Structure Analysis of Nonstandard Kernels for Multivariate Reconstructions

A Dissertation
submitted to the Department of Mathematics,
Faculty of Mathematics, Informatics and Natural Sciences,
Universität Hamburg

by

Juliane Entzian

Prof. Dr. Armin Iske & Prof. Ph.D. Elisabeth Larsson

Prof. Dr. Melanie Graf & Prof. Dr. Vicente Cortés & PD Dr. Alexander Lohse

Hamburg, October 17th, 2024

Declaration of Authorship | Eidesstattliche Erklärung

I, Juliane Entzian, hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated in the dissertation.

Furthermore, I declare that the electronic version of this dissertation coincides with the printed bound copy submitted to the faculty for archiving.

Hiermit versichere ich, Juliane Entzian, an Eides statt, dass ich die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Ich versichere außerdem, dass die elektronische Version dieser Dissertation mit der gebundenen, bei der Fakultät zur Archivierung eingereichten, Version übereinstimmt.

City, Date | Ort, Datum

Signature | Unterschrift

Declaration of Contributions | Eigenteilserklärung

This declaration describes which parts of the submitted dissertation were obtained without or with contributions of other people and what their respective contributions are. The thesis is concerned with the study of Summation Kernels (Chapter 3), Product Kernels (Chapter 4), Transformation Kernels (Chapter 5), Orthogonal Summation Kernels (Chapter 6), and Tensor Product Kernels (Chapter 7). Contributions are listed aligned with this partition:

(i) The results of Chapter 3, Chapter 4, and their texts are due to the candidate with no contribution of other people.

(ii) Chapter 5 deals with the same kernels, namely transformation kernels, as paper [AEI23a], co-authored with Kristof Albrecht and Armin Iske. However, none of the results from that paper are included in the chapter. Instead, Chapter 5 provides the broader theoretical foundation for the kernels used in the paper. The results and text of Chapter 5 are due to the candidate with no contribution of other people.

(iii) The results of Chapter 6 and its texts are due to the candidate with no contribution of other people.

(iv) Chapter 7 corresponds to the paper [AEI23b] on tensor product kernels, co-authored with Kristof Albrecht and Armin Iske. However, the candidate provides more detailed background and further results in Chapter 7. The proofs of Theorem 7.9, Theorem 7.10, Theorem 7.15 leading to the crucial statement of Theorem 7.25 (identical with [AEI23b, Theorem 2.3, Theorem 3.4, Theorem 4.4, and Theorem 4.6]), are joint work with equal contributions of Kristof Albrecht and the candidate. The parts on 'Newton Basis' and 'Convergence Rate' on page 144 ff. (aligning with [AEI23b, Section 5 and Section 6]) are reproduced in this dissertation with the permission of Kristof Albrecht. Section 7.3.3 extends [AEI23b, Theorem 4.9] and is due to the candidate, as is the text of Chapter 7.

List of Publications

[**AEI23a**] K. Albrecht, J. Entzian, and A. Iske.
'Anisotropic kernels for particle flow simulation.'
In *Modeling, Simulation and Optimization of Fluid Dynamic Applications.*
Springer, 2023.

[**AEI23b**] K. Albrecht, J. Entzian, and A. Iske.
'Product kernels are efficient and flexible tools for high-dimensional scattered interpolation.'
arXiv preprint arXiv:2312.09949 (2023).

Acknowledgments

First, I would like to thank my boyfriend, children, friends, and gym for providing the perfect balance in my life. You ensured that my time working on this thesis was kept short but productive.

I am grateful to my supervisor, Armin Iske, for providing me with the opportunity to pursue my PhD. I appreciate the freedom you gave me to explore my research independently, and thank you for supporting my ideas.

I was fortunate to work in a supportive, friendly, and appreciative environment, for which I thank all my colleagues. A special thanks goes to Kristof Albrecht, who was the ideal person to collaborate with during parental part-time work. Your companionship and support made this process not only manageable but enjoyable. I also thank him, Joshua Lampert, and Tizian Wenzel for their helpful comments and corrections in the process of refining this thesis. You helped to give it its final polish. Lastly, I want to thank my office mate Fabian Bleitner for providing help, feedback, and little chats whenever needed.

Abstract

This work concerns *adaptive kernel-based approximation methods*. We create a toolbox for adapting kernels to underlying problems, focusing on the interpolation of multivariate scattered data with an emphasis on anisotropies. By developing five nonstandard classes of flexible kernels – *transformation*, *summation*, and *product kernels*, as well as the anisotropic versions of the latter two *orthogonal summation*, and *tensor product kernels* – significant limitations of traditional radially symmetric kernels are addressed. These classes, some entirely new and others building on existing structures, provide the flexibility to select and combine kernels tailored to specific problems. Thus, they extend the variety of interpolation methods.

The theoretical analysis conducted on each kernel class's native space not only expands the understanding of native spaces in general but also enlightens underlying (name-giving) structures and their associated benefits. We investigate the interpolation method for each kernel, including impacts on accuracy and stability.

Numerical tests confirm the theoretical findings and show which kernel class is suitable for specific problem adaptations: We propose transformation or tensor product kernels for adapting to the point set; transformation kernels for adapting to the domain; and summation, transformation, or orthogonal summation kernels for adapting to the target function.

Zusammenfassung

Diese Arbeit befasst sich mit adaptiven, kernbasierten Approximationsmethoden. Wir entwickeln Werkzeuge um Kerne an das zugrunde liegende Problem anzupassen. Dabei fokussieren wir uns auf die Interpolation von mehrdimensionalen verstreuten Daten, mit besonderem Augenmerk auf Anisotropien. Durch die Entwicklung von fünf flexieblen Kernklassen – Transformationskerne, Summationskerne und Produktkerne sowie die anisotropen Versionen der letzten beiden, orthogonale Summationskerne und Tensorproduktkerne – werden wesentliche Einschränkungen traditioneller radial symmetrischer Kerne adressiert. Diese Klassen, teils völlig neu und teils auf bestehenden Strukturen bauend, bieten die Möglichkeit, Kerne anhand ihrer Eigenschaften und des zugrunde liegenden Problems auszuwählen oder zu kombinieren. Sie erweitern damit die Vielfalt und Anpassungsfähigkeit der Interpolationsmethoden.

Die theoretische Analyse der nativen Räume, die für jede Kernklasse durchgeführt wird, erweitert nicht nur das Verständnis von nativen Räumen im Allgemeinen, sondern deckt auch die zugrunde liegenden (namensgebenden) Strukturen und deren damit verbundenen Vorteile auf. Wir untersuchen die Interpolationsmethode für jeden Kern und betrachten die Auswirkungen auf Genauigkeit und Stabilität.

Numerische Tests bestätigen die theoretischen Ergebnisse und zeigen, welche Kernklassen für eine bestimmte Anpassung geeignet sind: Wir empfehlen Transformations- oder Tensorproduktkerne zur Anpassung an die Punktmenge; Transformationskerne zur Anpassung an den Definitionsbereich; sowie Summations-, Transformations- oder orthogonale Summationskerne zur Anpassung an die Zielfunktion.

Contents

Part I

Preliminaries

Chapter 1

Introduction

Imagine a growing child. Each year on their birthday, the parents measure the child and carve a corresponding notch into the doorframe. But how tall was the child at five and a half years old? What happened during the intervals between the notches on the doorframe? Answering these 'in-between' questions is known as interpolation, from the Latin *inter* meaning 'between'. The process of interpolation can become significantly more complex than our simple growth example. For instance, determining the positions of celestial bodies on ephemerides, which Babylonian astronomers achieved around 300 B.C. through interpolation, marking the beginning of the history of interpolation. While our example considers one-dimensional units of time (year) and height (meter), ephemerides involve a three-dimensional astronomic coordinate system. Today, more than ever, we should shift our focus from the celestial bodies back to our Earth. The reconstruction of functions is indispensable in climate research. Multidimensional spatial vectors of measurement stations are associated with data such as temperature, salinity, and humidity. We aim to reconstruct the underlying functions to determine the values of interest at any location on Earth and not just at the locations of the measuring stations. As the title indicates, this work is concerned with such multivariate reconstructions. Specifically, we focus on kernel-based interpolations, which approximate functions with a linear combination of basis functions generated by kernels. These methods offer the advantage of approximating functions using samples taken from an unorganized set of multidimensional points, known as scattered data fitting, and form the foundation for techniques used to handle noisy data.

