{\tau}\|Ee_{j-1}\|_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)},
\end{aligned}
$$

so this ultimately yields

$$
I_4 \leq 2^{\sigma+\tau}C_E^2c_2\mu_{\mathcal{D}}^{2\sigma}\mu_{\mathcal{T}}^{2\tau}\|Ee_{j-1}\|^2_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)}. \qquad (7.7)
$$

If we now collect the estimates (7.3), (7.5), (7.6) and (7.7), we obtain

$$\begin{aligned}
&\|Ee_j\|^2_{H_{\Phi_{j+1}\otimes\Psi_{j+1}}(\mathbb{R}^d\times\mathbb{R}^n)} \leq \frac{1}{c_1}(I_1+I_2+I_3+I_4)\\
&\leq \frac{2^{\sigma+\tau}c_2C_E^2}{c_1}\Big(\max\left\{C_{S,\mathcal{D}},C_{S,\mathcal{T}},(C_{S,\mathcal{D}}C_{S,\mathcal{T}})\right\}^2\left[\nu_{\mathcal{D}}^{-\sigma}+\nu_{\mathcal{T}}^{-\tau}+\nu_{\mathcal{D}}^{-\sigma}\nu_{\mathcal{T}}^{-\tau}\right]^2\\
&\qquad\qquad+\max\left\{1,2C_{S,\mathcal{D}}\right\}^2\left[\nu_{\mathcal{D}}^{-\sigma}\mu_{\mathcal{T}}^{\tau}+\mu_{\mathcal{T}}^{\tau}\right]^2+\max\left\{1,2C_{S,\mathcal{T}}\right\}^2\left[\mu_{\mathcal{D}}^{\sigma}\nu_{\mathcal{T}}^{-\tau}+\mu_{\mathcal{D}}^{\sigma}\right]^2\\
&\qquad\qquad+\mu_{\mathcal{D}}^{2\sigma}\mu_{\mathcal{T}}^{2\tau}\Big)\|Ee_{j-1}\|^2_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)}\\
&\leq \frac{2^{\sigma+\tau}c_2C_E^2}{c_1}\max\left\{1,2C_{S,\mathcal{D}},2C_{S,\mathcal{T}},(C_{S,\mathcal{D}}C_{S,\mathcal{T}})\right\}^2\|Ee_{j-1}\|^2_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)}\\
&\quad\cdot\left[\nu_{\mathcal{D}}^{-\sigma}+\nu_{\mathcal{T}}^{-\tau}+\nu_{\mathcal{D}}^{-\sigma}\nu_{\mathcal{T}}^{-\tau}+\nu_{\mathcal{D}}^{-\sigma}\mu_{\mathcal{T}}^{\tau}+\mu_{\mathcal{T}}^{\tau}+\mu_{\mathcal{D}}^{\sigma}\nu_{\mathcal{T}}^{-\tau}+\mu_{\mathcal{D}}^{\sigma}+\mu_{\mathcal{D}}^{\sigma}\mu_{\mathcal{T}}^{\tau}\right]^2\\
&= C_1^2\left[\left(\mu_{\mathcal{D}}^{\sigma}+\nu_{\mathcal{D}}^{-\sigma}\right)\left(\mu_{\mathcal{T}}^{\tau}+\nu_{\mathcal{T}}^{-\tau}\right)+\mu_{\mathcal{D}}^{\sigma}+\nu_{\mathcal{D}}^{-\sigma}+\mu_{\mathcal{T}}^{\tau}+\nu_{\mathcal{T}}^{-\tau}\right]^2\|Ee_{j-1}\|^2_{H_{\Phi_j\otimes\Psi_j}(\mathbb{R}^d\times\mathbb{R}^n)}
\end{aligned}$$

after factorizing the occurring terms, where C_1 can be chosen as

$$C_1 = 2^{\frac{\sigma+\tau}{2}}C_E\frac{\sqrt{c_2}}{\sqrt{c_1}}\max\left\{1,2C_{S,\mathcal{D}},2C_{S,\mathcal{T}},C_{S,\mathcal{D}}C_{S,\mathcal{T}}\right\}.$$

□

As in Theorem 6.3.1, this can be used to prove the error estimate of Theorem 7.1.4:

Proof. To start the recursion, we see that

$$\begin{aligned}
&\|f-\boldsymbol{A}_L(f)\|^2_{L^2(\mathcal{D}\times\mathcal{T})}\\
&\leq \|Ee_L\|^2_{L^2(\mathbb{R}^d\times\mathbb{R}^n)}\\
&= \|\widehat{Ee_L}\|_{L^2(\mathbb{R}^d\times\mathbb{R}^n)}\\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}^n}|\widehat{Ee_L}(\omega_1,\omega_2)|^2\,d\omega_2\,d\omega_1\\
&\leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^n}|\widehat{Ee_L}(\omega_1,\omega_2)|^2\left(1+\delta_{L+1}^2\|\omega\|_2^2\right)^{\sigma}\left(1+\kappa_{L+1}^2\|\omega_2\|_2^2\right)^{\tau}\,d\omega_2\,d\omega_1\\
&\leq c_{2,\Phi}c_{2,\Psi}(2\pi)^{\frac{d+n}{2}}\|Ee_L\|^2_{H_{\Phi_{L+1}\otimes\Psi_{L+1}}(\mathbb{R}^d\times\mathbb{R}^n)}.
\end{aligned}$$

This allows us to apply the recursion formula of Lemma 7.1.6 L-times to obtain

$$\begin{aligned}
\|f-\boldsymbol{A}_L(f)\|_{L^2(\mathcal{D}\times\mathcal{T})} &\leq \sqrt{c_{2,\Phi}c_{2,\Psi}}(2\pi)^{\frac{d+n}{4}}\|Ee_L\|_{H_{\Phi_{L+1}\otimes\Psi_{L+1}}(\mathbb{R}^d\times\mathbb{R}^n)}\\
&\leq \sqrt{c_{2,\Phi}c_{2,\Psi}}(2\pi)^{\frac{d+n}{4}}\alpha^L\|Ee_0\|_{H_{\Phi_1\otimes\Psi_1}(\mathbb{R}^d\times\mathbb{R}^n)}\\
&\leq C_E\frac{\sqrt{c_{2,\Phi}c_{2,\Psi}}}{\sqrt{c_{1,\Phi}c_{1,\Psi}}}\alpha^L\|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})},
\end{aligned}$$

which is the stated result. □

We now want to prove a pointwise error estimate. The analysis presented in the following suggests that such an estimate is only be possible if $\sigma > d$, $\tau > n$ and $\mu_{\mathcal{D}}$ is coupled to $\mu_{\mathcal{T}}$, in addition to the coupling of $\mu_{\mathcal{D}}$ with $\nu_{\mathcal{D}}$ and $\mu_{\mathcal{T}}$ with $\nu_{\mathcal{T}}$ as in the untensorized case. However, numerical examples suggest that this bound is not optimal.

Corollary 7.1.7.
Assume the setting of Theorem 7.1.4 holds. Additionally, let there be constants $\gamma_{\mathcal{D}}, \gamma_{\mathcal{T}} > 1$ such that

$$\frac{\gamma_{\mathcal{D}}}{\nu_{\mathcal{D}}} \leq \mu_{\mathcal{D}}, \quad \frac{\gamma_{\mathcal{T}}}{\nu_{\mathcal{T}}} \leq \mu_{\mathcal{T}}.$$

Then there exist two constants $C, \widetilde{C_1} > 0$ such that

$$\begin{aligned}\|f - \boldsymbol{A}_L(f)\|_{C(\mathcal{D}\times\mathcal{T})} &\leq C\left[\widetilde{C_1}\left(\mu_{\mathcal{D}}^{\sigma-\frac{d}{2}}\mu_{\mathcal{T}}^{\tau-\frac{n}{2}} + \mu_{\mathcal{D}}^{\sigma-\frac{d}{2}}\mu_{\mathcal{T}}^{-\frac{n}{2}} + \mu_{\mathcal{D}}^{-\frac{d}{2}}\mu_{\mathcal{T}}^{\tau-\frac{n}{2}}\right)\right]^L \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} \\ &\leq C\max\left\{\frac{\mu_{\mathcal{D}}^d}{\mu_{\mathcal{T}}^n}, \frac{\mu_{\mathcal{T}}^n}{\mu_{\mathcal{D}}^d}\right\}^{\frac{L}{2}} \beta^L \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})},\end{aligned}$$

where

$$\beta = \widetilde{C_1}\left(\mu_{\mathcal{D}}^{\sigma-\frac{d}{2}}\mu_{\mathcal{T}}^{\tau-\frac{n}{2}} + \mu_{\mathcal{D}}^{\sigma-d} + \mu_{\mathcal{T}}^{\tau-n}\right).$$

Thus, the multilevel method using tensorized kernels converges pointwise if $\sigma > d$, $\tau > n$, and $\sqrt{\max\left\{\frac{\mu_{\mathcal{D}}^d}{\mu_{\mathcal{T}}^n}, \frac{\mu_{\mathcal{T}}^n}{\mu_{\mathcal{D}}^d}\right\}}\beta < 1$.

Proof. In order to apply Lemma 7.1.6, we have to bound the embedding constant from $H_{\Phi_{L+1}\otimes\Psi_{L+1}}(\mathbb{R}^d \times \mathbb{R}^n)$ to $C(\mathcal{D}\times\mathcal{T})$, as we estimate

$$\begin{aligned}&\|f - \boldsymbol{A}_L(f)\|_{C(\mathcal{D}\times\mathcal{T})} \\ &= \|(Id_{\mathcal{D}}\overline{\otimes}Id_{\mathcal{T}})(E(f - \boldsymbol{A}_L(f)))\|_{C(\mathcal{D}\times\mathcal{T})} \\ &\leq \|Id_{\mathcal{D}}\overline{\otimes}Id_{\mathcal{T}}\|_{H_{\Phi_{L+1}(\mathbb{R}^d)}\overline{\otimes}H_{\Psi_{L+1}(\mathbb{R}^n)}\to C(\mathcal{D})\overline{\otimes}_\vee C(\mathcal{T})}\|Ee_L\|_{H_{\Phi_{L+1}\otimes\Psi_{L+1}}(\mathbb{R}^d\times\mathbb{R}^n)} \\ &\leq \|Id_{\mathcal{D}}\|_{H_{\Phi_{L+1}}(\mathbb{R}^d)\to C(\mathcal{D})}\|Id_{\mathcal{T}}\|_{H_{\Psi_{L+1}}(\mathbb{R}^n)\to C(\mathcal{T})}\|Ee_L\|_{H_{\Phi_{L+1}\otimes\Psi_{L+1}}(\mathbb{R}^d\times\mathbb{R}^n)},\end{aligned} \tag{7.8}$$

using the extension operator E on $\mathcal{D}\times\mathcal{T}$, and compatibility of the norms defined on $H_{\Phi_{L+1}\otimes\Psi_{L+1}}(\mathbb{R}^d \otimes \mathbb{R}^n)$ with the norm of $C(\mathcal{D}\times\mathcal{T})$. For bounds on the operator norms of the inclusions, we recapitulate (6.8) stated in the proof of Corollary 6.3.4, which showed that these norms are given by

$$\|Id_{\mathcal{D}}\|_{H_{\Phi_{L+1}}(\mathbb{R}^d)\to C(\mathcal{D})} = \sqrt{\Phi(0)}\delta_{L+1}^{-\frac{d}{2}}, \quad \|Id_{\mathcal{T}}\|_{H_{\Psi_{L+1}}(\mathbb{R}^n)\to C(\mathcal{T})} = \sqrt{\Psi(0)}\kappa_{L+1}^{-\frac{n}{2}}.$$

Inserting these into (7.8), we can apply the recursive estimate of Lemma 7.1.6 to obtain

$$\begin{aligned}&\|f - \boldsymbol{A}_L(f)\|_{C(\mathcal{D}\times\mathcal{T})} \\ &\leq \sqrt{\Phi(0)\Psi(0)}\delta_{L+1}^{-\frac{d}{2}}\kappa_{L+1}^{-\frac{n}{2}}\|Ee_L\|_{H_{\Phi_{L+1}\otimes\Psi_{L+1}}(\mathbb{R}^d\times\mathbb{R}^n)} \\ &\leq C\delta_{L+1}^{-\frac{d}{2}}\kappa_{L+1}^{-\frac{n}{2}}\alpha^L\|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}.\end{aligned}$$

We now proceed similarly to Corollary 6.3.4. First, we can bound δ_{L+1} and κ_{L+1} from below by

$$\begin{aligned}\delta_{L+1} &= \nu_{\mathcal{D}} h_{X_{L+1}} \geq \nu_{\mathcal{D}} c_{\mathcal{D}} \mu_{\mathcal{D}} h_{X_L} \geq \cdots \geq c_{\mathcal{D}}^L \mu_{\mathcal{D}}^L \nu_{\mathcal{D}} h_{X_1} = c_{\mathcal{D}}^L \mu_{\mathcal{D}}^L \delta_1, \\ \kappa_{L+1} &= \nu_{\mathcal{T}} h_{Y_{L+1}} \geq \nu_{\mathcal{T}} c_{\mathcal{T}} \mu_{\mathcal{T}} h_{Y_L} \geq \cdots \geq c_{\mathcal{T}}^L \mu_{\mathcal{T}}^L \nu_{\mathcal{T}} h_{Y_1} = c_{\mathcal{T}}^L \mu_{\mathcal{T}}^L \kappa_1.\end{aligned}$$

Next, we use the definition of $\gamma_{\mathcal{D}}$ and $\gamma_{\mathcal{T}}$ to express α given in (7.1) in terms of $\mu_{\mathcal{D}}$ and $\mu_{\mathcal{T}}$, as

$$\begin{aligned}\alpha &= C_1 \left[\left(\mu_{\mathcal{D}}^{\sigma} + \nu_{\mathcal{D}}^{-\sigma} \right) \left(\mu_{\mathcal{T}}^{\tau} + \nu_{\mathcal{T}}^{-\tau} \right) + \mu_{\mathcal{D}}^{\sigma} + \nu_{\mathcal{D}}^{-\sigma} + \mu_{\mathcal{T}}^{\tau} + \nu_{\mathcal{T}}^{-\tau} \right] \\ &\leq 4 \max \left\{ 1, \gamma_{\mathcal{D}}^{-\sigma}, \gamma_{\mathcal{T}}^{-\tau} \right\} C_1 \left(\mu_{\mathcal{D}}^{\sigma} \mu_{\mathcal{T}}^{\tau} + \mu_{\mathcal{D}}^{\sigma} + \mu_{\mathcal{T}}^{\tau} \right).\end{aligned}$$

Therefore, we can combine both bounds to obtain

$$\|f - \boldsymbol{A}_L(f)\|_{C(\mathcal{D}\times\mathcal{T})} \leq C \widetilde{C_1}^L \left(\mu_{\mathcal{D}}^{\sigma-\frac{d}{2}} \mu_{\mathcal{T}}^{\tau-\frac{n}{2}} + \mu_{\mathcal{D}}^{\sigma-\frac{d}{2}} \mu_{\mathcal{T}}^{-\frac{n}{2}} + \mu_{\mathcal{D}}^{-\frac{d}{2}} \mu_{\mathcal{T}}^{\tau-\frac{n}{2}} \right)^L \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}, \tag{7.9}$$

where $\widetilde{C_1} = C_1 4 \max \left\{ 1, \gamma_{\mathcal{D}}^{-\sigma}, \gamma_{\mathcal{T}}^{-\tau} \right\} (c_{\mathcal{D}} c_{\mathcal{T}})^{-1}$.
To compensate the negative powers of $\mu_{\mathcal{D}}$ and $\mu_{\mathcal{T}}$, which undermine the error result, we factorize

$$\begin{aligned}&\left(\mu_{\mathcal{D}}^{\sigma-\frac{d}{2}} \mu_{\mathcal{T}}^{\tau-\frac{n}{2}} + \mu_{\mathcal{D}}^{\sigma-\frac{d}{2}} \mu_{\mathcal{T}}^{-\frac{n}{2}} + \mu_{\mathcal{D}}^{-\frac{d}{2}} \mu_{\mathcal{T}}^{\tau-\frac{n}{2}} \right) \\ &= \left(\mu_{\mathcal{D}}^{\sigma-\frac{d}{2}} \mu_{\mathcal{T}}^{\tau-\frac{n}{2}} + \mu_{\mathcal{D}}^{\sigma-d} \left(\frac{\mu_{\mathcal{D}}^d}{\mu_{\mathcal{T}}^n} \right)^{\frac{1}{2}} + \mu_{\mathcal{T}}^{\tau-n} \left(\frac{\mu_{\mathcal{T}}^n}{\mu_{\mathcal{D}}^d} \right)^{\frac{1}{2}} \right) \\ &\leq \max \left\{ \left(\frac{\mu_{\mathcal{D}}^d}{\mu_{\mathcal{T}}^n} \right), \left(\frac{\mu_{\mathcal{T}}^n}{\mu_{\mathcal{D}}^d} \right) \right\}^{\frac{1}{2}} \left(\mu_{\mathcal{D}}^{\sigma-\frac{d}{2}} \mu_{\mathcal{T}}^{\tau-\frac{n}{2}} + \mu_{\mathcal{D}}^{\sigma-d} + \mu_{\mathcal{T}}^{\tau-n} \right),\end{aligned}$$

which yields the stated bounds. □

Corollary 7.1.7 indicates that pointwise convergence demands more restrictive assumptions. In particular, the convergence enforcing term in (7.9) has to be less than or equal to one. However, the numerical test below suggests that this bound is not optimal: On the left-hand side of Figure 7.1, we use $d = 5$ and $n = 1$, and set $\mathcal{D} = [-1, 1]^5$ and $\mathcal{T} = [-1, 1]$ as the domains. Additionally, we choose $\sigma := 3 > \frac{5}{2}$ and $\tau = 2 > \frac{1}{2}$ with the according Wendland kernels. For point sets $(X_j)_{j\in\mathbb{N}}$ and $(Y_j)_{j\in\mathbb{N}}$, we use uniform grids on $\mathcal{D}$ and $\mathcal{T}$, respectively, where we roughly use 2^d additional points on each subsequent level, i.e., the mesh norms h_{X_j} and h_{Y_j} get halved per level. This leads to $\mu_{\mathcal{D}} = \mu_{\mathcal{T}} = \frac{1}{2} =: \mu$, and we accordingly choose $\nu_{\mathcal{D}} = \nu_{\mathcal{T}} = 2$, thus $\gamma_{\mathcal{D}} = \gamma_{\mathcal{T}} = 1$. In this setting, the upper bound in Corollary 7.1.7 simplifies to

$$\begin{aligned}\|f - \boldsymbol{A}_L(f)\|_{C(\mathcal{D}\times\mathcal{T})} &\leq C \left(\widetilde{C}_1 \left(\mu^{\sigma-\frac{d}{2}+\tau-\frac{n}{2}} \right) + \mu^{\sigma-\frac{d}{2}-\frac{n}{2}} + \mu^{\tau-\frac{d}{2}-\frac{n}{2}} \right)^L \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} \\ &= C \left(\widetilde{C}_1 \left(\mu^2 + 1 + \mu^{-1} \right) \right)^L \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}.\end{aligned}$$

This indicates that the convergence fails using the reasonable assumption $\widetilde{C}_1 > \frac{4}{13}$. On the right-hand side, we proceed similary, choosing $d = 4$ with $\mathcal{D} = [-1,1]^4$, and $n = 1$ with $\mathcal{T} = [-1,1]$. We choose $\sigma = 5/2$ and $\tau = 2$, and the occuring point sets and parameters analogous to before. Here, the upper simplifies to

$$\begin{aligned}\|f - \boldsymbol{A}_L(f)\|_{C(\mathcal{D}\times\mathcal{T})} &\leq C\left(\widetilde{C}_1\left(\mu^{\sigma-\frac{d}{2}+\tau-\frac{n}{2}}\right) + \mu^{\sigma-\frac{d}{2}-\frac{n}{2}} + \mu^{\tau-\frac{d}{2}-\frac{n}{2}}\right)^L \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} \\ &= C\left(\widetilde{C}_1\left(\mu^2 + 1 + \mu^{-1/2}\right)\right)^L \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}.\end{aligned}$$

However, in both cases, the numerical behavior observed in Figure 7.1 suggests that an error reduction holds as well.