Currently, most kernels considered assign a radial influence area to a sample, meaning the influence of the sample is uniform in all directions. Such radial kernels, including Gaussians, (inverse) multiquadrics, and polyharmonic splines, have been proven to be powerful tools in various applications of multivariate scattered data approximation. However, no method is perfect. Kernel-based approximation methods also have their challenges. For example, trade-offs must be made concerning computational cost and storage versus accuracy, and stability versus accuracy, as discussed, e.g., in G.E. Fasshauer's work [Fas07].

Moreover, the radial nature of these kernels is a limitation for some applications. Many materials exhibit anisotropic properties that cannot be captured by radial kernels.

For instance, wood has different mechanical properties along the grain, across the grain, and tangential to the growth rings, affecting its strength, stiffness, and thermal expansion. Similarly, sedimentary rocks show anisotropies in permeability that affect the flow of groundwater through rock layers, further illustrating the need for more flexible kernel functions. Additionally, fluid flows, whether in medical or climate research contexts, often exhibit anisotropies due to varying conditions and forces acting in different directions. In medical research, blood flow through arteries can show anisotropic behavior because of the complex interactions between blood cells and the vessel walls, as well as variations in vessel diameter and curvature. In climate research, ocean currents and atmospheric flows exhibit anisotropic characteristics due to the Earth's rotation, temperature gradients, and varying wind or water pressures in different directions. These anisotropic properties affect measurements such as flow velocity, direction, and pressure distribution. Understanding the anisotropic nature of ocean currents is crucial for accurate climate modeling and predicting the transport of heat and nutrients in marine environments.

Therefore, novel adaptive kernel methods are required. To identify possible adjustments of the interpolation method, we take a closer look at the interpolant itself.

As mentioned, the interpolant s of a real-valued function f on $\Omega \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, consists of a linear combination of basis functions. In kernel-based interpolation, these basis functions are derived from bivariate kernels $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ evaluated at the points $\{x_1, \ldots, x_N\} = X \subset \Omega$. Thus,

$$s(x) := \sum_{i=1}^{N} c_i K(x_i, x) \quad \text{for all } x \in \Omega,$$

where $c = (c_1, \ldots, c_N)^T \in \mathbb{R}^N$ is chosen such that s equals f on X, i.e.,

$$s(x_j) = \sum_{i=1}^{N} c_i K(x_i, x_j) = f(x_j) \quad \text{for } j = 1, \ldots, N.$$

Hence, the interpolant s depends on three factors: firstly, the target function f evaluated at the points of X, secondly, the point set X, and thirdly, the kernel K. To improve the interpolation method, we can align these three factors. The target function f is immutable, and aside from any known properties, it is generally unknown.

The point set X can be freely chosen in some applications, such as sensor placements in environmental monitoring, or it can be so large that significant time and storage benefits are achieved by considering only a subset. These settings enable a suitable selection of the point set X. Since the late 1990s, so-called greedy methods have been developed for this purpose. Initially, the f-greedy algorithm [SW00] was developed. It selects the point set in regard to the target function f, resulting in many points being placed where significant changes in f occur. However, closely placed points lead to poor stability, prompting the development of a selection algorithm focused on stability, the P-greedy algorithm [DMSW05], and various combinations of both, as discussed in [WSH23] and the references therein. The selection of points in X depends on the choice of the kernel K for each of these greedy algorithms, providing a first reason to examine kernels. Moreover, in many situations, the point set X is predetermined and cannot be altered – for example, when sensor locations are fixed or when dealing with historical measurements. This underscores the need to investigate kernels, the third and final factor in adapting interpolants, where the option of employing greedy methods and their associated improvements remains at our disposal.

The theory on kernels dates back to the mid-20th century. J. Mercer's theorem has been known since 1909 [Mer09]. Building on earlier work by E.H. Moore [Moo39] and others, N. Aronszajn [Aro50] developed the theory of reproducing kernel Hilbert spaces in the 1940s. N. Aronszajn's article provides a thorough overview of the early history and the first applications of kernels. Comprehensive theoretical discussions on positive (semi)-definite kernels and their properties regarding scattered data interpolation can be found in the works of M. Buhmann [Buh03], H. Wendland [Wen05], and A. Iske [Isk18]. Up until now, kernels are adapted to the underlying problem in two ways: Firstly, the kernel's differentiability is inherited by the interpolant, so knowledge about the differentiability of the function f should influence the choice of the kernel. Secondly, shape parameters of radial kernels have been used to tailor the kernel to specific problems, e.g. in [KC92], [LF05], [MVHÖ23].

With this work, we aim to expand the toolbox for adapting kernels to underlying problems and provide a corresponding theoretical analysis. Our goal is to overcome the limitations of traditional radial kernels and offer accurate and flexible methods for multivariate scattered data approximation, with a focus on anisotropies.

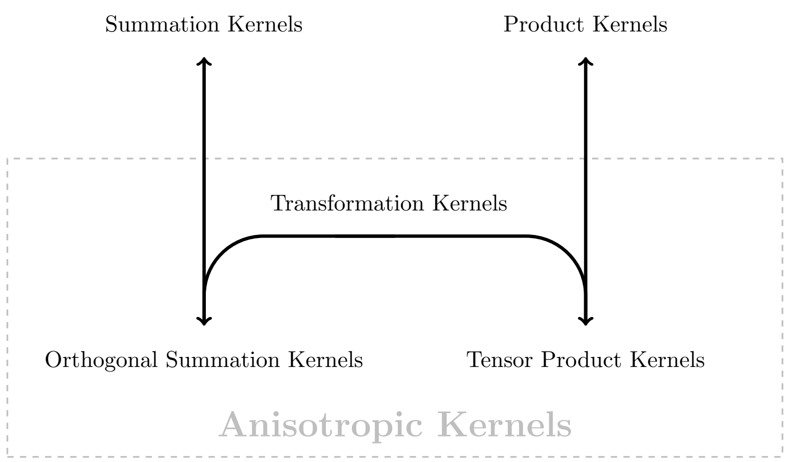

Figure 1.1: Schematic relation of summation, product, transformation, orthogonal summation, and tensor product kernels.

To this end, we construct novel kernels, analyze their structure, and examine their utility for anisotropic problems. We consider the product and summation of kernels that remain radial if their components are radial and focus on anisotropic kernels. These include transformation kernels as well as orthogonal summation kernels and tensor product kernels. Fig. 1.1 schematically depicts relations between these kernels.

Below, we summarize the main contributions of this thesis. A more detailed presentation of the contributions and benefits, as well as a distinction from previously known results of earlier papers regarding each kernel, can be found in the corresponding introductions.

Investigation on Approaches for Combining or Adapting Kernels

- Summation Kernels – Chapter 3: The summation of kernels provides insights into reproducing kernel Hilbert spaces and their norms, which we further develop.
- Product Kernels – Chapter 4: We place the product of kernels within a broader picture.
- Transformation Kernels – Chapter 5: We extend the principle of shape parameters to general transformations.
- Orthogonal Summation Kernels – Chapter 6: We present a novel kernel, where the summation of component kernels acting on low-dimensional spaces build a kernel acting on the high-dimensional Cartesian product of the component spaces. This allows a flexible adaptation of the component kernels to properties of the corresponding low-dimensional space.
- Tensor Product Kernels – Chapter 7: While these kernels are used in statistics, we provide an investigation with regard to kernel-based interpolation. A tensor product kernel on a high-dimensional space equals the product of various component kernels acting on low-dimensional spaces. This allows a flexible adaptation of a component kernel to properties of the corresponding low-dimensional space.

Examination of the Native Spaces' Structure

- Summation Kernels – Section 3.2: We further develop the structure analysis of summation kernels' native spaces by connecting N. Aronszajn's findings to Sobolev spaces (Section 3.2.2), examining intersections of native spaces (Section 3.2.3), linking Mercer's Theorem to summation kernels (Section 3.2.4), and classifying kernels into equivalence classes (Remark 3.21).
- Transformation Kernels – Section 5.2: We provide a relation between the transformation kernel's native space and the one corresponding to its initial kernel (Theorem 5.6).
- Orthogonal Summation Kernels – Section 6.2: We prove the native space to be structured as an orthogonal sum, providing the name of the kernel (Theorem 6.7).
- Tensor Product Kernels – Section 7.2: We present an alternative proof for the tensor structure of its native space, which lends the kernel its name (Section 7.2.2).

Analysis of Interpolation Methods

- Summation Kernels – Section 3.3: We provide a thorough analysis of the summation kernel's interpolation method, finding that interpolations using a kernel from a larger equivalence class result in inferior approximations (Section 3.3.1), and that a summation kernel exhibits the stability of its most stable component (Section 3.3.2). This yields a trade-off principle for kernels, which is schematically visualized in Remark 3.42.
- Transformation Kernels – Section 5.3: We provide a thorough analysis of accuracy and stability, and deduce conditions under which a special transformation kernel outperforms its initial kernel (Section 5.3.1 and 5.3.2).
- Orthogonal Summation Kernels – Section 6.3: We investigate the DC-strictly positive definiteness of the kernel, examine the conditions under which improved accuracy can be expected (Section 6.3.1), and find that the stability of the orthogonal summation kernel aligns with that of its most stable component (Section 6.3.2).
- Tensor Product Kernels – Section 7.3: We provide a thorough analysis of the kernel's positive definiteness by introducing the concept of grid-like structured data sets (resulting in Theorem 7.25), and analyze the interpolation process on these sets (Section 7.3.1). Furthermore, we develop statements regarding stability (Section 7.3.3).

Implementation and Evaluation

- Summation Kernels – Section 3.4: We demonstrate the effect of a kernel's equivalence class on the interpolation method.
- Transformation Kernels – Section 5.4: We demonstrate how accuracy or stability can be improved without significantly affecting the other by using transformation kernels adapted to the target (Section 5.4.1) and adapted to the domain and point set (Section 5.4.2).
- Orthogonal Summation Kernels – Section 6.4: We demonstrate the outstanding performance of orthogonal summation kernels in anisotropic sum structures.
- Tensor Product Kernels – Section 7.4: We demonstrate how to improve accuracy while maintaining controllable stability by using tensor product kernels (Section 7.4.1), and how these kernels enable a speedup of the interpolation process in grid like settings (Section 7.4.2).

In total, we investigate five novel classes of more flexible kernels. Some of these classes are entirely new inventions, while others already existed. For summation kernels and tensor product kernels, the existing structural analysis of the native spaces is developed further. In the case of transformation kernels and orthogonal summation kernels, the analysis is entirely new. Although some of these kernels have been applied occasionally in the past, a comprehensive investigation of all of them, with a focus on interpolation methods, was lacking. We provide this comprehensive examination, with a particular emphasis on accuracy and stability.

The thesis is divided into four parts. The Preliminaries (Part I) set the stage for the detailed exploration and contributions that follow. In Chapter 2, we provide foundational theoretical principles of kernel-based interpolation.

Part II investigates Combinations of Kernels, which are not necessarily anisotropic. This includes summation kernels (Chapter 3) and product kernels (Chapter 4).

Part III delves into constructions that lead to Anisotropic Kernels, encompassing transformation kernels (Chapter 5), orthogonal summation kernels (Chapter 6), and tensor product kernels (Chapter 7). We explore how general principles apply to these specific types of kernels, examine their unique characteristics and how they can be exploited for interpolation purposes.

The Final Remarks (Part IV) include a summary of the thesis, discusses broader implications of the research, and outlines potential directions for future work (Conclusion and Outlook). Additionally, reference materials are provided here.

Each of the Chapters 3 – 7, which deal with a specific nonstandard kernel, follow a consistent structure: starting with a clear definition and basic properties (first section), followed by a detailed analysis of the native spaces (second section), and culminating in their application for kernel-based interpolation (third section). The theoretical results are complemented by supportive numerical tests (fourth section), which validate and illustrate the practical implications of our findings. A detailed outline of each chapter can be found in the respective introduction.

Chapter 2

Kernel-Based Interpolation

This chapter provides fundamentals on kernel-based interpolation methods. We closely follow the structure of Chapter 2 and 3 in Kristof Albrecht's thesis [Alb24]. For a comprehensive and extensive treatment, we recommend the works of H. Wendland [Wen05], and A. Iske [Isk18].

Section 2.1 motivates using positive definite kernels for scattered data interpolations. In Section 2.2, we explore the properties of positive definiteness, translation-invariance, and radial symmetry of kernels. An examination of native spaces is conducted in Section 2.3, starting with reproducing kernel Hilbert spaces (Section 2.3.1), progressing through their construction (Section 2.3.2), and their structure and properties (Section 2.3.3). Additionally, we introduce the Sobolev space as a special native space in (Section 2.3.4). Understanding native spaces enables precise statements regarding the interpolation method, which is the focus of Section 2.4. Here, we emphasize accuracy (Section 2.4.1) and numerical stability (Section 2.4.2), culminating in a discussion on the trade-off principle between these two aspects (Section 2.4.3).

2.1 Essentials

We aim at finding a function $s : \Omega \longrightarrow \mathbb{R}$, $\Omega \subseteq \mathbb{R}^d$, that satisfies the *interpolation condition*

$$s(x_i) = f_i \qquad \text{for } i = 1, \dots, N \tag{2.1}$$

for $N \in \mathbb{N}$, arbitrarily given interpolation points $X = \{x_1, \dots, x_N\} \subseteq \Omega$, and arbitrary function values $f_1, \dots, f_N \in \mathbb{R}$. Such a function is called an *interpolant* of the target function f on the point set X. Here, we demand Ω to contain inner points, as this is the case for most applications. Still, theoretically, it is possible to interpolate on a finite set Ω. The interpolant s is commonly restricted to the span of predetermined basis functions $\mathcal{B} = \{b_1, \dots, b_N\} \subset \{f : \Omega \longrightarrow \mathbb{R}\}$, so that it is given by the linear combination

$$s = \sum_{i=1}^{N} c_i \, b_i,$$

where the coefficient vector $c = (c_1, \dots, c_N)^T \in \mathbb{R}^N$ solves the linear system

$$\big(b_i(x_j)\big)_{i,j=1}^{N} c = f_X \qquad \text{for } f_X = (f_1, \dots, f_N)^T \in \mathbb{R}^N. \tag{2.2}$$

Definition 2.1. The evaluation matrix of basis functions $\mathcal{B} = \{b_1, \ldots, b_N\}$ on a point set $X = \{x_1, \ldots, x_M\}$ is called *Vandermonde matrix*

$$V_{\mathcal{B},X} = \big(b(x)\big)_{\substack{x \in X \\ b \in \mathcal{B}}} = \begin{pmatrix} b_1(x_1) & \ldots & b_N(x_1) \\ \vdots & \ddots & \vdots \\ b_1(x_M) & \ldots & b_N(x_M) \end{pmatrix} \in \mathbb{R}^{M \times N}.$$

A unique coefficient vector $c \in \mathbb{R}^N$ for the interpolant s is guaranteed if the Vandermonde matrix $V_{\mathcal{B},X}$ is regular. This in turn is the case if the basis $\mathcal{B}$ spans a Haar space. See [Haa10] and [Wen05, Definition 2.1] for a detailed discussion.