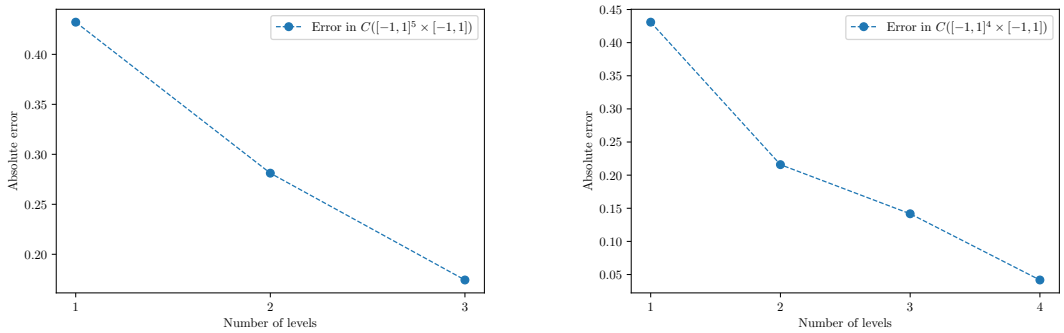


Figure 7.1: Multilevel method with tensorized kernels. The test function in both figures is given by $f(x,y) := \|x\|_2^{0.6}|y|^{1.6}$, which is an element of $H^{3,2}_{\mathrm{mix}}([-1,1]^5 \times [-1,1])$ and $H^{5/2,2}_{\mathrm{mix}}([-1,1]^4 \times [-1,1])$, respectively.

7.1.1 Properties of the method

We now want to generalize the ideas proven in Section 6.4 to the setting considered in this chapter. This can be accomplished in a similar manner as before. First, we observe that the multilevel approximation is again an element of

$$V_L = W_1 + \cdots + W_L,$$

where

$$W_j = \operatorname{span}\left\{\Phi_j(\cdot - x)\Psi_j(\cdot - y) \mid x \in X_j, y \in Y_j\right\}.$$

This sum is direct when appropriate kernels are chosen.

Corollary 7.1.8.
Let $X_1, \ldots, X_L \subseteq \mathcal{D}$ and $Y_1, \ldots, Y_L \subseteq \mathcal{T}$ be point sets with pairwise distinct fill distances $h_{X_j,\mathcal{D}} \neq h_{X_k,\mathcal{D}}$ and $h_{Y_j,\mathcal{T}} \neq h_{Y_k,\mathcal{T}}$ for all $1 \leq j \neq k \leq L$. Furthermore, let two Wendland kernels $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ and $\Psi \in L^1(\mathbb{R}^n) \cap C(\mathbb{R}^n)$ be given, with Hilbert spaces $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$, and $H_\Psi(\mathbb{R}^n) \cong H^\tau(\mathbb{R}^n)$, where $\tau > \frac{n}{2}$, respectively. Further assume that $\nu_\mathcal{D}, \nu_\mathcal{T} \geq 1$ are given. Define $\delta_j := \nu_\mathcal{D} h_{X_j,\mathcal{D}}$, $\kappa_j := \nu_\mathcal{T} h_{Y_j,\mathcal{T}}$, and the kernels Φ_j, Ψ_j as in Theorem 7.1.4. Then, the set of scaled and shifted tensorized kernel functions

$$\left\{\Phi_j(\cdot - x)\Psi_j(\cdot - y) \mid x \in X_j, y \in Y_j, 1 \leq j \leq L\right\}$$

is linearly independent. Hence, the sum

$$V_L = W_1 \oplus W_2 \oplus \cdots \oplus W_L$$

is direct.

Proof. By Corollary 6.4.1, we can see that both $\{\Phi_j(\cdot - x) \mid x \in X_j, 1 \leq j \leq L\}$ and $\{\Psi_j(\cdot - y) \mid y \in Y_j, 1 \leq j \leq L\}$ are linearly independent. Corollary 2.4.6 implies that

$$\{\Phi_j(\cdot - x)\Psi_k(\cdot - y) \mid x \in X_j, y \in Y_k, 1 \leq j, k \leq L\}$$

is linearly independent as well. However, this yields the claim, as independence stays valid when choosing a subset. □

Although Corollary 7.1.8 has only been proven for Wendland kernels up to now, we assume in the following that the occurring shifted and scaled kernel functions are also linearly independent. We then proceed as before with the observation that, for fixed $L \in \mathbb{N}$, we only use information at the points of $\boldsymbol{X} := \bigcup_{j=1}^{L}(X_j \times Y_j) \subseteq \mathcal{D} \times \mathcal{T}$. Hence, we can use a fixed ordering of this set and write again $\boldsymbol{A}_L$ as the composition $\boldsymbol{A}_L = \boldsymbol{\mathcal{A}}_L \circ T$, where

1. $T\colon C(\mathcal{D} \times \mathcal{T}) \to \mathbb{R}^{|\boldsymbol{X}|}, \quad f \mapsto f(\boldsymbol{X})$ is the *sampling* operator,
2. $\boldsymbol{\mathcal{A}}_L\colon \mathbb{R}^{|\boldsymbol{X}|} \to V_L$ is the *synthesis* operator.

This allows us to characterize the approximation space of $\boldsymbol{A}_L$.

Theorem 7.1.9.
In the setting of Theorem 7.1.4, we define, for fixed $L \in \mathbb{N}$,

$$\boldsymbol{X} = \bigcup_{j=1}^{L}(X_j \times Y_j), \quad N := |\boldsymbol{X}|.$$

Then we can write $\boldsymbol{A}_L = \boldsymbol{\mathcal{A}}_L \circ T$ as before. Moreover, T is surjective and $\boldsymbol{\mathcal{A}}_L$ is injective, so $\operatorname{Im}(\boldsymbol{A}_L) = \operatorname{Im}(\boldsymbol{\mathcal{A}}_L)$. This implies that the dimension of the approximation space is given by $\dim(\boldsymbol{A}_L(C(\mathcal{D} \times \mathcal{T}))) = N$ and a basis of $\boldsymbol{A}_L(C(\mathcal{D} \times \mathcal{T}))$ is given by

$$\{\boldsymbol{\mathcal{A}}_L(\boldsymbol{v}_j) \mid 1 \leq j \leq N\},$$

where $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\}$ denotes an arbitrary basis of $\mathbb{R}^N$.

Proof. The proof of this statement is analogous to the one of Theorem 6.4.2. □

With again the same proof as in the untensorized case, we can show the following representation as well:

Corollary 7.1.10.
Consider the setting of Theorem 7.1.4. Then the multilevel operator $\boldsymbol{A}_L$ can be written as

$$\boldsymbol{A}_L(f) = \sum_{\emptyset \neq \mathfrak{u} \subseteq \{1,\ldots,L\}} (-1)^{|\mathfrak{u}|+1} \prod_{j \in \mathfrak{u}} \mathcal{I}_{X_j \times Y_j, \Phi_j \otimes \Psi_j}(f),$$

where, as before, the product $\prod_{j \in \mathfrak{u}} \mathcal{I}_{X_j \times Y_j, \Phi_j \otimes \Psi_j}$ is sorted in descending order.

Proof. Apply Lemma 6.4.4 and proceed as in Corollary 6.4.5. □

Representations for nested sequences

Next, we present different basis representations of the approximation space in the special case that the point sets are nested, i.e,

$$X_1 \subseteq X_2 \subseteq \cdots \subseteq X_L, \quad Y_1 \subseteq Y_2 \subseteq \cdots \subseteq Y_L.$$

For the point sets, we use similar notations as in Section 6.4.2:

- For every $1 \leq j \leq L$, the point sets X_j, Y_j are given as

  $$X_j = \left\{x_1^{(j)}, x_2^{(j)}, \ldots, x_{N_j}^{(j)}\right\}, \quad Y_j = \left\{y_1^{(j)}, y_2^{(j)}, \ldots, y_{M_j}^{(j)}\right\},$$

  with $N_j, M_j \in \mathbb{N}$.

- Additionally, we assume that the first elements of X_j and Y_j are given by the ones in X_{j-1} and Y_{j-1}, respectively, i.e., we can write

  $$X_j = X_{j-1} \dot\cup \left\{x_{N_{j-1}+1}^{(j)}, \ldots, x_{N_j}^{(j)}\right\}, \quad Y_j = Y_{j-1} \dot\cup \left\{y_{M_{j-1}+1}^{(j)}, \ldots, y_{M_j}^{(j)}\right\}.$$

  As before, we suppress the upper index whenever there is no confusion. Moreover, for every $(x, y) \in X_L \times Y_L$, there exists a unique index $1 \leq k \leq L$ such that (x, y) is in $X_k \times Y_k \setminus X_{k-1} \times Y_{k-1}$, where $X_0 \times Y_0 := \emptyset$.

- We denote the Lagrange functions on $X_j \subseteq \mathcal{D}$ using the kernel Φ_j on level $1 \leq j \leq L$ by

  $$\chi_{\mathcal{D}}^{(j)} = \left(\chi_{1,\mathcal{D}}^{(j)}, \ldots, \chi_{N_j,\mathcal{D}}^{(j)}\right),$$

  and the ones on $Y_j \subseteq \mathcal{T}$ using the kernel Ψ_j for $1 \leq j \leq L$ by

  $$\chi_{\mathcal{T}}^{(j)} = \left(\chi_{1,\mathcal{T}}^{(j)}, \ldots, \chi_{M_j,\mathcal{T}}^{(j)}\right).$$

For convenience, we interpret the concatenation $\boldsymbol{A}_L = \boldsymbol{\mathcal{A}}_L \circ \boldsymbol{T}$ with a *matrix-valued* evaluation operator $\boldsymbol{T}$. Thus, we assume

- $\boldsymbol{T} \colon C(\mathcal{D} \times \mathcal{T}) \to \mathbb{R}^{N_L \times M_L}, \quad f \mapsto f(X_L, Y_L)$, which is defined as usual assuming an ordering of the point sets $X_L \subseteq \mathcal{D}$ and $Y_L \subseteq \mathcal{T}$, is the *sampling operator*.
- $\boldsymbol{\mathcal{A}}_L \colon \mathbb{R}^{N_L \times M_L} \to V_L$ is the *synthesis operator*.

With this setup, we can prove different representations of the multilevel scheme using tensorized kernels, leading to different basis representations of the approximation space.

Proposition 7.1.11.
Assume that Setting 7.1.1 holds, where the point sequences are additionally nested as above, and set $N_0 = M_0 := 0$. Then, we have the following basis representations:

1. *The* cardinal *representation: For every* $f \in H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D} \times \mathcal{T})$*, we can write*

$$\boldsymbol{A}_L(f) = \sum_{j=1}^{N_L} \sum_{k=1}^{M_L} f(x_j, y_k) b_{j,k},$$

where, for $(x_j, y_k) \in X_\ell \times Y_\ell \setminus X_{\ell-1} \times Y_{\ell-1}$*, the corresponding function* $b_{j,k}$ *is given by*

$$b_{j,k} = \boldsymbol{A}_L\left(\chi^{(L)}_{j,\mathcal{D}} \otimes \chi^{(L)}_{k,\mathcal{T}}\right) = \boldsymbol{A}_{[\ell+1:L]}\left(\chi^{(L)}_{j,\mathcal{D}} \otimes \chi^{(L)}_{k,\mathcal{T}}\right).$$

Here, $\boldsymbol{A}_{[\ell+1:L]}(f) := s_{\ell+1} + \cdots + s_L$ *is defined analogously to Proposition 6.4.6.*

2. *The* recursive *representation: For every* $f \in H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D} \times \mathcal{T})$*, we can write*

$$\begin{aligned}
\boldsymbol{A}_L(f)(x,y) = &\sum_{j=1}^{N_1} \sum_{k=1}^{M_1} f(x_j, y_k) \chi^{(1)}_{j,\mathcal{D}}(x) \chi^{(1)}_{k,\mathcal{T}}(y) \\
&+ \left(\sum_{j=1}^{N_1} \sum_{k=M_1+1}^{M_2} (f - \boldsymbol{A}_1(f))(x_j, y_k) \chi^{(2)}_{j,\mathcal{D}}(x) \chi^{(2)}_{k,\mathcal{T}}(y) \right. \\
&\quad \left. + \sum_{j=N_1+1}^{N_2} \sum_{k=1}^{M_2} (f - \boldsymbol{A}_1(f))(x_j, y_k) \chi^{(2)}_{j,\mathcal{D}}(x) \chi^{(2)}_{k,\mathcal{T}}(y) \right) \\
&+ \cdots + \\
&+ \left(\sum_{j=1}^{N_{L-1}} \sum_{k=M_{L-1}+1}^{M_L} (f - \boldsymbol{A}_L(f))(x_j, y_k) \chi^{(j)}_{j,\mathcal{D}}(x) \chi^{(k)}_{k,\mathcal{T}}(y) \right. \\
&\quad \left. + \sum_{j=N_{L-1}+1}^{N_L} \sum_{k=1}^{M_L} (f - \boldsymbol{A}_L(f))(x_j, y_k) \chi^{(L)}_{j,\mathcal{D}}(x) \chi^{(L)}_{k,\mathcal{T}}(y) \right).
\end{aligned}$$

Hence, the approximation space of $\boldsymbol{A}_L$ *is given by*

$$\boldsymbol{A}_1(C(\mathcal{D} \times \mathcal{T})) = \operatorname{span}\left\{ \chi^{(1)}_{j,\mathcal{D}} \otimes \chi^{(1)}_{k,\mathcal{D}} \mid 1 \le j \le N_1, 1 \le k \le M_1 \right\},$$

for $L = 1$*, and*

$$\begin{aligned}
\boldsymbol{A}_L(C(\mathcal{D} \times \mathcal{T})) = &\boldsymbol{A}_{L-1}(C(\mathcal{D} \times \mathcal{T})) \\
&\oplus \operatorname{span}\left\{ \chi^{(L)}_{j,\mathcal{D}} \otimes \chi^{(L)}_{k,\mathcal{T}} \mid 1 \le j \le N_{L-1}, M_{L-1} + 1 \le k \le M_L \right\} \\
&\oplus \operatorname{span}\left\{ \chi^{(L)}_{j,\mathcal{D}} \otimes \chi^{(L)}_{k,\mathcal{T}} \mid N_{L-1} + 1 \le j \le N_L, 1 \le k \le M_L \right\},
\end{aligned}$$

for $L \ge 2$*.*

Proof. We can prove the first case the same way as in the untensorized case using linearity. As we know that $\mathcal{I}_{X_L \times Y_L, \Phi_L \otimes \Psi_L}(f)$ interpolates f in $X_L \times Y_L$, we obtain

$$\boldsymbol{A}_L(f) = \boldsymbol{A}_L\left(\mathcal{I}_{X_L \times Y_L, \Phi_L \otimes \Psi_L}(f)\right) = \sum_{j=1}^{N_L} \sum_{k=1}^{M_L} f(x_j, y_k) \boldsymbol{A}_L\left(\chi^{(L)}_{j,\mathcal{D}} \otimes \chi^{(L)}_{k,\mathcal{T}}\right).$$

Similarly as in Proposition 6.4.6, we see that, for $(x_j, y_k) \in X_{\ell+1} \times Y_{\ell+1} \setminus X_\ell \times Y_\ell$, the cardinal functions $\chi^{(L)}_{j,\mathcal{D}} \otimes \chi^{(L)}_{k,\mathcal{T}}$ vanish on $X_\ell \times Y_\ell$, which leads to

$$\boldsymbol{A}_L \left(\chi^{(L)}_{j,\mathcal{D}} \otimes \chi^{(L)}_{k,\mathcal{T}} \right) = s_{\ell+1} + \cdots + s_L = \boldsymbol{A}_{[\ell+1:L]} \left(\chi^{(L)}_{j,\mathcal{D}} \otimes \chi^{(L)}_{k,\mathcal{T}} \right).$$

The second representation follows the ideas of Corollary 6.4.11. We first rewrite $\boldsymbol{A}_L(f)$ as

$$\boldsymbol{A}_L(f) = \mathcal{I}_{X_1 \times Y_1, \Phi_1 \otimes \Psi_1}(f) + \cdots + \mathcal{I}_{X_L \times Y_L, \Phi_L \otimes \Psi_L}(f - \boldsymbol{A}_{L-1}(f))$$

and insert the cardinal representations of the interpolation operator $\mathcal{I}_{X_j \times Y_j, \Phi_j \otimes \Psi_j}$

$$\begin{aligned}
\boldsymbol{A}_L(f)(x,y) = &\sum_{j=1}^{N_1} \sum_{k=1}^{M_1} f(x_j, y_k) \chi^{(1)}_{j,\mathcal{D}}(x) \chi^{(1)}_{k,\mathcal{T}}(y) \\
&+ \sum_{j=1}^{N_2} \sum_{k=1}^{M_2} (f - \boldsymbol{A}_1(f))(x_j, y_k) \chi^{(2)}_{j,\mathcal{D}}(x) \chi^{(2)}_{k,\mathcal{T}}(y) \\
&+ \cdots + \sum_{j=1}^{N_L} \sum_{k=1}^{M_L} (f - \boldsymbol{A}_{L-1}(f))(x_j, y_k) \chi^{(L)}_{j,\mathcal{D}}(x) \chi^{(L)}_{k,\mathcal{T}}(y).
\end{aligned}$$