Definition 2.2. Let $\Omega \subseteq \mathbb{R}^d$ contain at least N points and $V \subseteq C(\Omega)$ be an N-dimensional subspace of continuous functions on Ω. V is called a *Haar space* of dimension N on Ω if for arbitrary pairwise distinct point sets $\{x_1, \ldots, x_N\} \subseteq \Omega$ and arbitrary function values $f_1, \ldots, f_N$ there exists exactly one function $s \in V$ fulfilling the interpolation condition (2.1).

We see that V is an N-dimensional Haar space if and only if for any pairwise distinct point set $\{x_1, \ldots, x_N\} \subseteq \Omega$ and any basis $\{b_1, \ldots, b_N\}$ of V it is

$$\det \big(b_i(x_j)\big)_{i,j=1}^N \neq 0. \tag{2.3}$$

In the case where $\Omega \subseteq \mathbb{R}^d$, $d \geqslant 2$, contains an interior point, the following counterexample for the existence of a Haar space can be constructed.

Example 2.3. As $\Omega \subseteq \mathbb{R}^d$, $d \geqslant 2$, contains an interior point, there exists $x_0 \in \Omega$ and $\delta > 0$ so that $B(x_0, \delta) \subseteq \Omega$. It is possible to find a pairwise distinct point set $X = \{x_1, \ldots, x_N\} \subset B(x_0, \delta)$ for every $N \in \mathbb{N}$. Let $N \geqslant 2$. Since $d \geqslant 2$ we can move x_1 and x_2 along the continuous curves $x_1(t), x_2(t)$ in Ω, where $t \in [0, 1]$, such that the curves have no intersection with X except for $x_1 = x_1(0) = x_2(1)$ and $x_2 = x_2(0) = x_1(1)$. If $V = \operatorname{span}\{b_i : i = 1, \ldots, N\}$ is a Haar space, the function

$$D(t) = \det \Big(b_i\left(x_j(t)\right)\Big)_{i,j=1}^N$$

is continuous on $[0, 1]$ and never equal to zero. However, as the first two rows switch between $t = 0$ and $t = 1$, it is $D(0) = -D(1)$. By the intermediate value theorem, there exist a time $t_0 \in [0, 1]$ where $D(t_0) = 0$. Hence, the Vandermonde matrix for the point set $\{x_1(t_0), x_2(t_0)\} \cup X$ is not regular. This contradicts (2.3) and V cannot be a Haar space.

This results in the Mairhuber-Curtis theorem proven by J.C. Mairhuber and P.C. Curtis Jr in the 1950s (cf. [CJ59], [Mai56], [Wen05, Theorem 2.3]).

Theorem 2.4 (Mairhuber-Curtis)**.** *Let $\Omega \subseteq \mathbb{R}^d$, $d \geqslant 2$, contain an interior point, then there exists no Haar space on Ω of dimension $N \geqslant 2$.*

Still, a unique interpolation is possible if we restrict the point set X to be a grid (cf. [Wen05, Lemma 2.8]). In this thesis, however, we are interested in a more general interpolation scheme and focus on scattered point sets $X = \{x_1, \dots, x_N\}$. Hence, we need another approach and make each basis function b_i depend on one point $x_i \in X$ for $i = 1, \dots, N$. This gives rise to the concept of bivariate functions

$$K : \Omega \times \Omega \longrightarrow \mathbb{R}, \qquad \Omega \subseteq \mathbb{R}^d,$$

acting on the Cartesian product $\Omega \times \Omega$, which we call *kernels* acting on Ω.

Definition 2.5. The *Cartesian product* of two sets Ω_1 and Ω_2 is defined by

$$\Omega_1 \times \Omega_2 := \left\{(x, y) \ : \ x \in \Omega_1, \ y \in \Omega_2\right\}.$$

This definition can be extended to the Cartesian product of finitely many sets Ω_ℓ for $\ell = 1, \dots, M$, i.e.,

$$\bigtimes_{\ell=1}^{M} \Omega_\ell := \left\{(x_1, \dots, x_M) \ : \ x_\ell \in \Omega_\ell \text{ for } \ell = 1, \dots, M\right\}.$$

We define the kernel basis functions

$$\begin{aligned} b_i &:= K(x_i, \cdot) : \Omega \longrightarrow \mathbb{R} \qquad \text{for } i = 1, \dots, N \text{ and} \\ \mathcal{B}_{K,X} &:= \{b_1, \dots, b_N\}. \end{aligned} \tag{2.4}$$

The Vandermonde matrix $V_{\mathcal{B}_{K,X},X}$ of the kernel basis in (2.4) evaluated on the point set $X = \{x_1, \dots, x_N\}$ is given by

$$\mathbf{A}_{K,X} := \Big(K\left(x_i, x_j\right)\Big)_{i,j=1}^{N} \in \mathbb{R}^{N \times N}. \tag{2.5}$$

This matrix is called the *interpolation matrix* of the kernel K regarding X and the linear system (2.2) turns into

$$\mathbf{A}_{K,X}\, c = f_X. \tag{2.6}$$

To find a unique solution $c \in \mathbb{R}^N$ and hence a unique *interpolant*

$$s_{f,K,X} := \sum_{i=1}^{N} c_i\, K(x_i, \cdot), \tag{2.7}$$

lying in the *interpolation space*

$$S_{K,X} := \operatorname{span}\left\{K(x, \cdot) \ : \ x \in X\right\}, \tag{2.8}$$

the interpolation matrix $\mathbf{A}_{K,X}$ needs to be regular for any pairwise distinct point sets $X \subset \Omega$. If the context is clear, we occasionally omit parts of the subscripts of the interpolant $s_{f,K,X}$ in (2.7) for simplicity. We want to exclude the possibility of Example 2.3. Therefore, we demand $\det\left(\mathbf{A}_{K,X}\right) \neq 0$ for any pairwise distinct point set X. This certainly includes all subsets of X.

Definition 2.6. Let $r \in \{1, \ldots, N\}$ and $A \in \mathbb{R}^{N \times N}$ be a quadratic matrix. A *principal submatrix* $A_r \in \mathbb{R}^{N-r \times N-r}$ of A is given by deleting r times the same row and column of A.

By Sylvester's criterion a symmetric matrix is positive definite if and only if all its principal minors (determinants of principal submatrices) are positive, see [HJ91, Theorem 7.2.5]. Hence, we demand the interpolation matrix to be positive definite.

Let us summarize and conclude this discourse. For a kernel K, that induces positive definite interpolation matrices $\mathbf{A}_{K,X}$ for arbitrary pairwise distinct data sets X, the interpolant $s \in S_{K,X}$ of the form (2.7) is unique. These kernels are the primary focus of this thesis, and the subsequent sections provide fundamentals of the corresponding research.

2.2 Positive Semi-Definite Kernels

In Section 2.1 we saw that a positive definite interpolation matrix leads to a unique interpolant. Here, we state the definition of positive (semi-)definite kernels, fulfilling this requirement, and have a look into some basic properties. Furthermore, we introduce the eminent cases of translation-invariant and radial kernels in order to state Bochner's and Schoenberg's characterization for positive definiteness. Each characterization is equipped with one supporting example of a positive definite kernel.

Definition 2.7. A bivariate function $K : \Omega \times \Omega \longrightarrow \mathbb{R}$, $\Omega \subseteq \mathbb{R}^d$, is called *positive semi-definite kernel* on Ω if for any arbitrary set of pairwise distinct centers $X = \{x_1, \ldots, x_N\} \subset \Omega$, $N \in \mathbb{N}$, the interpolation matrix $\mathbf{A}_{K,X}$ is positive semi-definite, i.e.,

$$c^T \mathbf{A}_{K,X} c = \sum_{j=1}^{N} \sum_{k=1}^{N} c_j c_k \, K(x_j, x_k) \geqslant 0 \qquad \text{for all } c = (c_1, \ldots, c_N)^T \in \mathbb{R}^N. \tag{2.9}$$

We say the kernel is *positive definite* on Ω, if $\mathbf{A}_{K,X}$ is positive definite, i.e., (2.9) is strictly positive for all $c \in \mathbb{R}^N \backslash \{0\}$.

We can directly deduce the following properties.

Lemma 2.8. *Let K be a positive semi-definite kernel on $\Omega \subseteq \mathbb{R}^d$ then*

(i) $K(x,x) \geqslant 0$ and $K(x,x) > 0$ if K is positive definite for all $x \in \Omega$.

(ii) the kernel aK is positive semi-definite for all $a \geqslant 0$, and aK is positive definite for all $a > 0$ if K was positive definite.

In the following chapters, we will solely work with *symmetric* positive (semi-) definite kernels, i.e.,

$$K(x,y) = K(y,x) \qquad \text{for all } x, y \in \Omega.$$

This is because these kernels are closely related to reproducing kernel Hilbert spaces, which we consider in Section 2.3.

Lemma 2.9. *Let K be a symmetric positive semi-definite kernel on $\Omega \subseteq \mathbb{R}^d$ then*

(i) the estimate $K(x,y)^2 \leqslant K(x,x)K(y,y)$ holds for all $x, y \in \Omega$.

(ii) K is bounded on $\Omega \times \Omega$ if and only if K is bounded on the diagonal

$$\{(x,x) \,:\, x \in \Omega\} \subset \Omega \times \Omega.$$

(iii) K vanishes if and only if K equals zero on the diagonal.

Proof. Let $X = \{x, y\} \subset \Omega$. As $\mathbf{A}_{K,X}$ is positive semi-definite, it is

$$0 \leqslant \det\left(\mathbf{A}_{K,X}\right) = K(x,x)K(y,y) - K(x,y)^2 \qquad \text{for all } x, y \in \Omega,$$

which yields (i). The second statement is a direct consequence of (i), and property (iii) immediately follows from (ii). ∎

Given Section 2.1, investigating positive semi-definite kernels may seem superfluous since we cannot guarantee a unique solution to the interpolation problem without any further steps. In Section 2.3, however, it turns out that symmetric positive semi-definite kernels are reproducing kernels of Hilbert spaces, which in turn include the set of all possible interpolants. The analysis of these so-called native spaces provides insights into the interpolants' properties such as structure and approximation quality. Lemma 2.10 below, shows that the pairwise distinctiveness of X is not a necessary condition for the positive semi-definiteness of $\mathbf{A}_{K,X}$. This fundamental observation is important in Chapter 7 and Chapter 6, where anisotropic product and summation kernels are studied.

Lemma 2.10. *Let $X_N = \{x_1 \ldots, x_N\} \subseteq \Omega$ be a pairwise distinct point set and K a positive semi-definite kernel on Ω. Furthermore, let $x_1 = x_{N+1}$ and $X_{N+1} = X_N \cup \{x_{N+1}\}$. Then the interpolation matrix $\mathbf{A}_{K,X_{N+1}}$ is positive semi-definite.*

Proof. We can write the interpolation matrix as a block matrix

$$\mathbf{A}_{K,X_{N+1}} = \begin{pmatrix} \mathbf{A}_{K,X_N} & \mathbf{A}_{K,X_N}\, e_1 \\ e_1^T\, \mathbf{A}_{K,X_N} & K(x_{N+1}, x_{N+1}) \end{pmatrix},$$

where e_1 denotes the first unit vector. Let $c' \in \mathbb{R}^N, c_{N+1} \in \mathbb{R}$ be arbitrarily chosen and

$$c = \left(c', c_{N+1}\right) \in \mathbb{R}^{N+1}.$$

Then

$$\begin{aligned} c^T\, \mathbf{A}_{K,X_{N+1}}\, c &= c'^T\, \mathbf{A}_{K,X_N}\, c' + (c_{N+1}e_1)^T\, \mathbf{A}_{K,X_N}\, c' \\ &\qquad + c'^T\, \mathbf{A}_{K,X_N}\, c_{N+1}e_1 + c_{N+1}^2 K(x_{N+1}, x_{N+1}) \\ &= \left(c' + c_{N+1}e_1\right)^T\, \mathbf{A}_{K,X_N}\, \left(c' + c_{N+1}e_1\right) \\ &\geqslant 0, \end{aligned}$$

since $c_{N+1}^2 K(x_{N+1}, x_{N+1}) = c_{N+1}^2 K(x_1, x_1) = (c_{N+1}e_1)^T\, \mathbf{A}_{K,X_N}\, (c_{N+1}e_1)$. ∎

Remark 2.11. The above lemma implies two findings:

1. For the matrix $\mathbf{A}_{K,X}$ to be positive definite, it is necessary for X to be pairwise distinct. This is because with $c_{N+1} \neq 0$ and $c' = -c_{N+1}e_1$ the vector $c = (c', c_{N+1}) \neq 0$ but $c^T \, \mathbf{A}_{K,X_{N+1}} \, c = 0$.

2. Lemma 2.10 can quickly be generalized to the assertion: If K is a positive semi-definite kernel on Ω and $X \subseteq \Omega$ is an arbitrary finite set of data points, the interpolation matrix $\mathbf{A}_{K,X}$ is positive semi-definite.

Next, we introduce two subclasses of kernels. For each of them, a characterization regarding their positive definiteness is presented. This, in turn, helps in finding specific examples of positive definite kernels.

A *translation* is a mapping $T_\xi : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ given by $x \longmapsto \xi - x$ for a fixed value $\xi \in \mathbb{R}^d$. If K is invariant regarding all translations T_ξ it is in particular invariant regarding T_x. We obtain

$$K(x,y) = K(T_x(x), T_x(y)) = K(x-x, x-y) = K(0, x-y) \qquad \text{for all } x, y \in \mathbb{R}^d.$$

Hence, K can be viewed as a function acting on one variable only, i.e.,

$$\Phi(x) := K(0,x) \qquad \text{for all } x \in \mathbb{R}^d.$$

Remark 2.12. For every *translation-invariant* kernel $K : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$ there exists a univariate function $\Phi : \mathbb{R}^d \longrightarrow \mathbb{R}$ such that

$$K(x,y) = \Phi(x-y) \qquad \text{for all } x, y \in \mathbb{R}^d.$$

We call Φ a positive (semi-)definite function if the corresponding kernel K is positive (semi-)definite.

The subsequent statements directly follow from Remark 2.12, Lemma 2.8, and Lemma 2.9.

Lemma 2.13. *Let K be a positive semi-definite translation-invariant kernel on $\Omega \subseteq \mathbb{R}^d$ with univariate function Φ. Then $0 \leqslant \Phi(0)$ and*

$$K(x-z, y-z) = K(x,y) \qquad \text{for all } x, y, z \in \Omega.$$

If K is also symmetric, then

(i) Φ *is an even function, i.e.,* $\Phi(-x) = \Phi(x)$ *for all* $x \in \Omega$.

(ii) Φ *is bounded, i.e.,* $|\Phi(x)| \leqslant \Phi(0)$ *for all* $x \in \Omega$.

(iii) $\Phi(0) = 0$ *if and only if* $\Phi \equiv 0$.

We introduce a second, even smaller, yet important subclass of kernels. This is the class of kernels K that are invariant under translations and rotations, i.e., they satisfy

$$K(x,y) = K\left(T_\xi(x), T_\xi(y)\right) \text{ and } K(x,y) = K(Ax, Ay) \qquad \text{for all } x, y \in \mathbb{R}^d,$$

all translations T_ξ and all rotation matrices A. For every $\xi \in \mathbb{R}^d$ there exist a rotation matrix $A_\xi \in \mathbb{R}^{d\times d}$ such that $A_\xi \xi = \|\xi\|_2 e_1$, where e_1 denotes the first unit vector. This yields

$$\begin{aligned} K(x,y) &= K\left(A_{(x-y)}x, A_{(x-y)}y\right) = \Phi\left(A_{(x-y)}x - A_{(x-y)}y\right) \\ &= \Phi\left(A_{(x-y)}(x-y)\right) = \Phi\left(\|x-y\|_2 e_1\right) \qquad \text{for all } x, y \in \Omega. \end{aligned}$$

This enables us to describe the kernel K using a function ϕ acting on a single dimension

$$\phi(\|x-y\|_2) := \Phi\left(\|x-y\|_2 e_1\right) = K(x,y) \qquad \text{for all } x, y \in \Omega.$$

Remark 2.14. For a translation and rotation invariant kernel $K : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$ there exists a one-dimensional *radial basis function* (RBF) $\phi : [0,\infty) \longrightarrow \mathbb{R}$ such that

$$K(x,y) = \phi(\|x-y\|_2) \qquad \text{for all } x, y \in \Omega.$$

We call ϕ positive (semi-)definite if the corresponding kernel function K is positive (semi-)definite. Note, that these kernels are as well referred to as *radially symmetric* or simply *radial* kernels.