Next, we use that $f - \boldsymbol{A}_\ell(f)$ vanishes on $X_\ell \times Y_\ell$ in this representation, which leads to

$$\begin{aligned}
\boldsymbol{A}_L(f)(x,y) = &\sum_{j=1}^{N_1} \sum_{k=1}^{M_1} f(x_j, y_k) \chi^{(1)}_{j,\mathcal{D}}(x) \chi^{(1)}_{k,\mathcal{T}}(y) \\
&+ \left(\sum_{j=1}^{N_1} \sum_{k=M_1+1}^{M_2} (f - \boldsymbol{A}_1(f))(x_j, y_k) \chi^{(2)}_{j,\mathcal{D}}(x) \chi^{(2)}_{k,\mathcal{T}}(y) \right. \\
&\quad \left. + \sum_{j=N_1+1}^{N_2} \sum_{k=1}^{M_2} (f - \boldsymbol{A}_1(f))(x_j, y_k) \chi^{(2)}_{j,\mathcal{D}}(x) \chi^{(2)}_{k,\mathcal{T}}(y) \right) \\
&+ \cdots + \\
&+ \left(\sum_{j=1}^{N_{L-1}} \sum_{k=M_{L-1}+1}^{M_L} (f - \boldsymbol{A}_{L-1}(f))(x_j, y_k) \chi^{(L)}_{j,\mathcal{D}}(x) \chi^{(L)}_{k,\mathcal{T}}(y) \right. \\
&\quad \left. + \sum_{j=N_{L-1}+1}^{N_L} \sum_{k=1}^{M_L} (f - \boldsymbol{A}_{L-1}(f))(x_j, y_k) \chi^{(L)}_{j,\mathcal{D}}(x) \chi^{(L)}_{k,\mathcal{T}}(y) \right).
\end{aligned}$$

The representation of the image of $\boldsymbol{A}_L$ follows immediately. □

7.1.2 Approximation of positive definite functions

In Section 5.2 and Section 6.5, we applied kernel-based approximation methods on a given random field. With the use of Lemma 5.1.2, we then concluded that the emerging covariance function of the approximated random field was given by the tensor product of the used approximation schemes. This observation motivated us to study approximation methods on mixed regularity spaces. Of course, one can also consider this approach the

other way around: Assume that a covariance function Cov of a random field is given, which we want to approximate using the multilevel method with tensorized kernels. Naturally, also $\boldsymbol{A}_L(\mathrm{Cov})$ should give a covariance function of a random field. Unfortunately, this does not seem to be the case. As the following numerical example suggests, the approximation of a positive definite function using the multilevel method with tensorized kernels does not lead to a positive semi-definite function. Thus, by Definition and Remark 4.1.4, this approximation cannot be a covariance function.
To see this, let f be a positive definite function and assume that $\boldsymbol{A}_L(f)$ is positive semi-definite. This would imply that the matrix $\boldsymbol{A}_L(f)(X, X)$ is positive semi-definite for all evaluation point sets $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ and thus only possesses non-negative eigenvalues. This property can be tested numerically.
To make this concrete, we set $\mathcal{D} = \mathcal{T} = [-1, 1]$ and choose for multilevel method the Wendland kernel $\Phi = \Psi = \Phi_{1,2}$. The target function is a Matérn kernel with parameter $\nu = 2.5$. We then choose X consisting of 800 randomly chosen points in $\mathcal{D}$ and calculate the smallest eigenvalue on different levels. We observe the following:

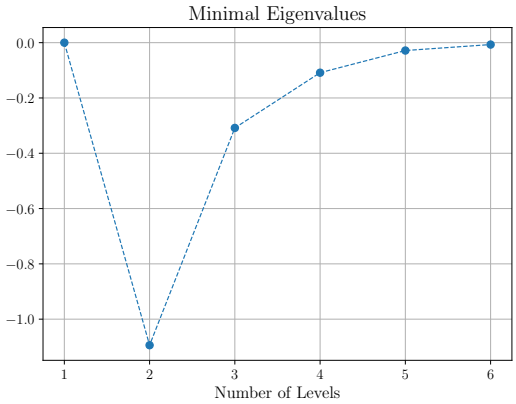


Figure 7.2: Minimal eigenvalues of $\boldsymbol{A}_L(f)(X, X)$ for $L \in \{1, \dots, 6\}$. On the first level, the multilevel method using the tensor kernel $\Phi_1 \otimes \Phi_1$ coincides with the tensor kernel interpolant in $X_1 \times X_1$. Thus, the operator is indeed positive definite.

We note that the eigenvalues of $\boldsymbol{A}_L(f)(X, X)$ tend to zero as L grows, but that there nonetheless still exist negative eigenvalues.
However, until now, we cannot prove or disprove analytically that the approximant $\boldsymbol{A}_L(f)$ is not positive semi-definite. As f is assumed to be positive definite, the integral operator of f possesses only positive eigenvalues λ_j for $j \in \mathbb{N}$. Lemma 5.3.5 allows us to bound the difference of the eigenvalues of f and the eigenvalues $\lambda_j^{(L)}$ of (the integral operator of) $\boldsymbol{A}_L(f)$ by

$$|\lambda_j - \lambda_j^{(L)}| \leq \|f - \boldsymbol{A}_L(f)\|_{L^2(\mathcal{D}\times\mathcal{D})} \leq C\alpha^L \|f\|_{H^{\sigma,\sigma}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{D})}.$$

Thus, we can choose a level L such that the right-hand side is less than or equal to $\varepsilon > 0$, and therefore all eigenvalues $\lambda_j^{(L)}$ are elements of $[\lambda_j - \varepsilon, \lambda_j + \varepsilon]$. However, as the eigenvalues λ_j tend to zero for $j \to \infty$, we cannot ensure that all eigenvalues $\lambda_j^{(L)}$ are positive or that a negative eigenvalue always exists.

7.2 Tensor product of multilevel methods

The other method we can consider was already mentioned beforehand, and is simply given as the tensor product of two multilevel methods on $\mathcal{D}$ and $\mathcal{T}$, respectively. This construction easily provides a convergence result.

Theorem 7.2.1 (Tensor product of multilevel methods)**.**
Assume Setting 7.1.1 holds, where the mesh norms of the point sets $(X_j)_{j\in\mathbb{N}}$ and $(Y_j)_{j\in\mathbb{N}}$ fulfill

$$\begin{aligned} h_{X_1,\mathcal{D}} > h_{X_2,\mathcal{D}} > \cdots > 0, \\ h_{Y_1,\mathcal{T}} > h_{Y_2,\mathcal{T}} > \cdots > 0. \end{aligned}$$

Additionally, assume there exist constants $c_\mathcal{D}, c_\mathcal{T} > 0$ and $\mu_\mathcal{D}, \mu_\mathcal{T} < 1$ such that

$$\begin{aligned} c_\mathcal{D}\mu_\mathcal{D} h_{X_j,\mathcal{D}} \leq & h_{X_{j+1},\mathcal{D}} \leq \mu_\mathcal{D} h_{X_j,\mathcal{D}}, \\ c_\mathcal{T}\mu_\mathcal{T} h_{Y_j,\mathcal{T}} \leq & h_{Y_{j+1},\mathcal{T}} \leq \mu_\mathcal{T} h_{Y_j,\mathcal{T}}. \end{aligned}$$

Now, choose $\nu_\mathcal{D}, \nu_\mathcal{T} \geq 1$ such that $\nu_\mathcal{D} h_{X_1,\mathcal{D}} \leq 1$ and $\nu_\mathcal{T} h_{Y_1,\mathcal{T}} \leq 1$. We set the scaling parameters as

$$\delta_j := \nu_\mathcal{D} h_{X_j,\mathcal{D}}, \quad \kappa_j := \nu_\mathcal{T} h_{Y_j,\mathcal{T}}, \quad j \in \mathbb{N},$$

and the sequence of kernels

$$\Phi_j := \Phi_{\delta_j}, \quad \Psi_j := \Psi_{\kappa_j}, \quad j \in \mathbb{N}$$

as in Lemma 6.1.1.
Denote by $A_{L_1,\mathcal{D}}$ the multilevel method on $X_1, \ldots, X_{L_1}$ using Φ, and by $A_{L_2,\mathcal{T}}$ the one on $Y_1, \ldots, Y_{L_2}$ using Ψ, given by Theorem 6.3.1. Then, for every $f \in H^{\sigma,\tau}_{\text{mix}}(\mathcal{D}\times\mathcal{T})$, we obtain the estimate

$$\begin{aligned} &\|f - A_{L_1,\mathcal{D}} \overline{\otimes} A_{L_2,\mathcal{T}}(f)\|_{L^2(\mathcal{D}\times\mathcal{T})} \\ &\leq C\left(\alpha_\mathcal{D}^{L_1}\|f\|_{H^{\sigma,0}_{\text{mix}}(\mathcal{D}\times\mathcal{T})} + \alpha_\mathcal{T}^{L_2}\|f\|_{H^{0,\tau}_{\text{mix}}(\mathcal{D}\times\mathcal{T})} + \alpha_\mathcal{D}^{L_1}\alpha_\mathcal{T}^{L_2}\|f\|_{H^{\sigma,\tau}_{\text{mix}}(\mathcal{D}\times\mathcal{T})}\right) \\ &\leq C\left(\alpha_\mathcal{D}^{L_1} + \alpha_\mathcal{T}^{L_2} + \alpha_\mathcal{D}^{L_1}\alpha_\mathcal{T}^{L_2}\right)\|f\|_{H^{\sigma,\tau}_{\text{mix}}(\mathcal{D}\times\mathcal{T})}, \end{aligned}$$

where

$$\alpha_\mathcal{D} = C_{1,\mathcal{D}}\left(\left(\frac{1}{\nu_\mathcal{D}}\right)^{2\sigma} + \mu_\mathcal{D}^{2\sigma}\right)^{\frac{1}{2}}, \quad \alpha_\mathcal{T} = C_{1,\mathcal{T}}\left(\left(\frac{1}{\nu_\mathcal{T}}\right)^{2\tau} + \mu_\mathcal{T}^{2\tau}\right)^{\frac{1}{2}},$$

with $C_{1,\mathcal{D}}, C_{1,\mathcal{T}} > 0$.
If we now additionally assume there are constants $\gamma_\mathcal{D}, \gamma_\mathcal{T} > 0$ such that

$$\frac{\gamma_\mathcal{D}}{\nu_\mathcal{D}} \leq \mu_\mathcal{D}, \quad \frac{\gamma_\mathcal{T}}{\nu_\mathcal{T}} \leq \mu_\mathcal{T}$$

holds, then we can also bound the error in $C(\mathcal{D}\times\mathcal{T})$ by

$$\begin{aligned} &\|f - A_{L_1,\mathcal{D}} \overline{\otimes} A_{L_2,\mathcal{T}}(f)\|_{C(\mathcal{D}\times\mathcal{T})} \\ &\leq C\left(\beta_\mathcal{D}^{L_1} + \beta_\mathcal{T}^{L_2} + \beta_\mathcal{D}^{L_1}\beta_\mathcal{T}^{L_2}\right)\|f\|_{H^{\sigma,\tau}_{\text{mix}}(\mathcal{D}\times\mathcal{T})}, \end{aligned}$$

where

$$\beta_{\mathcal{D}} = \widetilde{C}_{1,\mathcal{D}} \left(\left(\frac{1}{\nu_{\mathcal{D}}} \right)^{\sigma - \frac{d}{2}} + \mu_{\mathcal{D}}^{\sigma - \frac{d}{2}} \right), \quad \beta_{\mathcal{T}} = \widetilde{C}_{1,\mathcal{T}} \left(\left(\frac{1}{\nu_{\mathcal{T}}} \right)^{\tau - \frac{n}{2}} + \mu_{\mathcal{T}}^{\tau - \frac{n}{2}} \right).$$

Here, $\widetilde{C}_{1,\mathcal{D}} > 0$ depends on $\gamma_{\mathcal{D}}$ and $\widetilde{C}_{1,\mathcal{T}} > 0$ on $\gamma_{\mathcal{T}}$. This method is called the tensor product of multilevel methods.

Proof. The error estimate itself can be proven as in Theorem 3.5.6. We first use Lemma 3.5.5 to rewrite the operator as

$$\begin{aligned} &Id_{\mathcal{D}\times\mathcal{T}} - A_{L_1,\mathcal{D}} \overline{\otimes} A_{L_2,\mathcal{T}} \\ &= (Id_{\mathcal{D}} - A_{L_1,\mathcal{D}}) \overline{\otimes} Id_{\mathcal{T}} + Id_{\mathcal{D}} \overline{\otimes} (Id_{\mathcal{T}} - A_{L_2,\mathcal{T}}) - (Id_{\mathcal{D}} - A_{L_1,\mathcal{D}}) \overline{\otimes} (Id_{\mathcal{T}} - A_{L_2,\mathcal{T}}). \end{aligned}$$

With this, the triangle inequality, compatibility, and Theorem 6.3.1, we obtain

$$\begin{aligned} &\|f - A_{L_1,\mathcal{D}} \overline{\otimes} A_{L_2,\mathcal{T}}(f)\|_{L^2(\mathcal{D}\times\mathcal{T})} \\ &\leq \|(Id_{\mathcal{D}} - A_{L_1,\mathcal{D}}) \overline{\otimes} Id_{\mathcal{T}}(f)\|_{L^2(\mathcal{D}\times\mathcal{T})} + \|Id_{\mathcal{D}} \overline{\otimes} (Id_{\mathcal{T}} - A_{L_2,\mathcal{T}})(f)\|_{L^2(\mathcal{D}\times\mathcal{T})} \\ &\qquad + \|(Id_{\mathcal{D}} - A_{L_1,\mathcal{D}}) \overline{\otimes} (Id_{\mathcal{T}} - A_{L_2,\mathcal{T}})(f)\|_{L^2(\mathcal{D}\times\mathcal{T})} \\ &\leq C\alpha_{\mathcal{D}}^{L_1} \|f\|_{H^{\sigma,0}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} + C\alpha_{\mathcal{T}}^{L_2} \|f\|_{H^{0,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} + C\alpha_{\mathcal{D}}^{L_1}\alpha_{\mathcal{T}}^{L_2} \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})} \\ &\leq C \left(\alpha_{\mathcal{D}}^{L_1} + \alpha_{\mathcal{T}}^{L_2} + \alpha_{\mathcal{D}}^{L_1}\alpha_{\mathcal{T}}^{L_2} \right) \|f\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\mathcal{T})}. \end{aligned}$$

Again, the error in $C(\mathcal{D} \times \mathcal{T})$ is proven the same way, using Corollary 6.3.4. □

Remark 7.2.2.
If we assume $L_1 = L_2 =: L$, then a simple calculation proves that

$$\begin{aligned} &\alpha_{\mathcal{D}}^{L} + \alpha_{\mathcal{T}}^{L} + \alpha_{\mathcal{D}}^{L}\alpha_{\mathcal{T}}^{L} \\ &\leq \left[\max\left\{ C_{1,\mathcal{D}}, C_{1,\mathcal{T}}, C_{1,\mathcal{D}}C_{1,\mathcal{T}} \right\} \left(\mu_{\mathcal{D}}^{\sigma} + \nu_{\mathcal{D}}^{-\sigma} + \mu_{\mathcal{T}}^{\tau} + \nu_{\mathcal{T}}^{-\tau} + \left(\mu_{\mathcal{D}}^{\sigma} + \nu_{\mathcal{D}}^{-\sigma} \right) \left(\mu_{\mathcal{T}}^{\tau} + \nu_{\mathcal{T}}^{-\tau} \right) \right) \right]^{L}, \end{aligned}$$

which, up to constants, equals the convergence enforcing term of the multilevel method using tensorized kernels in $L^2(\mathcal{D} \times \mathcal{T})$ given in Theorem 7.1.4.

Similarly as in Corollary 3.5.4, we can use the basis representation (6.9) given in Remark 6.4.3 to derive an explicit representation of this method:

Corollary 7.2.3.
Assume the setting of Theorem 7.2.1 holds. We write

$$X := \bigcup_{j=1}^{L_1} X_j = \{x_1, \ldots, x_N\}, \quad Y = \bigcup_{j=1}^{L_2} Y_j = \{y_1, \ldots, y_M\},$$

and denote by $b_{j,\mathcal{D}}$ for $1 \leq j \leq N$ and $b_{j,\mathcal{T}}$ for $1 \leq j \leq M$ the basis functions of $A_{L_1,\mathcal{D}}$ and $A_{L_2,\mathcal{T}}$ given in (6.9)*. Then the tensor product of multilevel method can be represented as*

$$A_{L_1,\mathcal{D}} \overline{\otimes} A_{L_2,\mathcal{T}}(f) = \sum_{j=1}^{N} \sum_{k=1}^{M} f(x_j, y_k) b_{j,\mathcal{D}} b_{k,\mathcal{T}}$$

for all $f \in C(\mathcal{D} \times \mathcal{T})$.

Proof. For elementary tensors $v = v_1 \otimes v_2 \in C(\mathcal{D}) \otimes C(\mathcal{T})$, this representation is obviously true, and extends to all elements of $C(\mathcal{D}) \otimes C(\mathcal{T})$ by linearity. For an arbitrary element $f \in C(\mathcal{D} \times \mathcal{T})$, we can find a sequence $(f_j)_{j\in\mathbb{N}} \subseteq C(\mathcal{D}) \otimes C(\mathcal{T})$ that converges to f pointwise and extends the representation to $A_{L_1}\overline{\otimes}A_{L_2}(f)$. □

7.3 Comparison of the methods

We now want to compare both methods presented earlier. For this, we assume that $L_1 = L_2 =: L$ in Theorem 7.2.1 holds and the point sets $(X_j)_{j\in\mathbb{N}} \subseteq \mathcal{D}$ and $(Y_j)_{j\in\mathbb{N}} \subseteq \mathcal{T}$ are chosen such that Theorem 7.1.4 and Theorem 7.2.1 are applicable. Note that for both methods equal assumptions on the point sets are demanded. We also choose all occurring parameters to be the same.
First, we want to rigorously prove that $\boldsymbol{A}_L \neq A_{L,\mathcal{D}}\overline{\otimes}A_{L,\mathcal{T}}$ on $H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \mathcal{T})$, whenever $L \geq 2$. Of course, we already have a result whenever the dimension of the images is different, but we can state a universally applicable result.

Corollary 7.3.1.
Assume Setting 7.1.1 holds, with $\nu_{\mathcal{D}}, \nu_{\mathcal{T}} \geq 1$ properly chosen such that Theorem 7.1.4 and Theorem 7.2.1 are both applicable, and let $L \geq 2$. Then, the multilevel method with tensorized kernels $\boldsymbol{A}_L$ does not coincide with the tensor product of the multilevel methods $A_{L,\mathcal{D}}\overline{\otimes}A_{L,\mathcal{T}}$ on $C(\mathcal{D} \times \mathcal{T})$.