Let us focus on the question under which circumstances a bi-variate kernel is positive (semi-)definite. To do so, we state Bochner's characterization for translation-invariant kernels first and thereafter Schoenberg's characterization for radial kernels. Furthermore, these characterizations are provided with one example of a positive definite kernel each.

Already in the 1930s, S. Bochner linked the positive definiteness of translation-invariant kernels to a non-negative Fourier transform (cf. [Boc32], [Boc33]). Here, we state the L^1-version of Bochner's characterization, also presented in [Isk18, Theorem 8.7].

Theorem 2.15 (Bochner)**.** *Let $\Phi \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ be an even function. Then Φ is positive definite on $\mathbb{R}^d$ if and only if its Fourier transform $\widehat{\Phi}$ is non-negative and non-vanishing.*

Example 2.16 (Askey)**.** The radial characteristic kernel of R. Askey (cf. [Ask73]), also called truncated power function, is given by the RBF $\phi_\ell : [0,\infty) \longrightarrow \mathbb{R}$, where

$$\phi_\ell(r) = (1-r)_+^\ell = \begin{cases} (1-r)^\ell & \text{for } r \leqslant 1, \\ 0 & \text{else.} \end{cases}$$

By Bochner's characterization this yields a positive definite kernel on $\mathbb{R}^d$ provided that $\ell \in \mathbb{N}$ satisfies $\ell \geqslant \lfloor d/2 \rfloor + 1$. For a detailed proof we refer to [Wen05, Theorem 6.20].

In 1938, I.J. Schoenberg established a relation between positive semi-definite radial kernels and completely monotone functions (cf. [Sch38], [Wen05, Definition 7.4, Theorem 7.14]).

Definition 2.17. A function φ is called *completely monotone* on $[0, \infty)$ if it satisfies $\varphi \in C([0, \infty)) \cap C^\infty(0, \infty)$ and

$$(-1)^\ell \, \varphi^{(\ell)}(r) \geqslant 0 \qquad \text{for all } \ell \in \mathbb{N}_0,\ r > 0,$$

where $\varphi^{(\ell)}$ denotes the ℓ^{th} derivative of φ.

Theorem 2.18 (Schoenberg)**.** *A RBF $\phi : [0, \infty) \longrightarrow \mathbb{R}$ is positive definite on every $\mathbb{R}^d$ if and only if $\phi(\sqrt{\cdot})$ is completely monotone on $[0, \infty)$ and not constant.*

Example 2.19 (Gaussian)**.** The Gaussian kernel, visualized in Fig. 2.1, is given by

$$K(x, y) = \mathrm{e}^{-\alpha \, \|x-y\|_2^2}, \qquad \alpha > 0.$$

Obviously this is a radial kernel with the RBF $\phi(r) = \mathrm{e}^{-\alpha \, r^2}$. Since

$$\phi(\sqrt{r}) = \mathrm{e}^{-\alpha \, r} =: \varphi(r) \in C([0, \infty)) \cap C^\infty(0, \infty)$$

and

$$(-1)^\ell \, \varphi^{(\ell)}(r) = (-1)^\ell (-\alpha)^\ell \, \mathrm{e}^{-\alpha \, r} \geqslant 0 \qquad \text{for all } r > 0,$$

the requirements of Schoenberg's characterization hold. This implies positive definiteness of the Gaussian on every $\mathbb{R}^d$.

2.3 Native Spaces

With regard to the interpolation problem discussed in Section 2.1, it is crucial to determine which target function f can be approximated arbitrarily well by which kernel. It is immediately apparent that this is the case if f lies in the union of all possible interpolation spaces $S_{K,X}$ for $X \subseteq \Omega$,

$$S_{K,\Omega} := \operatorname{span}\{K(x, \cdot) \ : \ x \in \Omega\} = \bigcup_{X \subseteq \Omega} S_{K,X}, \tag{2.10}$$

or its closure. This section is concerned with the construction of the so-called native space of K. First, in Section 2.3.1, we introduce the general concept of reproducing kernel Hilbert spaces and draw the connection to symmetric positive definite kernels and their native spaces in Section 2.3.2. We go into the structure and properties of native spaces in Section 2.3.3 and finish in Section 2.3.4 with a special native space, the Sobolev space.

2.3.1 Reproducing Kernel Hilbert Spaces

This section addresses the general concept of reproducing kernel Hilbert spaces. We start with its definition, as given in [Aro50, p. 343], and deduce significant properties in Theorem 2.21. These are used to prove uniqueness results in Theorem 2.22.

Definition 2.20. The Hilbert space $\big(H, \langle\cdot,\cdot\rangle\big)$ of functions $f : \Omega \longrightarrow \mathbb{R}$ is called a *reproducing kernel Hilbert space* (RKHS) if there exists a *reproducing kernel* $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ that satisfies

(i) $K(x,\cdot) \in H$ for all $x \in \Omega$ and

(ii) the *reproducing property*, i.e., $f(x) = \langle f, K(x,\cdot)\rangle$ for all $f \in H$ and all $x \in \Omega$.

Let us state some basic properties of reproducing kernels, see [Wen05, Theorem 10.3 and 10.4] or [Aro50, p. 344].

Theorem 2.21. *Let* $\big(H, \langle\cdot,\cdot\rangle\big)$ *be a RKHS with reproducing kernel* $K : \Omega \times \Omega \longrightarrow \mathbb{R}$. *Then*

(i) K is symmetric and positive semi-definite.

(ii) the set $S_{K,\Omega}$ lies dense in H.

(iii) the sequence $(f_n)_{n\in\mathbb{N}} \subseteq H$ converges pointwise to $f \in H$ if $(f_n)_{n\in\mathbb{N}}$ converges normwise to f.

Proof.

(i) The symmetry of K follows from the symmetry of the inner product and the reproducing property of K,

$$K(x,y) = \big\langle K(x,\cdot), K(y,\cdot)\big\rangle = \big\langle K(y,\cdot), K(x,\cdot)\big\rangle = K(y,x) \qquad \text{for all } x,y \in \Omega.$$

Let $X = \{x_1,\ldots,x_N\} \subseteq \Omega$ and $c = (c_1,\ldots,c_N)^T \in \mathbb{R}^N$ then

$$\begin{aligned} c^T\, \mathbf{A}_{K,X}\, c &= \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j\, K(x_i,x_j) = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j\, \big\langle K(x_i,\cdot), K(x_j,\cdot)\big\rangle \\ &= \left\langle \sum_{i=1}^{N} c_i\, K(x_i,\cdot), \sum_{j=1}^{N} c_j\, K(x_j,\cdot) \right\rangle = \left\| \sum_{i=1}^{N} c_i\, K(x_i,\cdot) \right\|^2 \geqslant 0. \end{aligned}$$

(ii) By Definition 2.20 (i) and the fact that H as a Hilbert space is linear, we obtain $S_{K,\Omega} \subset H$. Since H is complete, $\overline{S_{K,\Omega}}$ is a closed and linear subset of H. From functional analysis (cf. [Mus14, Theorem 4.6]), we know that in this case, H can be orthogonally decomposed into

$$H = \overline{S_{K,\Omega}} \oplus \overline{S_{K,\Omega}}^{\perp},$$

where '$\perp$' denotes the orthogonal complement. Let $f \in \overline{S_{K,\Omega}}^{\perp}$ be arbitrarily chosen, then

$$f(x) = \langle f, K(x,\cdot)\rangle = 0 \qquad \text{for all } x \in \Omega,$$

by the reproducing property and the fact that $K(x,\cdot) \in S_{K,\Omega}$. This implies $\overline{S_{K,\Omega}}^{\perp} = \{0\}$ and henceforth $\overline{S_{K,\Omega}} = H$.

(iii) Let the sequence $(f_n)_{n\in\mathbb{N}} \subseteq H$ converge normwise to $f \in H$, i.e.,

$$\|f_n - f\| \longrightarrow 0 \qquad \text{for } n \to \infty.$$

This implies

$$|f_n(x) - f(x)| = \left|\left\langle f_n - f, K(x,\cdot)\right\rangle\right| \leqslant \|f_n - f\| \, \|K(x,\cdot)\| \longrightarrow 0$$

for all $x \in \Omega$ and $n \to \infty$.

■

The following theorem proves that every reproducing kernel has a unique RKHS and that every RKHS is has a unique reproducing kernel. For a detailed discussion, see [Wen05, Theorem 10.11] and [Aro50, p. 343].

Theorem 2.22. *The following uniqueness properties hold.*

(i) Let K be the reproducing kernel for $\left(H_1, \langle\cdot,\cdot\rangle_1\right)$ and $\left(H_2, \langle\cdot,\cdot\rangle_2\right)$, then $H_1 = H_2$ and $\langle\cdot,\cdot\rangle_1 = \langle\cdot,\cdot\rangle_2$.

(ii) Let K_1 and K_2 be reproducing kernels of $\left(H, \langle\cdot,\cdot\rangle\right)$ then $K_1 \equiv K_2$.

Proof.

(i) It is

$$\langle K(x,\cdot), K(\cdot,y)\rangle_1 = K(x,y) = \langle K(x,\cdot), K(\cdot,y)\rangle_2 \qquad \text{for all } x, y \in \Omega,$$

so that the inner products coincide on $S_{K,\Omega}$. Furthermore, $S_{K,\Omega} \subset H_1, H_2$ lies dense in both spaces, by Theorem 2.21 *(ii)*. Since norms coincide on $S_{K,\Omega}$, a sequence $(f_n)_{n\in\mathbb{N}} \subset S_{K,\Omega}$ is a Cauchy sequence in $\left(H_1, \langle\cdot,\cdot\rangle_1\right)$ if and only if it is a Cauchy sequence in $\left(H_2, \langle\cdot,\cdot\rangle_2\right)$. Then there exist functions $f \in H_1$, $g \in H_2$ with

$$\|f_n - f\|_1, \|f_n - g\|_2 \longrightarrow 0 \qquad \text{for } n \to \infty.$$

Using Theorem 2.21 *(iii)*, we conclude

$$f(x) = \lim_{n\to\infty} f_n(x) = g(x) \qquad \text{for all } x \in \Omega.$$

This yields $H_1 = H_2$. Moreover, we obtain

$$\|f\|_1 = \lim_{n\to\infty} \|f_n\|_1 = \lim_{n\to\infty} \|f_n\|_2 = \|f\|_2 \qquad \text{for all } f \in H_1 = H_2.$$

The polarization identity implies equality of the inner products.

(ii) By the assumption and the reproducing property of K_1 and K_2 it is

$$\begin{aligned}
&\|K_1(x,\cdot) - K_2(x,\cdot)\|^2 \\
&\quad = \left\langle (K_1 - K_2)(x,\cdot), (K_1 - K_2)(x,\cdot)\right\rangle \\
&\quad = \left\langle (K_1 - K_2)(x,\cdot), K_1(x,\cdot)\right\rangle - \left\langle (K_1 - K_2)(x,\cdot), K_1(x,\cdot)\right\rangle \\
&\quad = (K_1 - K_2)(x,x) - (K_1 - K_2)(x,x) \\
&\quad = 0 \qquad \text{for all } x \in \Omega.
\end{aligned}$$

This implies $K_1(x,\cdot) = K_2(x,\cdot)$ for all $x \in \Omega$ and hence $K_1 \equiv K_2$.

■

2.3.2 Native Space Construction

The preceding analysis establishes a one-to-one relations between symmetric positive semi-definite kernels and RKHSs. Every reproducing kernel is symmetric and positive semi-definite by Theorem 2.21. This prompts the question of whether every symmetric and positive semi-definite kernel reproduces a RKHS. Indeed, it can. To demonstrate this, we follow the approach outlined in [Wen05, Chapter 10.2].

Let K be a symmetric and positive semi-definite kernel on $\Omega \subseteq \mathbb{R}^d$. We equip the space of linear combinations

$$S_{K,\Omega} = \operatorname{span}\{K(\cdot, x) \ : \ x \in \Omega\},$$

introduced in (2.10), with the bilinear form

$$\langle f, g\rangle_K := \sum_{i=1}^{N}\sum_{j=1}^{M} c_i d_j \ K(x_i, y_j), \tag{2.11}$$

where $f, g \in S_{K,\Omega}$ are of the form

$$f = \sum_{i=1}^{N} c_i \ K(x_i, \cdot) \qquad \text{and} \qquad g = \sum_{j=1}^{M} d_j \ K(y_j, \cdot). \tag{2.12}$$

Obviously, the reproducing property holds on $S_{K,\Omega}$, i.e.,

$$f(x) = \sum_{i=1}^{N} c_i \ K(x_i, x) = \langle f, K(x, \cdot)\rangle_K.$$

Lemma 2.23. *The mapping $\langle \cdot, \cdot\rangle_K : S_{K,\Omega} \times S_{K,\Omega} \longrightarrow \mathbb{R}$ defined in (2.11) is well-defined, bilinear, symmetric and positive definite.*

Proof. In addition to (2.12) let $f, g \in S_{K,\Omega}$ have the representations

$$f = \sum_{i=1}^{\tilde{N}} \tilde{c}_i \ K(\tilde{x}_i, \cdot) \qquad \text{and} \qquad g = \sum_{j=1}^{\tilde{M}} \tilde{d}_j \ K(\tilde{y}_j, \cdot)$$

then

$$\begin{aligned} \langle f, g\rangle_K &= \sum_{i=1}^{N}\sum_{j=1}^{M} c_i d_j \ K(x_i, y_j) = \sum_{i=1}^{N} c_i \ g(x_i) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{\tilde{M}} c_i \tilde{d}_j \ K(\tilde{y}_j, x_i) = \sum_{j=1}^{\tilde{M}} \tilde{d}_j \ f(\tilde{y}_j) = \sum_{i=1}^{\tilde{N}}\sum_{j=1}^{\tilde{M}} \tilde{c}_i \tilde{d}_j \ K(\tilde{x}_i, \tilde{y}_j). \end{aligned}$$

Hence, the mapping $\langle \cdot, \cdot\rangle_K$ is well-defined. Clearly it is bilinear, and its symmetry is given by the symmetry of K. The property of definiteness is due to the reproducing property and the Cauchy-Schwartz inequality, which is valid as $\langle \cdot, \cdot\rangle_K$ is bilinear. Let $f \in S_{K,\Omega}$ and $\langle f, f\rangle_K = 0$. Then

$$\left|f(x)\right|^2 = \left|\langle f, K(x, \cdot)\rangle_K\right|^2 \leqslant \langle f, f\rangle_K \ K(x, x) = 0 \qquad \text{for all } x \in \Omega.$$