Proof. First, we observe that, for every $f \in C(\mathcal{D} \times \mathcal{T})$, the multilevel method using tensorized kernels can be written as

$$\boldsymbol{A}_L(f) = s_1 + \cdots + s_L = \mathcal{I}_{X_1\times Y_1,\Phi_1\otimes\Psi_1}(f) + \cdots + \mathcal{I}_{X_L\times Y_L,\Phi_L\otimes\Psi_L}(f - \boldsymbol{A}_{L-1}(f)),$$

which implies that $\boldsymbol{A}_L(f)$ is an element of

$$\mathcal{B} := \operatorname{span}\{\Phi_j(\cdot - x)\Psi_j(\cdot - y) \mid x \in X_j, y \in Y_j, 1 \leq j \leq L\}.$$

Our goal is to construct a function f such that $A_{L,\mathcal{D}}\overline{\otimes}A_{L,\mathcal{T}}(f) \neq \boldsymbol{A}_L(f)$. To accomplish this, we choose $x^* \in X_1$ and $y^* \in Y_2 \setminus Y_1$ fixed (note that by assumption $h_{Y_1,\mathcal{T}} > h_{Y_2,\mathcal{T}}$, thus such an element exists) and a function $f_1 \in C(\mathcal{D})$ that vanishes on $X_1 \setminus \{x^*\}$ and satisfies $f_1(x^*) = 1$, and a second function $f_2 \in C(\mathcal{T})$ that vanishes on $Y_1 \cup Y_2 \setminus \{y^*\}$ and fulfills $f_2(y^*) = 1$. Then we set $f := f_1 \otimes f_2$. We want to prove that $A_{L\mathcal{D}}\overline{\otimes}A_{L,\mathcal{T}}$ has a non-vanishing term $\Phi_1(\cdot - x^*)\Psi_2(\cdot - y^*)$ and thus cannot be represented as a linear combination of elements in $\mathcal{B}$.
As f is an elementary tensor, the tensor product of multilevel methods decomposes into $A_{L,\mathcal{D}}\overline{\otimes}A_{L,\mathcal{T}}(f) = A_{L,\mathcal{D}}(f_1)A_{L,\mathcal{T}}(f_2)$, which allows us to examine both terms separately. The multilevel method in $\mathcal{D}$ can be rewritten as

$$\begin{aligned} A_{L,\mathcal{D}}(f_1) &= \mathcal{I}_{X_1,\Phi_1}(f_1) + \cdots + \mathcal{I}_{X_L,\Phi_L}(f_1 - A_{L-1,\mathcal{D}}(f_1)) \\ &= \mathcal{I}_{X_1,\Phi_1}(f_1) + \sum_{j=2}^{L} \mathcal{I}_{X_j,\Phi_j}(f_1 - A_{j-1,\mathcal{D}}(f_1)). \end{aligned}$$

The latter term is an element of $\operatorname{span}\{\Phi_j(\cdot - x) \mid 2 \leq j \leq L, x \in X_j\}$ and will be denoted by g_1. The first one is just given as the Lagrange function centered in $x^* \in X_1$ using

Φ_1 and is therefore a non-vanishing element of span $\{\Phi_1(\cdot - x) \mid x \in X_1\}$. For the multilevel method $A_{L,\mathcal{T}}(f_2)$, we use the same decomposition as before. Here, we observe that $\mathcal{I}_{Y_1,\Psi_1}(f_2) = 0$ since f_2 vanishes on Y_1. Thus, we obtain the representation

$$\begin{aligned} A_{L,\mathcal{T}}(f_2) &= \mathcal{I}_{Y_2,\Psi_2}(f_2 - A_{1,\mathcal{T}}(f_2)) + \cdots + \mathcal{I}_{Y_L,\Psi_L}(f_2 - A_{L-1,\mathcal{T}}(f_2)) \\ &= \mathcal{I}_{Y_2,\Psi_2}(f_2) + g_2, \end{aligned}$$

where, similarly to before, g_2 denotes the remainder of the multilevel method and is given as an element of span $\{\Psi_j(\cdot - y) \mid 3 \leq j \leq L, y \in Y_j\}$.
Again, the first term yields the Lagrange function in Y_2 centered at y^*, i.e., is given as a non-vanishing element of span $\{\Psi_2(\cdot - y) \mid y \in Y_2\}$. With this, the tensor product of both methods is given as

$$\begin{aligned} A_{L,\mathcal{D}}\overline{\otimes}A_{L,\mathcal{T}}(f) &= \left(\mathcal{I}_{X_1,\Phi_1}(f_1) + g_1\right)\left(\mathcal{I}_{Y_2,\Psi_2}(f_2) + g_2\right) \\ &= \mathcal{I}_{X_1,\Phi_1}(f_1)\mathcal{I}_{Y_2,\Psi_2}(f_2) + \mathcal{I}_{X_1,\Phi_1}(f_1)g_2 + g_1\mathcal{I}_{Y_2,\Psi_2}(f_2) + g_1 g_2. \end{aligned}$$

Here, only the first term contains elements in $\{\Phi_1(\cdot - x)\Psi_2(\cdot - y) \mid x \in X_1, y \in Y_2\}$, and at least one of them is non-vanishing, as $\mathcal{I}_{X_1,\Phi_1}(f_1)\mathcal{I}_{Y_2,\Psi_2}(f_2) \neq 0$, by construction. Thus, the other occurring terms cannot cancel the first term by linear independence, which yields that $A_{L,\mathcal{D}}\overline{\otimes}A_{L,\mathcal{T}}(f)$ consists of at least one element which is not in $\mathcal{B}$. Therefore, $A_{L,\mathcal{D}}\overline{\otimes}A_{L,\mathcal{T}}(f) \neq \boldsymbol{A}_L(f)$ and the methods do not coincide. □

When comparing the convergence results, we can see that both methods converge linearly to the target function. Moreover, as already seen in Remark 7.2.2, we can estimate the convergence enforcing term of the tensor product of multilevel methods by the one of the multilevel method using tensor kernels, i.e.,

$$\begin{aligned} &\alpha_{\mathcal{D}}^L + \alpha_{\mathcal{T}}^L + \alpha_{\mathcal{D}}^L\alpha_{\mathcal{T}}^L \\ &\leq \left[C\left(\mu_{\mathcal{D}}^{\sigma} + \nu_{\mathcal{D}}^{-\sigma} + \mu_{\mathcal{T}}^{\tau} + \nu_{\mathcal{T}}^{-\tau} + \left(\mu_{\mathcal{D}}^{\sigma} + \nu_{\mathcal{D}}^{-\sigma}\right)\left(\mu_{\mathcal{T}}^{\tau} + \nu_{\mathcal{T}}^{-\tau}\right)\right)\right]^L, \end{aligned}$$

in the $L^2(\mathcal{D} \times \mathcal{T})$ case. We expect the convergence rate to be similar, especially if $\mathcal{D} = \mathcal{T}$, which is also supported by the numerical examples given in Figure 7.3 and Figure 7.4:
On the left-hand side of both figures, the domains $\mathcal{D} = [-1,1]^3$ and $\mathcal{T} = [-1,1]$ are used. For the interpolation we chose the Wendland kernels $\Phi = \Phi_{3,0}$, with native space equivalent to $H^2(\mathbb{R}^3)$, and $\Psi = \Phi_{1,1}$, with native space equivalent to $H^2(\mathbb{R})$. The target function is given by $f(x,y) := \|x\|_2^{3/4}|y|^{7/4} \in H^{2,2}_{\text{mix}}([-1,1]^3 \times [-1,1])$.
On the right-hand side, we chose the domains $\mathcal{D} = \mathcal{T} = [-1,1]^2$, with Wendland kernels $\Phi = \Psi = \Phi_{3,1}$ and native space equivalent to $H^{5/2}(\mathbb{R}^2)$. The target function is given by $f(x,y) = \|x\|_2^{1.6}\|y\|^{1.6} \in H^{5/2,5/2}_{\text{mix}}([-1,1]^2 \times [-1,1]^2)$. In both cases, the point sets are uniform refining grids on $\mathcal{D} \times \mathcal{T}$, where $\mu_{\mathcal{D}} = \mu_{\mathcal{T}} = \frac{1}{2}$ and accordingly $\nu_{\mathcal{D}} = \nu_{\mathcal{T}} = 2$ was chosen.

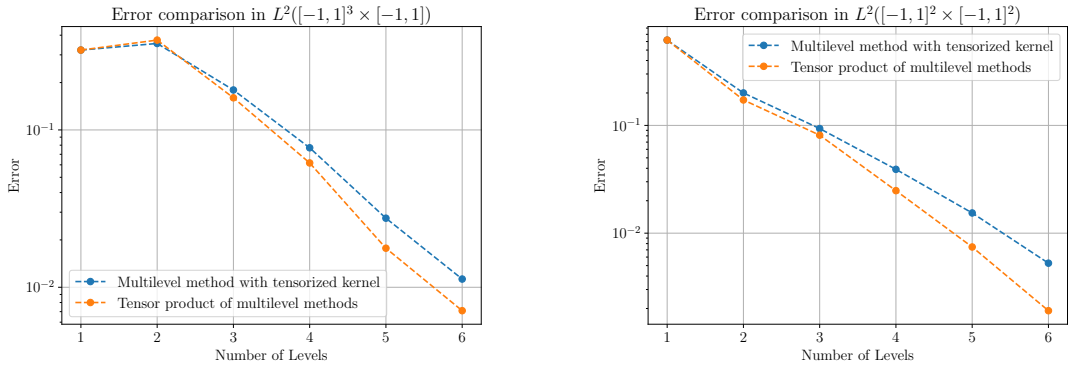


Figure 7.3: Comparison of the error decay in $L^2(\mathcal{D} \times \mathcal{T})$.

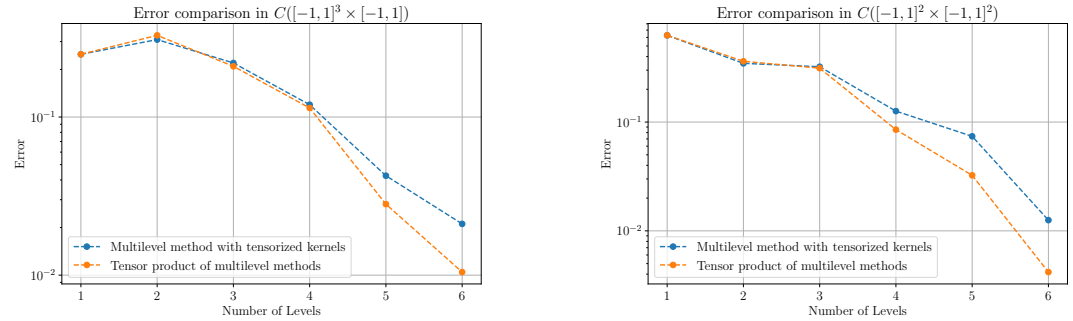


Figure 7.4: The absolute error comparison of both approximations considered in Figure 7.3.

The main difference of both methods are pointwise error bounds, as for the tensor product multilevel method a pointwise error estimate in Theorem 7.2.1 is easily concluded, while for the multilevel method using tensor kernels, we have to assume more restrictive assumptions, as stated in Corollary 7.1.7. Nonetheless, there seems to hold a convergence result with weaker assumptions, as Figure 7.1 suggests.

7.3.1 Computational and evaluation cost

In order to compare the asymptotic computational costs, we denote the number of points by $N_j := |X_j|$ and $M_j := |Y_j|$, for $j \in \mathbb{N}$. For simplicity, we assume in the following that the point sets are also nested. We first want to determine the computational cost to calculate the approximation, followed by the cost for evaluating the given approximation in one point $(x, y) \in \mathcal{D} \times \mathcal{T}$.

Multilevel method using tensorized kernels

Following the description given in Algorithm 2, we calculate in every step $1 \leq j \leq L$ the kernel interpolant $\mathcal{I}_{X_j \times Y_j, \Phi_j \otimes \Psi_j}(e_{j-1})$. Then, for the residual step, we have to evaluate this interpolant in all points of $X_L \times Y_L$.

In order to calculate the cost for one step j, we proceed as described in Section 3.5.1 for the calculation of the interpolation, i.e., we compute the coefficient matrix given by $C_j = \Phi_j(X_j, X_j)^{-1} e_{j-1}(X_j, Y_j) \Psi_j(Y_j, Y_j)^{-1}$ in a total of $\mathcal{O}(N_j M_j^3 + M_j N_j^3)$ time. For the residual, on the other hand, we have to calculate

$$e_j(X_L, Y_L) = e_{j-1}(X_L, Y_L) - \Phi_j(X_L, X_j) C_j \Psi_j(Y_j, Y_L). \tag{7.10}$$

By recursion, we assume that $e_{j-1}(X_L, Y_L)$ is already known and was computed in the previous step. The transition matrices $\Phi_j(X_L, X_j) \in \mathbb{R}^{N_L \times N_j}$ and $\Psi_j(Y_j, Y_L) \in \mathbb{R}^{M_j \times M_L}$ can be constructed in $\mathcal{O}(N_j\, N_L)$ and $\mathcal{O}(M_j\, M_L)$ time, respectively. Thus, the computation of the matrix product on the right-hand side of (7.10) can be calculated, depending on the order of multiplications, in $\mathcal{O}(N_j\, M_j\, N_L + M_j\, M_L\, N_L)$ or $\mathcal{O}(M_j\, N_j\, M_L + N_j\, M_L\, N_L)$ time, respectively. The difference step in (7.10) can be calculated in $\mathcal{O}(N_L\, M_L)$ time, which leads to a total cost of $\mathcal{O}\left(\min\{N_j\, M_j\, N_L + M_j\, M_L\, N_L,\ M_j\, N_j\, M_L + N_j\, M_L\, N_L\}\right)$ for the residual per level. The total cost of this method can now be obtained by adding up the individual costs per level and is therefore given as

$$\begin{aligned} \operatorname{cost}(\boldsymbol{A}_L) = \mathcal{O}\Bigg(& \sum_{j=1}^{L} (N_j\, M_j^3 + M_j\, N_j^3) \\ & + \sum_{j=1}^{L} \min\{N_j\, M_j\, N_L + M_j\, M_L\, N_L,\ M_j\, N_j\, M_L + N_j\, M_L\, N_L\} \Bigg). \end{aligned} \tag{7.11}$$

To analyze the evaluation cost for one point $(x, y) \in \mathcal{D} \times \mathcal{T}$, we have to evaluate the interpolants $\mathcal{I}_{X_j \times Y_j, \Phi_j \otimes \Psi_j}(e_{j-1})$ in $(x, y) \in \mathcal{D} \times \mathcal{T}$, for $1 \leq j \leq L$. Having done this, we have to add up the results. To calculate the evaluation per level, we have to construct the vectors $\Phi_j(x, X_j) \in \mathbb{R}^{N_j}$ and $\Psi_j(Y_j, y) \in \mathbb{R}^{M_j}$, which can be done in $\mathcal{O}(N_j)$ and $\mathcal{O}(M_j)$ time, respectively. By assumption, the matrix $C_j \in \mathbb{R}^{N_j \times M_j}$ has already been computed. The point evaluation on this level can be obtained by

$$\mathcal{I}_{X_j \times Y_j, \Phi_j \otimes \Psi_j}(x, y) = \Phi_j(x, X_j) C_j \Psi_j(Y_j, y),$$

i.e., we have to calculate the given vector-matrix-vector product. This can be accomplished in $\mathcal{O}(N_j\, M_j + M_j)$ or $\mathcal{O}(N_j\, M_j + N_j)$ time, respectively, depending on the order of multiplication. However, both costs are dominated by $\mathcal{O}(N_j\, M_j)$, which leads to a total asymptotical evaluation time of

$$\mathcal{O}\left(\sum_{j=1}^{L} N_j\, M_j \right)$$

for one point evaluation.

Tensor product of multilevel methods

For the second method, we use the Lagrange type representation proven in Corollary 7.2.3, which is given as

$$A_{L,\mathcal{D}} \overline{\otimes} A_{L,\mathcal{T}}(f)(x, y) = \sum_{k=1}^{N_L} \sum_{\ell=1}^{M_L} f(x_k, y_\ell) b_{k,\mathcal{D}}(x) b_{\ell,\mathcal{T}}(y),$$

with $X_L = \{x_1, \ldots, x_{N_L}\}$ and $Y_L = \{y_1, \ldots, y_{M_L}\}$. Therefore, the computational cost is given by the construction of all basis functions $b_{k,\mathcal{D}}$, for $1 \leq k \leq N_L$, and $b_{\ell,\mathcal{T}}$, for $1 \leq \ell \leq M_L$, respectively. Following the ideas developed in Section 6.4, a basis function $b_{k,\mathcal{D}}$ can be calculated by the multilevel method $A_{L,\mathcal{D}}$ itself using a proper right-hand side (given by the unit vector $\boldsymbol{e}_k \in \mathbb{R}^{N_L}$). We therefore have to calculate the computational cost of this scheme. Similar to before, the cost for one level $1 \leq j \leq L$ of $A_{L,\mathcal{D}}$ consists of the kernel interpolation step and the residual. For the kernel interpolation, we have to solve one linear equation system with the kernel matrix $\Phi_j(X_j, X_j)$. This matrix can be constructed in $\mathcal{O}(N_j^2)$ time, and the system of linear equations can be solved in $\mathcal{O}(N_j^3)$ time using a CG-solver. Afterwards, we have to calculate the residual on all points in X_L, which yields a total computation time of $\mathcal{O}(N_L\, N_j)$ for the occurring matrix vector product, with additional $\mathcal{O}(N_L\, N_j)$ time to construct the transition matrix $\Phi_j(X_L, X_j)$, and $\mathcal{O}(N_L)$ time to perform the difference step. All together, the computational cost for one step of the calculation of one basis function is given by $\mathcal{O}(N_j^3 + N_L\, N_j)$. To calculate one basis function $b_{k,\mathcal{D}}$, we have to add the cost of all steps $1 \leq j \leq L$. As we have to proceed this way for every basis function $b_{k,\mathcal{D}}$ with $1 \leq k \leq N_L$, the total cost for computing all basis functions is thus given by

$$\mathcal{O}\left(N_L \left(\sum_{j=1}^{L} N_j^3 + N_L \sum_{j=1}^{L} N_j\right)\right).$$

Of course, the same considerations hold for the computational cost of the other basis functions $b_{\ell,\mathcal{T}}$, for $1 \leq \ell \leq M_L$, which leads to a total cost of

$$\mathcal{O}\left(N_L \left(\sum_{j=1}^{L} N_j^3 + N_L \sum_{j=1}^{L} N_j\right) + M_L \left(\sum_{j=1}^{L} M_j^3 + M_L \sum_{j=1}^{L} M_j\right)\right). \tag{7.12}$$

In order to evaluate the approximation in one point $(x, y) \in \mathcal{D} \times \mathcal{T}$, we first have to evaluate all basis functions $b_{k,\mathcal{D}}$, for $1 \leq k \leq N_L$, in x, all functions $b_{\ell,\mathcal{T}}$, for $1 \leq \ell \leq M_L$, in y, and then compute the total approximation using the formula given in Corollary 7.2.3. The cost of evaluating one basis function $b_{k,\mathcal{D}}$ in x is again calculated levelwise. On each level $1 \leq j \leq L$, we construct the vector $\Phi_j(x, X_j)$ in $\mathcal{O}(N_j)$ time. The evaluation of the kernel interpolation is then simply given by a dot product with the corresponding solution vector, which can also be evaluated in $\mathcal{O}(N_j)$ time. Summing over all levels yields the evaluation cost for one basis function, and is given by

$$\mathcal{O}\left(N_L \sum_{j=1}^{L} N_j\right).$$

As before, the same arguments also hold for the basis functions $b_{\ell,\mathcal{T}}$, for $1 \leq \ell \leq M_L$. The evaluation formula of Corollary 7.2.3, which costs a total of $\mathcal{O}(N_L M_L)$ time when all coefficients are known, leads to a total evaluation time of

$$\mathcal{O}\left(N_L \sum_{j=1}^{L} N_j + M_L \sum_{j=1}^{L} M_j + N_L\, M_L\right).$$

Comparison of the computational costs

When comparing both methods in terms of computation and evaluation time, the first major difference is that the asymptotic computation time for the multilevel method using tensor kernels given in (7.11) couples the number of points N_j and M_j used in each step, while for the tensor product of multilevel methods these values are strictly separated. This suggests that it is advantageous to use the multilevel method using tensorized kernels whenever one of the quantities is significantly smaller, i.e., if, for example, $M_j \ll N_j$ for $1 \leq j \leq L$, which may occur whenever the dimensions d and n differ, in this example $n \ll d$. In this case, if additionally $N_j M_j^3 \leq M_j N_j^3$, the costs given in (7.11) are dominated by

$$\operatorname{cost}(\boldsymbol{A}_L) = \mathcal{O}\left(L M_L N_L^3\right),$$

while the cost of the tensor product of multilevel methods given in (7.12) is dominated by

$$\operatorname{cost}(A_{L,\mathcal{D}} \overline{\otimes} A_{L,\mathcal{T}}) = \mathcal{O}\left(L N_L^4\right).$$

On the other hand, when $N_j \approx M_j$ or particularly $N_j = M_j$ holds, there is no asymptotic advantage of one over the other.
The same observation also seems to hold when considering the evaluation time in one point. For the multilevel method using tensor kernels, the evaluation time is dominated by $\mathcal{O}(L M_L N_L)$, while the one for the tensor product of multilevel methods satisfies $\mathcal{O}(L(N_L^2 + M_L^2))$. Again, the inherent structure of the first method provides an advantage whenever the number of points N_j and M_j used per level differ significantly, while both are asymptotically the same when $N_j \approx M_j$ for all $1 \leq j \leq L$.

8 Application to PDEs with randomness

8.1 Formalizing the problem

In this chapter, we head back to the motivational problem in a more rigorous manner, before we apply the techniques introduced so far. The fundamental ideas used in this chapter can, for example, be found in [35, Chapter 9].
We consider the following setting: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{D} \subseteq \mathbb{R}^d$ a bounded domain (the *spatial domain*). The dimension usually fulfills $d = 2$ or $d = 3$, however, we just demand $d \in \mathbb{N}$. Assume that a random field $a\colon \Omega \times \mathcal{D} \to \mathbb{R}$ and a right-hand side $f\colon \mathcal{D} \to \mathbb{R}$ are given. We consider the following second order elliptic partial differential equation with Dirichlet boundary condition, including randomness:
Find $u\colon \Omega \times \mathcal{D} \to \mathbb{R}$, which solves the following equation:

$$\begin{aligned} -\operatorname{div}_x\left(a(\omega,x)\nabla_x u(\omega,x)\right) &= f(x), \qquad x \in \mathcal{D}, \\ u(\omega,x) &= 0, \qquad x \in \partial\mathcal{D}, \end{aligned} \tag{8.1}$$

for all $\omega \in \Omega$.
The occurring random field a is also called *diffusion coefficient*, as already mentioned in the introduction. In order to derive classical solutions of (8.1), one has to assume classical differentiability of the paths of a to state a well-defined problem. However, a typical approach to solve this equation are *weak formulations* of the problem: Assume for the time being that u denotes a suitable, i.e., at least differentiable solution of equation (8.1). Our goal is to find a reformulation of the differential equation that only demands weak differentiability of the solution. In the literature, there are mainly two different approaches, which differ in the treatment of the random parameter $\omega \in \Omega$:
In the *pathwise weak formulation*, we fix $\omega \in \Omega$, multiply the first equation of (8.1) by a test function $v \in C^1(\mathcal{D})$, and integrate the resulting equation over $\mathcal{D}$. On the left-hand side, we can then apply the integration by parts formula. As u solves equation (8.1) and thus, in particular, vanishes on the boundary $\partial\mathcal{D}$, the boundary term occurring in the integration by parts formula also vanishes. This leads to the equation

$$\int_{\mathcal{D}} a(\omega,x)\nabla_x u(\omega,x)\nabla_x v(\omega,x)\,dx = \int_{\mathcal{D}} f(x)v(x)\,dx.$$

This formulation is well-defined when $f \in L^2(\mathcal{D})$, $u(\omega,\cdot)$ and v are elements of $H^1(\mathcal{D})$, and the diffusion coefficient is (uniformly) bounded, i.e., there exists a constant $a_{max} > 0$ such that

$$|a(\omega,x)| \le a_{max}, \quad x \in \mathcal{D}, \quad \omega \in \Omega, \tag{8.2}$$

holds. To include the Dirichlet boundary condition, we additionally restrict the function space $H^1(\mathcal{D})$ to $H_0^1(\mathcal{D})$, which leads to the pathwise weak formulation:

Find for every $\omega \in \Omega$ a solution $u(\omega, \cdot) \in H_0^1(\mathcal{D})$ that satisfies, for all $v \in H_0^1(\mathcal{D})$, the equation

$$B_1(u(\omega,\cdot),v)(\omega) := \int_{\mathcal{D}} a(\omega,x)\nabla_x u(\omega,x)\cdot\nabla_x v(x)\,dx = \int_{\mathcal{D}} f(x)v(x)\,dx. \tag{8.3}$$

In short, this equation can be written as

$$B_1(u(\omega,\cdot),v)(\omega) = \langle f, v\rangle_{L^2(\mathcal{D})}, \quad v \in H_0^1(\mathcal{D}), \quad \omega \in \Omega.$$

The *fully weak formulation* uses similar ideas, with the main difference that the integration of (8.1) is both over the spatial domain $\mathcal{D}$ and the probability space Ω. This leads to test functions v as elements of the Bochner space $L^2(\Omega; H_0^1(\mathcal{D}))$ and to the following formulation:
Find a solution $u \in L^2(\Omega; H_0^1(\mathcal{D}))$ that satisfies, for all $v \in L^2(\Omega; H_0^1(\mathcal{D}))$, the equation

$$B_2(u,v) := \int_\Omega \int_{\mathcal{D}} a(\omega,x)\nabla_x u(\omega,x)\cdot\nabla_x v(\omega,x)\,dx\,d\mathbb{P}(\omega) = \int_\Omega \int_{\mathcal{D}} f(x)v(\omega,x)\,dx\,d\mathbb{P}(\omega). \tag{8.4}$$

In principle, one could also allow $f \in L^2(\Omega; L^2(\mathcal{D}))$ in (8.4), but we will not consider this case. In order for B_2 to be well-defined, we once again have to assume that the diffusion coefficient a is (uniformly) bounded, i.e., fulfills (8.2). We could relax this assumption (see [35, Assumption 9.5]), but, for simplicity, this will not be considered here. In order to ensure the existence of a unique solution of (8.3) and (8.4), respectively, we will use the Lemma of Lax-Milgram (see [15, Chapter 6.2.1]):

Theorem 8.1.1.
Let $(H, \|\cdot\|_H)$ be a Hilbert space and $B\colon H \times H \to \mathbb{R}$ a bilinear form, which fulfills the following two assumptions:

(i) B is bounded, i.e., there exists a constant $C > 0$ such that for all $u, v \in H$ the estimate

$$|B(u,v)| \le C\|u\|_H\|v\|_H$$

holds;

(ii) B is coercive, i.e., there exists an $\eta > 0$ such that for all $u \in H$ the estimate

$$B(u,u) \ge \eta\|u\|_H^2$$

holds.

Then, for every linear and bounded functional $g \in H^*$, there exists a unique element $u \in H$ such that, for all $v \in H$,

$$B(u,v) = g(v)$$

holds. Moreover, the norm of u satisfies the stability estimate

$$\|u\|_H \le \frac{1}{\eta}\|g\|_{H^*}.$$

In order to apply this theorem to (8.3) and (8.4), we have to ensure that the bilinear forms B_1 and B_2 are bounded and coercive. Again, this can be achieved with another assumption on the diffusion coefficient. From now on, we work with the following assumption:

Assumption 8.1.2.
The diffusion coefficient $a \in L^2(\Omega; L^2(\mathcal{D}))$ is a second order random field. Additionally, there are constants $0 < a_{min} \leq a_{max}$ such that, for $\mathbb{P}$-almost all $\omega \in \Omega$ and $x \in \mathcal{D}$, the estimate

$$0 < a_{min} \leq a(\omega, x) \leq a_{max}$$

holds.

With this we can prove the following:

Lemma 8.1.3.
Let $f \in L^2(\mathcal{D})$ be fixed. Then there exist a uniquely determined solution u_1 for the pathwise weak formulation (8.3)*, and a uniquely determined solution u_2 for the fully weak formulation* (8.4)*. Both solutions fulfill the stability estimate*

$$\|u_1(\omega, \cdot)\|_{H_0^1(\mathcal{D})} \leq \frac{\|f\|_{L^2(\mathcal{D})}}{a_{\min}}, \quad \omega \in \Omega, \quad \|u_2\|_{L^2(\Omega; L^2(\mathcal{D}))} \leq \frac{\|f\|_{L^2(\mathcal{D})}}{a_{\min}}.$$

Proof. To prove the existence, we want to apply Theorem 8.1.1. For the pathwise weak formulation, let $\omega \in \Omega$ be fixed. With the estimates assumed for the random field a, we obtain the boundedness of B_1 by

$$|B_1(u, v)| \leq \left| \int_{\mathcal{D}} a(\omega, x) \nabla u(x) \cdot \nabla v(x) \, dx \right| \leq a_{max} |\langle u, v \rangle_{H_0^1(\mathcal{D})}| \leq a_{max} \|u\|_{H_0^1(\mathcal{D})} \|v\|_{H_0^1(\mathcal{D})},$$

and the coercivity by

$$B_1(u, u) = \int_{\mathcal{D}} a(\omega, x) \nabla u(x) \cdot \nabla u(x) \, dx \geq a_{min} \int_{\mathcal{D}} \nabla u(x) \cdot \nabla u(x) \, dx = a_{min} \|u\|^2_{H_0^1(\mathcal{D})}.$$

For B_2, the boundedness

$$\begin{aligned} B_2(u, v) &= \int_\Omega \int_{\mathcal{D}} a(\omega, x) \nabla_x u(\omega, x) \cdot \nabla_x v(\omega, x) \, dx \, d\mathbb{P}(\omega) \\ &\leq a_{max} \langle u, v \rangle_{L^2(\Omega; H_0^1(\mathcal{D}))} \\ &\leq a_{max} \|u\|_{L^2(\Omega; H_0^1(\mathcal{D}))} \|v\|_{L^2(\Omega; H_0^1(\mathcal{D}))}, \end{aligned}$$

and coercivity

$$B_2(u, u) = \int_\Omega \int_{\mathcal{D}} a(\omega, x) \nabla_x u(\omega, x) \cdot \nabla_x u(\omega, x) \, dx \, d\mathbb{P}(\omega) \geq a_{min} \|u\|^2_{L^2(\Omega; H_0^1(\mathcal{D}))}$$

can be proven similarly. Hence, by Theorem 8.1.1, there exists, for every $\omega \in \Omega$, a unique solution $u_1(\omega, \cdot) \in H_0^1(\mathcal{D})$ of (8.3) and a unique solution $u_2 \in L^2(\Omega; H_0^1(\mathcal{D}))$ of (8.4). The stability estimates of the solutions follow directly by the bound given in Theorem 8.1.1. □

While this theorem proved in particular that, for fixed $\omega \in \Omega$, a uniquely determined solution $u_1(\omega, \cdot) \in H_0^1(\mathcal{D})$ of (8.3) exists, our next goal is to prove that the mapping $\Omega \ni \omega \mapsto u_1(\omega, \cdot)$ is an element of $L^2(\Omega; H_0^1(\mathcal{D}))$, in particular that it is measurable as an $H_0^1(\mathcal{D})$-valued random variable. This can then be used to prove that u_1 coincides with the solution of the fully weak formulation (8.4).
The following lemma turns out to be useful to prove measurability.

Lemma 8.1.4.
Consider the set

$$\mathcal{S} := \{g \in C_b(\mathcal{D}) \mid a_{\min} < \sup_{x \in \overline{\mathcal{D}}} g(x) < a_{\max}\} \subseteq C_b(\mathcal{D}),$$

and equip with the inherited norm and Borel σ-Algebra of $C_b(\mathcal{D})$. Let a right-hand side $f \in L^2(\mathcal{D})$ be fixed. Then, the solution mapping

$$\mathcal{A}\colon \mathcal{S} \to H_0^1(\mathcal{D}), \quad g \mapsto u_g,$$

where u_g is defined as the solution of

$$\int_{\mathcal{D}} g(x) \nabla_x u_g(x) \cdot \nabla v(x)\, dx = \int_{\mathcal{D}} f(x) v(x)\, dx, \quad v \in H_0^1(\mathcal{D}),$$

is well-defined and Lipschitz continuous with constant $\frac{\|f\|_{L^2(\mathcal{D})}}{a_{\min}^2}$.

Proof. First, we prove that $\mathcal{A}$ is well-defined. This, however, follows directly from Lax-Milgram Theorem 8.1.1, as the bilinear form is bounded and coercive for every coefficient $g \in \mathcal{S}$. To prove that this mapping is Lipschitz continuous, we use the ideas of [12, Lemma 4.1]. For this, let $g_1, g_2 \in \mathcal{S}$ be arbitrary, and denote by $u_j := \mathcal{A}(g_j)$ for $j = 1, 2$ the corresponding solutions. For $v \in H_0^1(\mathcal{D})$, we see that

$$\int_{\mathcal{D}} g_1(x) \nabla u_1(x) \cdot \nabla v(x)\, dx = \int_{\mathcal{D}} f(x) v(x)\, dx = \int_{\mathcal{D}} g_2(x) \nabla u_2(x) \nabla v(x)\, dx.$$

Thus, coercivity yields

$$\begin{aligned}
a_{\min} \|u_1 - u_2\|_{H_0^1(\mathcal{D})}^2 &\leq \int_{\mathcal{D}} g_1(x) \nabla (u_1 - u_2)(x) \cdot \nabla (u_1 - u_2)(x)\, dx \\
&= \int_{\mathcal{D}} g_1(x) \nabla u_1(x) \cdot \nabla (u_1 - u_2)(x)\, dx \\
&\quad - \int_{\mathcal{D}} g_1(x) \nabla u_2(x) \cdot \nabla (u_1 - u_2)(x)\, dx \\
&= \int_{\mathcal{D}} g_2(x) \nabla u_2(x) \cdot \nabla (u_1 - u_2)(x)\, dx \\
&\quad - \int_{\mathcal{D}} g_1(x) \nabla u_2(x) \cdot \nabla (u_1 - u_2)(x)\, dx \\
&= \int_{\mathcal{D}} (g_2 - g_1)(x) \nabla u_2(x) \cdot \nabla (u_1 - u_2)(x)\, dx.
\end{aligned}$$

This can be bounded using Hölder's inequality to obtain

$$\begin{aligned}&\int_{\mathcal{D}} (g_2 - g_1)(x)\nabla u_2(x)\nabla(u_1 - u_2)(x)\,dx\\ &\leq \|g_1 - g_2\|_{C(\mathcal{D})}\|u_2\|_{H_0^1(\mathcal{D})}\|u_1 - u_2\|_{H_0^1(\mathcal{D})}\\ &\leq \|g_1 - g_2\|_{C(\mathcal{D})}\frac{\|f\|_{L^2(\mathcal{D})}}{a_{\min}}\|u_1 - u_2\|_{H_0^1(\mathcal{D})}.\end{aligned} \tag{8.5}$$

If we now resubstitute this estimate, we can divide $\|u_1 - u_2\|_{H_0^1(\mathcal{D})}$ on both sides and thus obtain

$$\|u_1 - u_2\|_{H_0^1(\mathcal{D})} \leq \frac{\|f\|_{L^2(\mathcal{D})}}{a_{\min}^2}\|g_1 - g_2\|_{C_b(\mathcal{D})},$$

which is the stated result. □

In particular, the previous lemma proves that the solution mapping $\mathcal{A}\colon \mathcal{S} \to H_0^1(\mathcal{D})$ is measurable, as both spaces are equipped with the Borel σ-algebra. With this result, we can show that the solutions of (8.3) and (8.4) coincide:

Proposition 8.1.5.
Let Assumption 8.1.2 be valid, where additionally $a \in L^2(\Omega; C_b(\mathcal{D}))$. Then the solution u_1 of the pathwise problem (8.3)*, interpreted as function $u_1\colon \Omega \to H_0^1(\mathcal{D})$, is $\mathbb{P}$-measurable and additionally an element of $L^2(\Omega; H_0^1(\mathcal{D}))$. Furthermore, it solves the fully weak formulation* (8.4)*. Hence, u_1 is a representative of the solution $u_2 \in L^2(\Omega; H_0^1(\mathcal{D}))$ of* (8.4)*. In the following, we call the solution $u := u_1 = u_2 \in L^2(\Omega; H_0^1(\mathcal{D}))$ the* weak solution of the partial differential equation (8.1)*.*

Proof. As $a \in L^2(\Omega; C_b(\mathcal{D}))$ by assumption, we see that the mapping

$$\mathcal{J}\colon \Omega \to \mathcal{S}, \quad \omega \mapsto a(\omega, \cdot)$$

is measurable, as it is measurable on $C_b(\mathcal{D})$ but has values in $\mathcal{S}$ and thus remains measurable with respect to the restricted Borel σ-algebra. Moreover, Lemma 8.1.4 states that the mapping $\mathcal{A}\colon \mathcal{S} \to H_0^1(\mathcal{D})$ is well-defined and continuous. These statements imply that the solution of (8.3) is measurable as an $H_0^1(\mathcal{D})$-valued random variable, since it can be represented as

$$u_1(\omega, \cdot) = \mathcal{A}(a(\omega, \cdot)) = \mathcal{A} \circ \mathcal{J}(\omega), \quad \omega \in \Omega,$$

which is the composition of a measurable and a continuous mapping. With Proposition 2.3.4, we additionally obtain that u_1 is also $\mathbb{P}$-measurable with values in $H_0^1(\mathcal{D})$.
Next, we want to prove that $u_1 \in L^2(\Omega; H_0^1(\mathcal{D}))$. For this, let $\omega \in \Omega$ be fixed. The stability estimate of Lemma 8.1.3 implies that

$$\|u_1(\omega, \cdot)\|_{H_0^1(\mathcal{D})}^2 \leq \frac{\|f\|_{L^2(\mathcal{D})}^2}{a_{\min}^2},$$

where the upper bound is now independent of $\omega \in \Omega$. Hence, we obtain

$$\int_\Omega \|u_1(\omega, \cdot)\|_{H_0^1(\mathcal{D})}^2 \, d\mathbb{P}(\omega) \leq \frac{\|f\|_{L^2(\mathcal{D})}^2}{a_{\min}^2} < \infty$$

and u_1 is an element of $L^2(\Omega; H_0^1(\mathcal{D}))$ by Definition 2.3.8.
To show that u_1 coincides with the solution $u_2 \in L^2(\Omega; H_0^1(\mathcal{D}))$ of the fully weak formulation (8.4), let $v \in L^2(\Omega; H_0^1(\mathcal{D}))$ be arbitrary. This implies that $v(\omega, \cdot) \in H_0^1(\mathcal{D})$ almost surely. Thus, we see

$$\begin{aligned} B_2(u_1, v) &= \int_\Omega \int_\mathcal{D} a(\omega, x) \nabla_x u_1(\omega, x) \cdot \nabla_x v(\omega, x) \, dx \, d\mathbb{P}(\omega) \\ &= \int_\Omega B_1(u_1(\omega, \cdot), v(\omega, \cdot)) \, d\mathbb{P}(\omega) \\ &= \int_\Omega \int_\mathcal{D} f(x) v(\omega, x) \, dx \, d\mathbb{P}(\omega) \\ &= \langle f, v \rangle_{L^2(\Omega; L^2(\mathcal{D}))}. \end{aligned}$$

This proves that u_1 also solves the fully weak formulation. Since this solution is uniquely determined, both solutions coincide, hence u_1 is a representative of u_2. □

8.2 Exchanging the diffusion coefficient

While the previous section yields the *solvability* of the weak formulation of the elliptic partial differential equation (8.1), we next seek to compute the emerging solution. Here, it is often necessary to construct a surrogate to the diffusion coefficient a, as it may not be known on all of $\mathcal{D}$ but only on discrete sampling points, where the sampling itself may additionally be expensive or difficult. Therefore, we assume in the following that the diffusion coefficient a of (8.1) is only known at finitely many sampling points $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$.
When we now replace the random field a by a surrogate a^*, which, for example, stems from methods presented in Section 5.2 or Section 6.5, we have to ensure that the differential equation with the surrogate a^* as diffusion coefficient also possesses a solution u^* and, additionally, that this solution is a suitable approximation to u as well. If we first assume for the time being that both random fields a and a^* fulfill Assumption 8.1.2, we can prove, similarly to the Lipschitz continuity in Lemma 8.1.4, that the error of the solutions u and u^* are bounded as follows:

Lemma 8.2.1.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain and $a, a^* \in L^2(\Omega; C_b(\mathcal{D}))$ two pointwise defined second order random fields with constants $a_{min}, a_{max}, a^*_{min}, a^*_{max} > 0$ such that

$$\begin{aligned} 0 < a_{min} \leq & a(\omega, x) \leq a_{max}, \\ 0 < a^*_{min} \leq & a^*(\omega, x) \leq a^*_{max}, \end{aligned}$$

holds for all $\omega \in \Omega$ and $x \in \mathcal{D}$. Then, the weak solutions $u, u^* \in L^2(\Omega; H_0^1(\mathcal{D}))$ of (8.1) with diffusion coefficients a and a^*, respectively, and the same right-hand side $f \in L^2(\mathcal{D})$ exist, are unique, and their difference can be bounded by

$$\|u - u^*\|_{L^2(\Omega; H_0^1(\mathcal{D}))} \leq \frac{\|f\|_{L^2(\mathcal{D})}}{a_{min} a^*_{min}} \|a - a^*\|_{L^2(\Omega; C_b(\mathcal{D}))}.$$

Proof. The existence of both solutions with diffusion coefficient a and a^* follows directly from Lemma 8.1.3 and Proposition 8.1.5. The error bound uses the same techniques as Lemma 8.1.4. In particular, let $\omega \in \Omega$ be fixed and set $g_1 := a(\omega, \cdot)$ and $g_2 := a^*(\omega, \cdot)$ as in Lemma 8.1.4, and denote the corresponding solutions with $u_1 = u$ and $u_2 = u^*$, respectively. We then can proceed as described in Lemma 8.1.4 until (8.5), where we have to consider that, in our case, the stability estimate of Lemma 8.1.3 yields

$$\begin{aligned}
&a_{min}\|u(\omega,\cdot) - u^*(\omega,\cdot)\|^2_{H_0^1(\mathcal{D})} \\
&\leq \|a^*(\omega,\cdot) - a(\omega,\cdot)\|_{C(\mathcal{D})}\|u^*(\omega,\cdot)\|_{H_0^1(\mathcal{D})}\|u(\omega,\cdot) - u^*(\omega,\cdot)\|_{H_0^1(\mathcal{D})} \\
&\leq \|a^*(\omega,\cdot) - a(\omega,\cdot)\|_{C(\mathcal{D})}\frac{\|f\|_{L^2(\mathcal{D})}}{a^*_{\min}}\|u(\omega,\cdot) - u^*(\omega,\cdot)\|_{H_0^1(\mathcal{D})}.
\end{aligned}$$

Dividing by a_{min} and $\|u(\omega,\cdot) - u^*(\omega,\cdot)\|_{H_0^1(\mathcal{D})}$ on both sides, we obtain

$$\|u(\omega,\cdot) - u^*(\omega,\cdot)\|_{H_0^1(\mathcal{D})} \leq \|a^*(\omega,\cdot) - a(\omega,\cdot)\|_{C(\mathcal{D})}\frac{\|f\|_{L^2(\mathcal{D})}}{a_{min}a^*_{min}}$$

for all $\omega \in \Omega$. With this bound, we further estimate

$$\begin{aligned}
\|u - u^*\|_{L^2(\Omega;H_0^1(\mathcal{D}))} &= \left(\int_\Omega \|u(\omega,\cdot) - u^*(\omega,\cdot)\|^2_{H_0^1(\mathcal{D})}\, d\mathbb{P}(\omega)\right)^{\frac{1}{2}} \\
&\leq \frac{\|f\|_{L^2(\mathcal{D})}}{a_{min}a^*_{min}}\left(\int_\Omega \|a(\omega,\cdot) - a^*(\omega,\cdot)\|^2_{C(\mathcal{D})}\, d\mathbb{P}(\omega)\right)^{\frac{1}{2}} \\
&= \frac{\|f\|_{L^2(\mathcal{D})}}{a_{min}a^*_{min}}\|a - a^*\|_{L^2(\Omega;C_b(\mathcal{D}))},
\end{aligned}$$

which proves the claim. □

In order to prove that a surrogate a^* possesses a weak solution of the differential equation (8.1), we have to ensure that it also fulfills Assumption 8.1.2. This, however, can be established by considering pointwise error bounds: To be precise, we have $\mathbb{P}$-almost surely the estimate

$$\|a(\omega,\cdot) - a^*(\omega,\cdot)\|_{C(\mathcal{D})} \leq Ch_{X,\mathcal{D}}^{\tau-\frac{d}{2}}\rho_{X,\mathcal{D}}^{\sigma-\tau}\|a(\omega,\cdot)\|_{H^\tau(\mathcal{D})} \leq Ch_{X,\mathcal{D}}^{\tau-\frac{d}{2}}\rho_{X,\mathcal{D}}^{\sigma-\tau}a^{(\sigma)}_{max},$$

provided $a^{(\sigma)}_{max} := \sup_{\omega\in\Omega}\|a(\omega,\cdot)\|_{C(\mathcal{D})} < \infty$. If the right hand side is less than $\frac{a_{min}}{2}$, we see that

$$\begin{aligned}
a^*(\omega,x) &= a(\omega,x) + a^*(\omega,x) - a(\omega,x) \\
&\geq |a(\omega,x)| - |a(\omega,x) - a^*(\omega,x)| \\
&\geq a_{min} - \frac{a_{min}}{2} = \frac{a_{min}}{2} =: a^*_{min} > 0,
\end{aligned} \tag{8.6}$$

which implies that $a^*(\omega,x) \geq a^*_{min} > 0$ $\mathbb{P}$-almost surely. The bound from above follows similarly but easier. Fortunately, the error estimates proven in Section 5.2 and Corollary 6.5.1 guarantee these assumptions.

Corollary 8.2.2.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain with Lipschitz boundary and $f \in L^2(\mathcal{D})$ Furthermore, assume that the diffusion coefficient $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$, where $\sigma > \frac{d}{2}$, fulfills Assumption 8.1.2. Denote the weak solution of (8.1) with diffusion coefficient a and right-hand side f by $u \in L^2(\Omega; H_0^1(\mathcal{D}))$.

1. *Let $X \subseteq \mathcal{D}$ be a finite point set and $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ a positive definite function with $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$.*

   If $h_{X,\mathcal{D}} < h_0$ and $h_{X,\mathcal{D}}$ is additionally small enough such that the error estimate fulfills $\sup_{\omega\in\Omega} \|a(\omega,\cdot) - \mathcal{I}_{X,\Phi}(a)\|_{C(\mathcal{D})} < \frac{a_{min}}{2}$, then the weak solution u^* of (8.1) with diffusion coefficient $\mathcal{I}_{X,\Phi}(a)$ is well-defined and satisfies the bound

$$\|u - u^*\|_{L^2(\Omega;H_0^1(\mathcal{D}))} \leq C \frac{\|f\|_{L^2(\mathcal{D})}}{a_{min}^2} h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}} \|a\|_{L^2(\Omega;H^\sigma(\mathcal{D}))}.$$

2. *Let $(X_j)_{j\in\mathbb{N}} \subseteq \mathcal{D}$ be a sequence of finite point sets as in Corollary 6.5.1. If the multilevel approximation $A_L(a)$ fulfills $\sup_{\omega\in\Omega} \|a - A_L(a)\|_{C(\mathcal{D})} < \frac{a_{min}}{2}$, the weak solution $u_L \in L^2(\Omega; H_0^1(\mathcal{D}))$ of (8.1) with diffusion coefficient $A_L(a)$ is well-defined and fulfills*

$$\|u_L - u\|_{L^2(\Omega;H_0^1(\mathcal{D}))} \leq C \frac{\|f\|_{L^2(\mathcal{D})}}{a_{min}^2} \alpha^L \|a\|_{L^2(\Omega;H^\sigma(\mathcal{D}))}.$$

Proof. Under these assumptions, we can use (8.6) and Lemma 8.1.3 to prove that the weak solutions $u^*, u_L \in L^2(\Omega; H_0^1(\mathcal{D}))$ exist. For the error estimates of the solutions, we use Lemma 8.2.1 and apply the bounds given in Corollary 5.2.3 for the error of kernel-based interpolation, and Corollary 6.5.1 for the multilevel error. This yields the claim. □

Note that the assumption in the first case corresponds to the first case in Section 5.2.1, i.e., that the regularity of the random field $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ is given. An analogous result can additionally be proven for the second case by replacing Corollary 5.2.3 with the appropriate error estimate Theorem 5.2.5 or Theorem 5.2.6, respectively. For the sake of brevity, we omit these cases.

8.2.1 Parameterization of the differential equation

The canonical way to tackle the uncertainty is the parameterization of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a (usually high dimensional) *parameter space* $\Gamma \subseteq \mathbb{R}^n$. We explain this procedure in the following and also see that the methods proposed earlier naturally fit into this framework. Nonetheless, for the pathwise methods presented in Section 5.2 or Section 6.5, we also can omit this procedure and directly proceed with the approximation of the solution presented in Section 8.3 below. However, the parameterization allows us to apply, for example, the multilevel method using tensorized kernels to approximate the (parameterized) diffusion coefficient in both the spatial domain $\mathcal{D}$ and the parameter space Γ simultaneously.
In order to be able to parameterize the partial differential equation (8.1), we have to consider specific diffusion coefficients, which are also called *finite noise* (see [35, Definition 9.38]).

Definition 8.2.3 (Finite noise)**.**
A second order random field $a \in L^2(\Omega; L^2(\mathcal{D}))$ is called *(n-)finite noise* if there are $n \in \mathbb{N}$ real-valued random variables $\xi_k \colon \Omega \to \Gamma_k$, with $\Gamma_k \subseteq \mathbb{R}$ for all $1 \leq k \leq n$, such that a can be written as

$$a(\omega, x) = \widetilde{a}((\xi_1(\omega), \dots, \xi_n(\omega)), x), \quad \omega \in \Omega, \quad x \in \mathcal{D}.$$

If we set $\Gamma := \Gamma_1 \times \cdots \times \Gamma_n$ and $\xi := (\xi_1, \dots, \xi_n)$, we identify a with $\widetilde{a}$ and again write $a(\omega, x) = a(\xi(\omega), x)$ for the emerging random field.

Remark 8.2.4.
If we consider, for example, the kernel-based interpolation of a random field, we can use the Lagrange representation to write

$$\mathcal{I}_{X,\Phi}(a)(\omega, x) = \sum_{j=1}^{N} a(\omega, x_j)\chi_j(x),$$

with $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ and a suitable positive definite function Φ. Using this representation, we can see that a is measurable with respect to the σ-algebra created by $a(\cdot, x_1), \dots, a(\cdot, x_N)$. Hence, this random field is N-finite noise with random variables $\xi_j = a(\cdot, x_j)$, for $1 \leq j \leq N$. Obviously, the argument can be generalized for the penalized least squares approximant. For the pathwise multilevel approximant of the random field a, we can use the explicit basis representation given in (6.9), i.e.,

$$A_L(a)(\omega, x) = \sum_{j=1}^{N} a(\omega, x_j)b_j(x),$$

where $N = |X|$, with $X = \bigcup_{j=1}^{L} X_j = \{x_1, \dots, x_N\}$, endowed with a suitable ordering. Thus, the multilevel approximation is also N-finite noise, with $\xi_j = a(\cdot, x_j)$ for $1 \leq j \leq N$. Another way to end up with finite noise random fields is by using the Karhunen-Loève expansion of a random field and truncating the series representation as shown in Corollary 4.2.7. To be precise, under the assumptions of Corollary 4.2.5, we can expand the random field $a \in L^2(\Omega; L^2(\mathcal{D}))$ as

$$a(\omega, x) = m(x) + \sum_{j=1}^{\infty} \sqrt{\lambda_j}\xi_j(\omega)e_j(x), \quad \omega \in \Omega, \quad x \in \mathcal{D}$$

and then define $a_n(\omega, x) := m(x) + \sum_{j=1}^{n} \sqrt{\lambda_j}\xi_j(\omega)e_j(x)$, for a fixed natural number $n \in \mathbb{N}$. With this setup, it is clear that a_n is n-finite noise.
Of course, we could also combine both methods, as by Theorem 5.3.3 the Karhunen-Loève expansion of the interpolated random field $\mathcal{I}_{X,\Phi}(a)$ can be calculated explicitly and truncated afterwards.

If the diffusion coefficient a is n-finite noise, the weak solution $u \in L^2(\Omega; H_0^1(\mathcal{D}))$ is finite noise as well. The main tool to prove this statement is the so-called *Doob-Dynkin Lemma* (see [27, Lemma 1.14]):

Theorem 8.2.5 (Doob-Dynkin Lemma)**.**
Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $(S, \mathcal{S})$ and $(T, \mathcal{T})$ two measurable spaces, where $\mathcal{S}$ is the Borel $\sigma-$algebra on S. Then, for two measurable functions $f\colon \Omega \to S$ and $g\colon \Omega \to T$, the following statements are equivalent:

1. *f is g-measurable, i.e., $f^{-1}(A) \in \sigma(g)$ for all $A \in \mathcal{S}$,*
2. *There exists a measurable function $h\colon T \to S$ such that $f = h \circ g$.*

In order to apply this theorem, we have to show that if the diffusion coefficient $a\colon \Omega \to V$ is measurable with respect to a sub-σ-algebra $\mathcal{A} \subseteq \mathcal{F}$, then the weak solution u of (8.1) is measurable with respect to $\mathcal{A}$ as well ([35, Theorem 9.30]).

Lemma 8.2.6.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a bounded domain, $V \subseteq L^2(\mathcal{D})$ a separable Banach space endowed with the Borel σ-algebra $\mathcal{V} = \mathcal{B}(V)$, and $a \in L^2(\Omega; V)$ a second order random field which fulfills Assumption 8.1.2. Assume there exists a σ-algebra $\mathcal{A} \subseteq \mathcal{F}$ such that a is $(\mathcal{A}, \mathcal{V})-$measurable. Then the weak solution $u \in L^2(\Omega; H_0^1(\mathcal{D}))$ of (8.1) is $(\mathcal{A}, \mathcal{B}(H_0^1(\mathcal{D})))$-measurable as well.

Both statements together imply that when the diffusion coefficient $a \in L^2(\Omega; L^2(\mathcal{D}))$ is n-finite noise, the weak solution of (8.1) is also measurable with respect to these random variables.

Theorem 8.2.7.
Let $a \in L^2(\Omega; L^2(\mathcal{D}))$ be a second order random field that satisfies Assumption 8.1.2. If $a(\omega, x) = a(\xi_1(\omega), \dots, \xi_n(\omega), x)$ is n-finite noise, with real-valued random variables ξ_j for $1 \leq j \leq n$, and is measurable with respect to $\sigma((\xi_1, \dots, \xi_n))$, the weak solution $u \in L^2(\Omega; H_0^1(\mathcal{D}))$ of (8.1) is measurable with respect to $\sigma((\xi_1, \dots, \xi_n))$ and allows the representation

$$u(\omega, x) = u(\xi_1(\omega), \dots, \xi_n(\omega), x), \quad \omega \in \Omega, \quad x \in \mathcal{D}.$$

Proof. This statement is proven in [35, Lemma 9.40]. We know by assumption that a is measurable with respect to $\sigma((\xi_1, \dots, \xi_N))$. Hence, we can apply Lemma 8.2.6 to infer that the solution u is also measurable with respect to this σ-algebra. The representation then follows by the Doob-Dynkin Theorem 8.2.5. □

This theoretical groundwork allows us to parameterize the solution:
Assume that the diffusion coefficient $a(\omega, x) \in L^2(\Omega; L^2(\mathcal{D}))$ is n-finite noise for a fixed natural number $n \in \mathbb{N}$, i.e., that a can be written as

$$a(\omega, x) = a(\xi_1(\omega), \dots, \xi_n(\omega), x), \quad \omega \in \Omega, \quad x \in \mathcal{D}.$$

We now consider the $\mathbb{R}^n$-valued random variable $\xi\colon \Omega \to \mathbb{R}^n$ and assume that the Lebesgue density $\rho\colon \mathbb{R}^n \to \mathbb{R}$ of the pushforward measure $\mathbb{P}^\xi$ of ξ is known. If we denote the support of ρ by

$$\Gamma := \overline{\{x \in \mathbb{R}^n \mid \rho(x) > 0\}} \subseteq \mathbb{R}^n,$$

$\left(\Gamma, \mathcal{B}(\Gamma), \mathbb{P}^\xi\right)$ defines a probability space on Γ. The $L^2(\Omega; L^2(\mathcal{D}))$ norm of the random field a can be written as a weighted norm on Γ:

$$\|a\|^2_{L^2(\Omega; L^2(\mathcal{D}))} = \int_\Omega \|a(\omega, \cdot)\|^2_{L^2(\mathcal{D})}\, d\mathbb{P}(\omega) = \int_\Gamma \|a(y, \cdot)\|^2_{L^2(\mathcal{D})} \rho(y)\, dy =: \|a\|_{L^2_\rho(\Gamma; L^2(\mathcal{D}))}.$$

Similar statements also hold for $L^2_\rho(\Omega; H^1_0(\mathcal{D}))$. This allows us to replace the unknown probability space $(\Omega, \mathcal{F}, \mathbb{P})$ in the weak formulation (8.4) by the known, but probably high dimensional *parameter space* $\left(\Gamma, \mathcal{B}(\Gamma), \mathbb{P}^\xi\right)$. Therefore, we seek to solve the following problem:
Find a solution $u \in L^2_\rho(\Gamma; H^1_0(\mathcal{D}))$ that satisfies, for all $v \in L^2_\rho(\Gamma; H^1_0(\mathcal{D}))$, the equation

$$\int_\Gamma \int_\mathcal{D} a(y,x) \nabla_x u(y,x) \nabla_x v(y,x)\, dx\, \rho(y)\, dy = \int_\Gamma \int_\mathcal{D} f(x) v(y,x)\, dx\, \rho(y)\, dy. \tag{8.7}$$

We call u the *parameterized weak solution* of (8.1).
Last, we transform Assumption 8.1.2:

Assumption 8.2.8.
Let $\rho \in L^1(\Gamma)$ with $\rho \geq 0$ on Γ. Assume $a \in L^2_\rho(\Gamma; L^2(\mathcal{D}))$ is a second order random field. Additionally, let there be constants $0 < a_{\min} \leq a_{\max} < \infty$ such that, for all $y \in \Gamma$ and $x \in \mathcal{D}$, the bound

$$0 < a_{\min} \leq a(y,x) \leq a_{\max} < \infty$$

holds.

With this assumption, we can prove, analogously to Lemma 8.1.3, the unique solvability of (8.7).

Surrogate for the parameterized diffusion coefficient

Although we parameterized the diffusion coefficient $a \in L^2(\Gamma; L^2(\mathcal{D}))$ in the last section, there is still the possibility that the evaluation of the diffusion coefficient is expensive from a numerical point of view, which makes solving the differential equation unfeasible. Thus, even after parameterizing, it can be advisable to choose a surrogate to the random field.
In contrast to Ω, the parameter space $\Gamma \subseteq \mathbb{R}^n$ is explicitly known. This allows us to apply approximation methods using tensorized kernels, provided the diffusion coefficient permits sufficient regularity in both arguments. If the parameterization, for example, stems from a truncated Karhunen-Loève expansion, as proposed in Remark 8.2.4, the regularity assumption on Γ is obvious. Thus, we assume in the following that

$$a \in L^2_\rho(\Gamma; H^\sigma(\mathcal{D})) \cap H^{\sigma,\tau}_{\text{mix}}(\mathcal{D} \times \Gamma),$$

where we note that, for the sake of convenience, we swap the arguments of the spatial domain $x \in \mathcal{D}$ and the parameter space Γ in a, i.e., we write $a = a(x,y)$, for $x \in \mathcal{D}$ and $y \in \Gamma$.
In order to choose a surrogate to this parameterized random field, we focus on one generalization of the multilevel method for mixed regularity spaces as presented in Chapter 7. As already pointed out in the previous section, the parameter space $\Gamma \subseteq \mathbb{R}^n$ is usually (a subset of) a high dimensional space, while the spatial domain $\mathcal{D} \subseteq \mathbb{R}^d$ only is of moderate size with $d = 2$ or $d = 3$. Paired with our numerical considerations in Section 7.3.1, the natural choice, in this case, is the multilevel method using tensorized kernels as presented in Section 7.1. In order to apply Lemma 8.2.1, we have to consider the pointwise error estimate given in Corollary 7.1.7, which leads to the following theorem:

Theorem 8.2.9.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ and $\Gamma \subseteq \mathbb{R}^n$ be two bounded domains with Lipschitz boundary, where the measurable space $(\Gamma, \mathcal{B}(\Gamma))$ is equipped with a probability measure that possesses a Lebesgue density $\rho \in L^1(\Gamma)$. Let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ and $\Psi \in L^1(\mathbb{R}^n) \cap C(\mathbb{R}^n)$ be two positive definite functions with $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$, and $H_\Psi(\mathbb{R}^n) \cong H^\tau(\mathbb{R}^n)$, where $\tau > \frac{n}{2}$. Furthermore, let two sequences of point sets $(X_j)_{j\in\mathbb{N}} \subseteq \mathcal{D}$ and $(Y_j)_{j\in\mathbb{N}} \subseteq \Gamma$, which fulfill the assumptions of Corollary 7.1.7, be given. Then, for every random field $a \in H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\Gamma)$ and $L \in \mathbb{N}$, the error estimate

$$\begin{aligned}&\|a - \boldsymbol{A}_L(a)\|_{L^2_\rho(\Gamma;C(\mathcal{D}))}\\ &\leq C\left(\widetilde{C_1}\max\left\{\frac{\mu_{\mathcal{D}}^d}{\mu_\Gamma^n}, \frac{\mu_\Gamma^n}{\mu_{\mathcal{D}}^d}\right\}^{\frac{1}{2}}\left(\mu_{\mathcal{D}}^{\sigma-\frac{d}{2}}\mu_\Gamma^{\tau-\frac{n}{2}} + \mu_{\mathcal{D}}^{\sigma-d} + \mu_\Gamma^{\tau-n}\right)\right)^L \|a\|_{H^{\sigma,\tau}_{\mathrm{mix}}(\mathcal{D}\times\Gamma)}\end{aligned}$$

holds, where $\boldsymbol{A}_L$ denotes the multilevel approximation using tensorized kernels, constructed as in Corollary 7.1.7, with L levels.

Proof. Under the given assumptions, we obtain the estimate

$$\begin{aligned}\|a - \boldsymbol{A}_L(a)\|^2_{L^2_\rho(\Gamma;C(\mathcal{D}))} &= \int_\Gamma \|a(\cdot,y) - \boldsymbol{A}_L(a)(\cdot,y)\|^2_{C(\mathcal{D})}\rho(y)\,dy\\ &= \int_\Gamma \left(\sup_{x\in\mathcal{D}}|a(x,y) - \boldsymbol{A}_L(a)(x,y)|\right)^2 \rho(y)\,dy\\ &\leq \int_\Gamma \rho(y)\,dy\left(\sup_{x\in\mathcal{D}}\sup_{z\in\Gamma}|a(x,z) - \boldsymbol{A}_L(a)(x,z)|\right)^2\\ &= \|a - \boldsymbol{A}_L(a)\|^2_{C(\mathcal{D}\times\Gamma)},\end{aligned}$$

since ρ is a Lebesgue density and therefore its integral equals one. Hence, we can apply the error estimate given in Corollary 7.1.7, which proves the claim. □

With this, we can now proceed as in Section 8.2 and use this bound to exchange the diffusion coefficient in (8.7) by the given approximation.

8.3 Approximating the solution

While we discussed how to parameterize the diffusion coefficient and hence the weak solution to our elliptic PDE in the last section, we now focus on computing the solutions approximately. Here, we consider *Galerkin type* methods, which are mainly divided into two categories: The *pathwise* Galerkin and the *stochastic* Galerkin methods. The main difference between these two methods lies in the choice of the weak formulation to be discretized that in turn influences the choice of the underlying approximation space. While the pathwise Galerkin method uses the pathwise weak formulation (8.3) and discretizes only in the spatial domain $\mathcal{D} \subseteq \mathbb{R}^d$, the stochastic Galerkin method demands a parameterized problem as given in (8.7), but then discretizes in both the spatial domain $\mathcal{D}$ and in the parameter space $\Gamma \subseteq \mathbb{R}^N$ simultaneously. Here, for simplicity, we focus on a pathwise method, but refer to [35, Chapter 9.5] for an overview of stochastic Galerkin methods, as these are also applicable in our setting.

Pathwise Finite Element methods

In the following, we mainly use the results given in [9], to obtain statements similar to [32, Chapter 7] for random fields.
We use the setting given in [41, Chapter 2] as an assumption for the finite element space.

Setting 8.3.1.
Let $W_\hbar \subseteq H_0^1(\mathcal{D})$ be a standard finite element space with dimension $N_\hbar$ which contains continuous piecewise polynomials defined on a regular triangulation $\mathcal{T}_\hbar$ with $\hbar > 0$. Furthermore, we assume $W_\hbar(\mathcal{D})$ fulfills the following approximation property:

$$\inf_{v \in W_\hbar(\mathcal{D})} \|u - v\| \leq C\hbar^s,$$

where $s > 0$ is determined by the smoothness of $u \in H_0^1(\mathcal{D})$ and the degree of the approximating finite element subspace. $C = C(s, u)$ denotes a constant independent of $\hbar$, and $\|\cdot\|$ denotes a suitable norm on $H_0^1(\mathcal{D})$, usually $\|\cdot\|_{H^1(\mathcal{D})}$ or $\|\cdot\|_{L^2(\mathcal{D})}$.

Obviously, these are rather abstract assumptions, as for finite element methods one usually obtains estimates as given in [9, Theorem 4.4.20, Remark 4.4.27].

Corollary 8.3.2.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a convex polyhedron, $W_\hbar$ a finite element space consisting of piecewise linear polynomials on a regular triangulation $\mathcal{T}_\hbar$ with dimension $N_\hbar$.
Then, for all $u \in H_0^1(\mathcal{D})$, the estimate

$$\inf_{v \in W_\hbar} \|u - v\|_{L^2(\mathcal{D})} \leq C\hbar \|u\|_{H_0^1(\mathcal{D})},$$

and, for all $u \in H^2(\mathcal{D})$, the estimate

$$\inf_{v \in W_\hbar} \|u - v\|_{H_0^1(\mathcal{D})} \leq C\hbar \|u\|_{H^2(\mathcal{D})}$$

holds.

This result leads to the following approximation result for finite elements (cf. [9, Theorem 5.7.6]), which we apply to the paths of random fields.

Theorem 8.3.3.
Let the assumptions of Corollary 8.3.2 be valid. Assume that $a \in L^2(\Omega; L^2(\mathcal{D}))$ is a second order random field that fulfills Assumption 8.1.2, and denote the weak solution of (8.1) by $u \in L^2(\Omega; H_0^1(\mathcal{D}))$. Then the paths of u are $\mathbb{P}$-almost surely in $H_0^1(\mathcal{D})$. Moreover, there exists a constant $C > 0$ such that for almost all $\omega \in \Omega$ and $0 < \hbar \leq 1$ the finite element approximation $u_\hbar(\omega, \cdot) \in H_0^1(\mathcal{D})$ to $u(\omega, \cdot)$ satisfies the bound

$$\|u(\omega, \cdot) - u_\hbar(\omega, \cdot)\|_{L^2(\mathcal{D})} \leq C\hbar \|u(\omega, \cdot)\|_{H_0^1(\mathcal{D})}.$$

If, additionally, $u \in L^2(\Omega; H^2(\mathcal{D}))$, the estimate

$$\|u(\omega, \cdot) - u_\hbar(\omega, \cdot)\|_{H_0^1(\mathcal{D})} \leq C \inf_{v \in W_\hbar} \|u(\omega, \cdot) - v\|_{H^1(\mathcal{D})} \leq C\hbar \|u(\omega, \cdot)\|_{H^2(\mathcal{D})}$$

holds.

Proof. Under our assumptions, we know that $u(\omega,\cdot) \in H_0^1(\mathcal{D})$ $\mathbb{P}$-almost surely and that for every $0 < \hbar \leq 1$ the solution $u_\hbar(\omega,\cdot) \in H_0^1(\mathcal{D})$ exists. Hence, we can apply the finite element result given in [9, Theorem 5.7.6]. Note that all the assumptions made are fulfilled by our choice of finite elements and the error estimate given in Corollary 8.3.2. □

Finally, we are interested in some *quantity of interest* of the solution, which will be denoted by

$$Q(\omega) := \mathcal{Q}(u(\omega,\cdot)),$$

for a linear and bounded operator $\mathcal{Q}\colon L^2(\mathcal{D}) \to V$, where $(V, \|\cdot\|_V)$ is a separable Banach space. A typical choice can, for example, be the evaluation of the solution in one point (i.e., $V = \mathbb{R}$), or simply the identity operator ($V = L^2(\mathcal{D})$).
We now give full error estimates for all different kernel choices considered in Section 5.2 and, additionally, for the multilevel approximation considered in Section 6.5. Similarly as in Corollary 8.2.2, we only prove the results for the first case, as, for the second case, similar error estimates as in Corollary 8.2.2 can be derived, which lead to analogous results for the diffusion coefficient and the constructed surrogate. To further shorten this, we state the general assumptions separately.

Assumption 8.3.4.
Let $\mathcal{D} \subseteq \mathbb{R}^d$ be a convex polyhedron. For $0 < \hbar < 1$, let $\mathcal{T}_\hbar$ be a sequence of uniform triangulations on $\mathcal{D}$. Denote by $W_\hbar$ the corresponding finite element space using piecewise linear and continuous elements. Additionally, let $(V, \|\cdot\|_V)$ be a separable Hilbert space and $\mathcal{Q}\colon L^2(\mathcal{D}) \to V$ a linear and bounded operator.

We begin with the case that the kernel is chosen to match the smoothness of the random field.

Theorem 8.3.5.
Suppose Assumption 8.3.4 holds. Let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ be a positive definite function with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$, and $X = \{x_1, \dots, x_N\} \subseteq \mathcal{D}$ a finite point set with mesh norm $h_{X,\mathcal{D}} \leq h_0$. Let $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ be a second order random field which fulfills Assumption 8.1.2. Assume further that $h_{X,\mathcal{D}}$ is small enough such that

$$\sup_{\omega\in\Omega} \|a - \mathcal{I}_{X,\Phi}(a)\|_{C_b(\mathcal{D})} \leq \frac{a_{\min}}{2}$$

holds. Now, denote by $u \in L^2(\Omega; H_0^1(\mathcal{D}))$ the weak solution of (8.1) with diffusion coefficient a and, for $0 < \hbar < 1$, by $u_\hbar^* \in L^2(\Omega; H_0^1(\mathcal{D}))$ the finite element approximation of the weak solution of (8.1) with diffusion coefficient $\mathcal{I}_{X,\Phi}(a)$. Then the error estimate

$$\|\mathcal{Q}(u) - \mathcal{Q}(u_\hbar^*)\|_{L^2(\Omega;V)} \leq C\|\mathcal{Q}\|_{L^2(\mathcal{D})\to V}\left(h_{X,\mathcal{D}}^{\sigma-\frac{d}{2}}\|a\|_{L^2(\Omega;H^\sigma(\mathcal{D}))} + \hbar\frac{\|f\|_{L^2(\mathcal{D})}}{a_{min}}\right)$$

holds.

Proof. As $\mathcal{Q}$ is bounded, we obtain the estimate

$$\begin{aligned}
\|\mathcal{Q}(u) - \mathcal{Q}(u_\hbar^*)\|_{L^2(\Omega;V)}^2 &= \int_\Omega \|\mathcal{Q}(u(\omega,\cdot) - u_\hbar^*(\omega,\cdot))\|_V^2 \, d\mathbb{P}(\omega) \\
&\leq \|\mathcal{Q}\|_{L^2(\mathcal{D})\to V} \int_\Omega \|u(\omega,\cdot) - u_\hbar^*(\omega,\cdot)\|_{L^2(\mathcal{D})}^2 \, d\mathbb{P}(\omega) \\
&\leq \|\mathcal{Q}\|_{L^2(\mathcal{D})\to V}\|u - u_\hbar^*\|_{L^2(\Omega;L^2(\mathcal{D}))}.
\end{aligned}$$

If we denote by $u^*(\omega,\cdot) \in H_0^1(\mathcal{D})$ the true weak solution of (8.1) with diffusion coefficient $\mathcal{I}_{X,\Phi}(a)$, we can bound the latter norm by

$$\|u - u_\hbar^*\|_{L^2(\Omega;L^2(\mathcal{D}))} \leq C\|u - u^*\|_{L^2(\Omega;H_0^1(\mathcal{D}))} + \|u^* - u_\hbar^*\|_{L^2(\Omega;L^2(\mathcal{D}))}.$$

Here, we can use Corollary 8.2.2 for the first term, while the latter can be bounded with finite element error presented in Theorem 8.3.3. Both estimates together yield the claim. □

Next, we assume to know the regularity of the covariance function $\mathrm{Cov} \in H_{\mathrm{mix}}^{\sigma,\sigma}(\mathcal{D}\times\mathcal{D})$ and choose the kernel to match this smoothness.

Theorem 8.3.6.
Suppose Assumption 8.3.4 holds. Let $a \in L^2(\Omega;L^2(\mathcal{D}))$ be a second order random field with mean $m \in H^\beta(\mathcal{D})$, where $\beta > \frac{d}{2}$, and covariance function $\mathrm{Cov} \in H_{\mathrm{mix}}^{\sigma,\sigma}(\mathcal{D}\times\mathcal{D})$, where $\sigma > d$. Choose a positive definite function $\Phi \in L^1(\mathbb{R}^d)\cap C(\mathbb{R}^d)$ with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, and let $X = \{x_1,\dots,x_N\} \subseteq \mathcal{D}$ be a finite point set such that its mesh norm fulfills $h_{X,\mathcal{D}} \leq h_0$, and, additionally,

$$\sup_{\omega\in\Omega} \|a - \mathcal{I}_{X,\Phi}(a)\|_{C(\mathcal{D})} \leq \frac{a_{\min}}{2}$$

holds. Now denote by $u \in L^2(\Omega;H_0^1(\mathcal{D}))$ the weak solution of (8.1) with diffusion coefficient a and, for $0 < \hbar < 1$, by $u_\hbar^* \in L^2(\Omega;H_0^1(\mathcal{D}))$ the finite element approximation of the weak solution of (8.1) with diffusion coefficient $\mathcal{I}_{X,\Phi}(a)$. Then, for all $\frac{d}{2} < \tau < \min\left\{\beta, \sigma - \frac{d}{2}\right\}$, the error estimate

$$\|\mathcal{Q}(u) - \mathcal{Q}(u_\hbar^*)\|_{L^2(\Omega;V)} \leq C\|\mathcal{Q}\|_{L^2(\mathcal{D})\to V}\left(h_{X,\mathcal{D}}^{\tau-\frac{d}{2}}\rho_{X,\mathcal{D}}^{\sigma-\tau}\|a\|_{L^2(\Omega;H^\tau(\mathcal{D}))} + \hbar\frac{\|f\|_{L^2(\mathcal{D})}}{a_{min}}\right)$$

holds.

In the next theorem, we again assume the regularity of $\mathrm{Cov} \in H_{\mathrm{mix}}^{\sigma,\sigma}(\mathcal{D}\times\mathcal{D})$ is known, but now use the theory for sample path regularity developed in Theorem 4.3.5 to choose the kernel Φ to match regularity of the paths of the diffusion coefficient.

Theorem 8.3.7.
Suppose Assumption 8.3.4 holds. Let $a \in L^2(\Omega;L^2(\mathcal{D}))$ be a second order random field with mean $m \in H^\beta(\mathcal{D})$, where $\beta > \frac{d}{2}$, and covariance function $\mathrm{Cov} \in H_{\mathrm{mix}}^{\sigma,\sigma}(\mathcal{D}\times\mathcal{D})$, where $\sigma > d$. Choose $\frac{d}{2} < \tau < \min\left\{\beta, \sigma - \frac{d}{2}\right\}$ fixed and a positive definite function $\Phi \in L^1(\mathbb{R}^d)\cap C(\mathbb{R}^d)$ with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\tau(\mathbb{R}^d)$, and let $X = \{x_1,\dots,x_N\} \subseteq \mathcal{D}$ be a finite point set such that its mesh norm fulfills $h_{X,\mathcal{D}} \leq h_0$ and, additionally,

$$\sup_{\omega\in\Omega} \|a - \mathcal{I}_{X,\Phi}(a)\|_{C(\mathcal{D})} \leq \frac{a_{\min}}{2}$$

holds. Now, denote by $u \in L^2(\Omega;H_0^1(\mathcal{D}))$ the weak solution of (8.1) with diffusion coefficient a and, for $0 < \hbar < 1$, by $u_\hbar^* \in L^2(\Omega;H_0^1(\mathcal{D}))$ the finite element approximation of the weak solution of (8.1) with diffusion coefficient $\mathcal{I}_{X,\Phi}(a)$. Then the error estimate

$$\|\mathcal{Q}(u) - \mathcal{Q}(u_\hbar^*)\|_{L^2(\Omega;V)} \leq C\|\mathcal{Q}\|_{L^2(\mathcal{D})\to V}\left(h_{X,\mathcal{D}}^{\tau-\frac{d}{2}}\|a\|_{L^2(\Omega;H^\tau(\mathcal{D}))} + \hbar\frac{\|f\|_{L^2(\mathcal{D})}}{a_{min}}\right)$$

holds.

Finally, we can also consider the multilevel method as a surrogate to the diffusion coefficient.

Theorem 8.3.8.
Suppose Assumption 8.3.4 holds. Let $\Phi \in L^1(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ be a positive definite function with Hilbert space $H_\Phi(\mathbb{R}^d) \cong H^\sigma(\mathbb{R}^d)$, where $\sigma > \frac{d}{2}$. Suppose $(X_j)_{j\in\mathbb{N}}$ is a sequence of point sets that fulfills the assumptions of Corollary 6.3.4. Let $a \in L^2(\Omega; H^\sigma(\mathcal{D}))$ be a second order random field which fulfills Assumption 8.1.2 and choose $L \in \mathbb{N}$ such that

$$\sup_{\omega\in\Omega} \|a - A_L(a)\|_{C(\mathcal{D})} \leq \frac{a_{\min}}{2}$$

holds, where $A_L(a)$ denotes the pathwise multilevel approximant of a given in Corollary 6.5.1. Denote by $u \in L^2(\Omega; H_0^1(\mathcal{D}))$ the weak solution of (8.1) with diffusion coefficient a and, for $0 < \hbar < 1$, by $u_\hbar^* \in L^2(\Omega; H_0^1(\mathcal{D}))$ the finite element approximation of the weak solution of (8.1) with diffusion coefficient $A_L(a)$. Then the error estimate

$$\|\mathcal{Q}(u) - \mathcal{Q}(u_\hbar^*)\|_{L^2(\Omega;V)} \leq C\|\mathcal{Q}\|_{L^2(\mathcal{D})\to V} \left(\beta^L \|a\|_{L^2(\Omega;H^\sigma(\mathcal{D}))} + \hbar \frac{\|f\|_{L^2(\mathcal{D})}}{a_{min}}\right)$$

holds, where β is given as in (6.6).

9 Outlook

In this thesis, we introduced several well-known kernel-based techniques, namely interpolation, penalized least squares approximation, and the multilevel method. Then, we applied these deterministic approximation methods to second order random fields and obtained a proper surrogate for this field. This was then used to replace the diffusion coefficient of a given elliptic partial differential equation. Afterwards, we showed how the error estimates for replacing the diffusion coefficient influences the error estimates for the corresponding solutions.

To this end, the main tool is Lemma 5.1.2, which allowed us to apply any linear and bounded operator B on the paths of the random field a. Of course, it is possible to use arbitrary approximation operators, as long as they satisfy the assumptions made in this lemma. Therefore, one is not restricted to the kernel-based methods presented here. It is also clear that one can calculate the covariance function of the approximated process explicitly, as long as the approximation method permits a Lagrange type representation. By proceeding as in Section 6.4, but particularly as in Remark 6.4.3, such a representation can easily be obtained for every linear approximation operator that only depends on finitely many point evaluations of the target function. Thus, an obvious generalization of the presented framework is its application with different approximation techniques.

However, the sample path regularity result stated in Theorem 4.3.5, which was needed in our approach, may also be of use in other context, although it is just a slight generalization of the known theorem given in [46, Theorem 1]. Up to this point, the author is not aware that this statement has already been proven.

From an approximation-theoretical point of view, there are several open questions regarding the multilevel methods: In the untensorized case, we have proven a novel basis representation for the method, which allows us to obtain a deeper understanding of its approximation space and the method itself. However, in most cases, we restricted ourselves to the case of nested point sets, which also played a crucial role in most of the representation results, while, for non-nested point sets, there seems to be no direct generalization of the presented techniques. Another interesting question is closely tied to the basis functions themselves: As proposed before, although the existence of such Lagrange type functions b_j is clear by linearity, little is known of numerical and analytical properties of these basis functions. In our case, such a Lagrange type representation can be proven for non-nested point sets as well, but we lose cardinality of the corresponding functions. Moreover, to properly show that the given functions are indeed independent and hence form a basis of the approximation space, we used linear independence of shifted and scaled translation invariant basis functions. While this can be shown for Wendland kernels, this still seems to be an open question for other kernels, e.g., Matérn kernels.

Secondly, in the tensorized case, particularly the multilevel method using tensorized kernels, there seems to be a gap in the proof of Corollary 7.1.7, the pointwise convergence of the method. Here, numerics suggest a convergence result using milder assumptions as

the ones stated. However, up to this point, there is no proof that circumvents the occurring problem concerning negative powers of the coupling parameters $\mu_{\mathcal{D}}$ and $\mu_{\mathcal{T}}$ and the restrictive smoothness assumptions. This error estimate probably needs a more refined analysis, with different or improved techniques.

Literature

[1] R. J. Adler and J. Taylor, *Random Fields and Geometry* (Springer monographs in mathematics). Springer, 2007. DOI: 10.1007/978-0-387-48116-6.

[2] M. S. Agranovich, *Sobolev Spaces, Their Generalizations and Elliptic Problems in Smooth and Lipschitz domains* (Springer monographs in mathematics). Springer International Publishing, 2015. DOI: 10.1007/978-3-319-14648-5.

[3] A. Alexanderian, "A brief note on the Karhunen-Loève expansion," 2015. arXiv: 1509.07526.

[4] R. Arcangéli, M. C. López de Silanes, and J. J. Torrens, "An extension of a bound for functions in Sobolev spaces, with applications to (m, s)-spline interpolation and smoothing," *Numerische Mathematik*, vol. 107, pp. 181–211, 2007.

[5] R. Arcangéli, M. C. López de Silanes, and J. J. Torrens, "Extension of sampling inequalities to Sobolev semi-norms of fractional order and derivative data," *Numerische Mathematik*, vol. 121, pp. 587–608, 2012. DOI: 10.1007/s00211-011-0439-3.

[6] N. Aronszajn, "Theory of reproducing kernels," *Transactions of the American Mathematical Society*, vol. 68, pp. 337–404, 1950.

[7] H. Bauer, *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. De Gruyter, 1978. DOI: 10.1515/9783111682457.

[8] P. Billingsley, *Probability and Measure* (A Wiley-Interscience publication), 3rd ed. Wiley, 1995.

[9] S. C. Brenner and R. Scott, *The Mathematical Theory of Finite Element Methods* (Texts in applied mathematics), 3rd ed. Springer, 2007. DOI: 10.1007/978-0-387-75934-0.

[10] T. T. Cai, Z. Ren, and H. H. Zhou, "Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation," *Electronic Journal of Statistics*, vol. 10, pp. 1–59, 2016. DOI: 10.1214/15-EJS1081.

[11] T. T. Cai, C.-H. Zhang, and H. H. Zhou, "Optimal rates of convergence for covariance matrix estimation," *The Annals of Statistics*, vol. 38, pp. 2118–2144, 2010. DOI: 10.1214/09-AOS752.

[12] A. Cohen, R. DeVore, and C. Schwab, "Convergence rates of best N-term Galerkin approximations for a class of elliptic sPDEs," *Found. Comput. Math.*, vol. 10, pp. 615–646, 2010. DOI: 10.1007/s10208-010-9072-2.

[13] S. De Marchi and H. Wendland, "On the convergence of the rescaled localized radial basis function method," *Applied Mathematics Letters*, vol. 99, p. 105 996, 2020. DOI: 10.1016/j.aml.2019.105996.

[14] R. Eubank and T. Hsing, *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators* (Wiley Series in Probability and Statistics). Wiley-Blackwell, 2015. DOI: 10.1002/9781118762547.

[15] L. Evans, *Partial Differential Equations*, 2nd ed. American Mathematical Society, 2010.

[16] M. Floater and A. Iske, "Multistep scattered data interpolation using compactly supported radial basis functions," *Journal of Computational and Applied Mathematics*, vol. 73, pp. 65–78, 1996. DOI: 10.1016/0377-0427(96)00035-0.

[17] E. Fuselier and G. B. Wright, "Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates," *SIAM Journal on Numerical Analysis*, vol. 50, pp. 1753–1776, 2012.

[18] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, 1st ed. Springer, 1991. DOI: 10.1007/978-1-4612-3094-6.

[19] Q. T. L. Gia, I. Sloan, and H. Wendland, "Multiscale analysis in Sobolev spaces on the sphere," *SIAM Journal on Numerical Analysis*, vol. 48, 2065–2090, 2010. DOI: 10.1137/090774550.

[20] L. Gollwitzer, R. Kempf, and H. Wendland, *Nodal representations for kernel-based multilevel interpolation*, 2026. arXiv: 2606.16643 [math.NA].

[21] W. Greub, *Multilinear Algebra*. Springer New York, 2012. DOI: 10.1007/978-1-4613-9425-9.

[22] M. Griebel and G. Li, "On the decay rate of the singular values of bivariate functions," *SIAM Journal on Numerical Analysis*, vol. 56, pp. 974–993, 2018. DOI: 10.1137/17M1117550.

[23] M. Griebel, G. Li, and C. Rieger, "On the numerical approximation of the Karhunen-Loève expansion for random fields with random discrete data," 2021. arXiv: 2112.02526 [math.NA].

[24] W. Hackbusch, *Tensor Spaces and Numerical Tensor Calculus*. Springer, 2012. DOI: 10.1007/978-3-642-28027-6.

[25] D. Haroske and H. Triebel, *Distributions, Sobolev Spaces, Elliptic Equations*. EMS, European Mathematical Society, 2007. DOI: 10.4171/042.

[26] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis, "Bochner spaces," in *Analysis in Banach Spaces : Volume I: Martingales and Littlewood-Paley Theory*. Springer International Publishing, 2016, pp. 1–66. DOI: 10.1007/978-3-319-48520-1_1.

[27] O. Kallenberg, *Foundations of Modern Probability* (Probability Theory and Stochastic Modelling), 3rd ed. Springer Nature, 2021. DOI: 10.1007/978-3-030-61871-1.

[28] T. Kato, *Perturbation Theory for Linear Operators* (Classics in Mathematics), 2nd ed. Springer, 1995.

[29] A. Kechris, *Classical Descriptive Set Theory*. Springer New York, 1995.

[30] R. Kempf and H. Wendland, "High-dimensional approximation with kernel-based multilevel methods on sparse grids," *Numerische Mathematik*, vol. 154, pp. 485–519, 2023. DOI: 10.1007/s00211-023-01363-x.

[31] H. König, *Eigenvalue Distribution of Compact Operators* (Operator Theory). Springer, 1986. DOI: 10.1007/978-3-0348-6278-3.

[32] F. Y. Kuo, C. Schwab, and I. H. Sloan, "Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients," *SIAM Journal on Numerical Analysis*, vol. 50, pp. 3351–3374, 2012.

[33] Q. T. Le Gia, I. H. Sloan, and H. Wendland, "Vector-valued Gaussian processes for approximating divergence- or rotation-free vector fields," 2025. arXiv: 2511.12535 [math.NA].

[34] W. A. Light and E. W. Cheney, *Approximation Theory in Tensor Product Spaces* (Lecture notes in mathematics). Springer, 1985. DOI: 10.1007/BFb0075391.

[35] G. J. Lord, C. E. Powell, and T. Shardlow, *An Introduction to Computational Stochastic PDEs*. Cambridge University Press, 2014. DOI: 10.1017/CBO9781139017329.

[36] M. Lukić and J. Beder, "Stochastic processes with sample paths in reproducing kernel Hilbert spaces," *Transactions of the American Mathematical Society*, pp. 3945–3969, 2001. DOI: 10.2307/2693779.

[37] H. Meschkowski, *Hilbertsche Räume mit Kernfunktionen*. Springer Berlin, Heildeberg, 1962. DOI: 10.1007/978-3-642-94848-0.

[38] F. J. Narcowich, J. D. Ward, and H. Wendland, "Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting," *Mathematics of Computation*, vol. 74, no. 250, pp. 743–763, 2005.

[39] F. J. Narcowich, J. D. Ward, and H. Wendland, "Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions," *Constructive Approximation*, vol. 24, pp. 175–186, 2006. DOI: 10.1007/s00365-005-0624-7.

[40] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, 1992.

[41] F. Nobile, R. Tempone, and C. G. Webster, "A sparse grid stochastic collocation method for partial differential equations with random input data," *SIAM Journal on Numerical Analysis*, vol. 46, pp. 2309–2345, 2008. DOI: 10.1137/060663660.

[42] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2005. DOI: 10.7551/mitpress/3206.001.0001. eprint: https://direct.mit.edu/book-pdf/2514321/book_9780262256834.pdf.

[43] T. Runst and W. Sickel, *Sobolev Spaces of Fractional Order, Nemytskij Operators, and Nonlinear Partial Differential Equations*. De Gruyter, 1996. DOI: 10.1515/9783110812411.

[44] R. A. Ryan, *Introduction to Tensor Products of Banach Spaces*, 2002nd ed. Springer. DOI: 10.1007/978-1-4471-3903-4.

[45] Y. Sawano, *Theory of Besov spaces* (Developments in Mathematics), 1st ed. Springer, 2018. DOI: 10.1007/978-981-13-0836-9.

[46] M. Scheuerer, "Regularity of the sample paths of a general second order random field," *Stochastic Processes and their Applications*, vol. 120, pp. 1879–1897, 2010. DOI: 10.1016/j.spa.2010.05.009.

[47] M. Stein, *Interpolation of Spatial Data*. Springer New York, 1999.

[48] I. Steinwart, "Convergence types and rates in generic Karhunen-Loève expansions with applications to sample path properties," 2017. arXiv: 1403.1040 [math.NA].

[49] I. Steinwart and A. Christmann, *Support Vector Machines*. Springer, 2008. DOI: 10.1007/978-0-387-77242-4.

[50] I. Steinwart and C. Scovel, "Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs," *Constructive Approximation*, vol. 35, pp. 363–417, 2012. DOI: 10.1007/s00365-012-9153-3.

[51] T. Sullivan, *Introduction to Uncertainty Quantification* (Texts in applied mathematics). Springer International Publishing, 2015.

[52] A. Townsend and H. Wendland, "Multiscale analysis in Sobolev spaces on bounded domains with zero boundary values," *IMA Journal of Numerical Analysis*, vol. 33, pp. 1095–1114, 2012. DOI: 10.1093/imanum/drs036.

[53] H. Triebel, "Interpolation theory, function spaces, differential operators," in ser. North-Holland mathematical library, Elsevier, 1978.

[54] N. Vakhania, V. Tarieladze, and S. Chobanyan, *Probability Distributions on Banach Spaces* (Mathematics and its Applications). Springer, 2011. DOI: 10.1007/978-94-009-3873-1.

[55] G. Wahba, *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, 1990. DOI: 10.1137/1.9781611970128.

[56] H. Wendland, *Scattered Data Approximation* (Cambridge Monographs on Applied and Computational Mathematics). Cambridge University Press, 2004. DOI: 10.1017/CBO9780511617539.

[57] H. Wendland, "Multiscale analysis in Sobolev spaces on bounded domains," *Numerische Mathematik*, vol. 116, pp. 493–517, 2010. DOI: 10.1007/s00211-010-0313-8.

[58] H. Wendland, "Multiscale radial basis functions," in *Frames and Other Bases in Abstract and Function Spaces*, Springer International Publishing, 2017, pp. 265–299.

[59] D. Werner, *Funktionalanalysis*, 8th ed. Springer Spektrum Berlin, Heidelberg, 2018. DOI: 10.1007/978-3-662-55407-4.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet habe.

Weiterhin erkläre ich, dass ich die Hilfe von gewerblichen Promotionsberatern bzw. -vermittlern oder ähnlichen Dienstleistern weder bisher in Anspruch genommen habe, noch künftig in Anspruch nehmen werde.

Zusätzlich erkläre ich hiermit, dass ich keinerlei frühere Promotionsversuche unternommen habe.

Bayreuth, den 25. Juni 2026

Lorenz Gollwitzer