This implies $f \equiv 0$ on Ω. ∎

We deduce that for any symmetric and positive definite kernel K, the set $S_{K,\Omega}$, equipped with the norm $\|\cdot\|_K^2 = \langle\cdot,\cdot\rangle_K$, is a pre-Hilbert space. Furthermore, every metric space can be completed by [Mus14, Theorem 4.6]. Consequently, we denote the completion of the pre-Hilbert space $S_{K,\Omega}$ as the Hilbert space

$$\mathcal{H}_{K,\Omega} := \big\{\text{normwise limits of a Cauchy sequence } (s_n)_{n\in\mathbb{N}} \subset S_{K,\Omega}\big\}$$

equipped with the inner product

$$\langle f,g\rangle_K := \lim_{n\to\infty}\langle s_n,t_n\rangle_K,$$

where $(s_n)_{n\in\mathbb{N}},\ (t_n)_{n\in\mathbb{N}} \subset S_{K,\Omega}$ are Cauchy sequences that converge normwise to f and g in $\mathcal{H}_{K,\Omega}$. However, the elements of the completion are abstract elements, and we need to interpret them as functions. To do so, we define

$$f(x) := \lim_{n\to\infty} s_n(x) = \lim_{n\to\infty}\langle s_n, K(x,\cdot)\rangle_K = \langle f, K(x,\cdot)\rangle_K \qquad \text{for all } f\in\mathcal{H}_{K,\Omega},\ x\in\Omega.$$

In fact, any Cauchy sequence in $S_{K,\Omega}$ that converges pointwise to zero also converges to zero in the norm sense, see [BTA11, Theorem 2]. Hence, we are able to characterize the native space $\mathcal{H}_{K,\Omega}$ of a symmetric and positive semi-definite kernel K as done in Theorem 2.24 below. The existence of native spaces dates back to N. Aronszajn who based his thoughts on E.H. Moores results. For more information and detailed proofs we refer to [BTA11, Chapter 3] or [Wen05, Chapter 10].

Theorem 2.24 (Moore-Aronszajn)**.** *Let K be a symmetric and positive semi-definite kernel on $\Omega \subseteq \mathbb{R}^d$. Then*

$$\mathcal{H}_{K,\Omega} = \big\{f \textit{ is the pointwise limit of a Chauchy sequence } (s_n)_{n\in\mathbb{N}} \subset S_{K,\Omega}\big\}$$

equipped with the inner product

$$\langle f,g\rangle_K := \lim_{n\to\infty}\langle s_n,t_n\rangle_K,$$

where $(s_n)_{n\in\mathbb{N}},\ (t_n)_{n\in\mathbb{N}} \subset S_{K,\Omega}$ are Cauchy sequences that converge pointwise to f and g in $\mathcal{H}_{K,\Omega}$, is a reproducing kernel Hilbert space with reproducing kernel K. The pre-Hilbert space $S_{K,\Omega}$ lies dense in $\mathcal{H}_{K,\Omega}$.

The reproducing kernel Hilbert space $(\mathcal{H}_{K,\Omega},\langle\cdot,\cdot\rangle_K)$ is called native space *of K.*

We note, that the Moore-Aronszajn theorem, together with Theorem 2.21 and Theorem 2.22, establishes a one-to-one relation between symmetric and positive semi-definite kernels K and reproducing kernel Hilbert spaces $\big(H,\langle\cdot,\cdot\rangle\big)$. In the following, we use the notation $\big(\mathcal{H}_{K,\Omega},\langle\cdot,\cdot\rangle_K\big)$ for a reproducing kernel Hilbert space with symmetric and positive semi-definite reproducing kernel K on Ω. When it is clear which space is being referred to, we omit Ω in the subscript for simplicity. Conversely, if we wish to emphasize the domain, we use $\langle\cdot,\cdot\rangle_{K,\Omega}$.

2.3.3 Structure and Properties of Native Spaces

In the context of the interpolation problem described in Section 2.1, certain properties of the target function f, such as continuity or differentiability, may be known. To ensure that these properties are preserved in the interpolant, it is crucial to understand which properties functions in a reproducing kernel Hilbert space (RKHS) inherit from its reproducing kernel. This section, drawing on [SC08, Chapter 4], examines how the kernel K influences the properties of $\mathcal{H}_{K,\Omega}$. Additionally, we explore issues related to the separability and dimensions of native spaces, as well as uniqueness.

Definition 2.25. Let $\Omega \subseteq \mathbb{R}^d$. A kernel K on Ω is

(i) *bounded* if $\|K\|_\infty := \max_{x\in\Omega} \sqrt{K(x,x)} < \infty$.

(ii) *separately continuous* if $K(x,\cdot)$ is continuous for all $x \in \Omega$.

(iii) *m-times continuously differentiable* for $m \geqslant 0$ if $\partial^{\alpha,\alpha} K : \Omega \times \Omega \longrightarrow \mathbb{R}$ exists and is continuous for all multi-indices $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leqslant m$, where

$$\partial^{\alpha,\alpha} := \partial_1^{\alpha_1} \dots \partial_d^{\alpha_d} \partial_{d+1}^{\alpha_1} \dots \partial_{2d}^{\alpha_d} \qquad \text{for a multi-index } \alpha \in \mathbb{N}_0^d.$$

Equipped with these definitions, we state [SC08, Theorem 4.23, 4.28, 4.33, and Corollary 4.36] and [BTA11, Corollary 4] in the following theorem. For the corresponding proofs we refer to the given sources.

Theorem 2.26. *Let K on $\Omega \subseteq \mathbb{R}^d$ be the reproducing kernel of $\left(\mathcal{H}_{K,\Omega}, \langle\cdot,\cdot\rangle_K\right)$. Then the following statements hold.*

(i) K is bounded if and only if every $f \in \mathcal{H}_{K,\Omega}$ is bounded.

(ii) K is bounded and separately continuous if and only if every $f \in \mathcal{H}_{K,\Omega}$ is bounded and continuous.

(iii) If K is continuous, $\mathcal{H}_{K,\Omega}$ is separable.

(iv) If K is continuous and bounded,

$$K(x,y) = \sum_{i=1}^{\infty} e_i(x) e_i(y) \qquad \text{for all } x, y \in \Omega,$$

where $(e_i)_{i\in\mathbb{N}}$ is any orthonormal system in $(\mathcal{H}_{K,\Omega}, \langle\cdot,\cdot\rangle_K)$.

(v) If Ω is an open set and K is m-times continuously differentiable for $m \geqslant 0$, it is $\mathcal{H}_{K,\Omega} \subset C^m$ and

$$|\partial^\alpha f(x)| \leqslant \|f\|_K \left(\partial^{\alpha,\alpha} K(x,x)\right)^{1/2} \qquad \text{for all } f \in \mathcal{H}_{K,\Omega},\ x \in \Omega$$

and for every $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leqslant m$.

Let us take a closer look at separable native spaces. While, more general results can be found in [OS17], the following focuses on exploring the dimensions of $\mathcal{H}_{K,\Omega}$ here. This focus is due to the fact that separable Hilbert spaces of identical dimensions share equivalent structural properties. Although the concepts of the *power function* and the *trade-off principle* are discussed in detail in Section 2.4, we introduce them here in order to show Theorem 2.28. We adopt the definition of the power function provided by [SH17] in this context.

Definition 2.27. Let K be a positive semi-definite kernel on Ω and $X \subset \Omega$ a finite set of data points. The *power function* $P_{K,X}$ is given by

$$P_{K,X}(x) := \|K(\cdot, x) - s_{K(\cdot,x),K,X}\|_K \qquad \text{for all } x \in \Omega,$$

where $s_{K(\cdot,x),X}$ denotes the orthogonal projection of $K(\cdot, x)$ onto $S_{K,X}$.

The power function $P_{K,X}(x)$ equals zero if and only if the function $K(x, \cdot)$ lies in $S_{K,X}$. In other words, the power function is positive if and only if $K(x, \cdot)$ is linearly independent of the set of functions $\{K(y, \cdot)\}_{y\in X}$. Section 2.4, concerned with the trade-off principle, provides a lower bound for the power function while simultaneously offering an upper bound for the minimal eigenvalue $\lambda_{\min}\left(\mathbf{A}_{K,X}\right)$ of $\mathbf{A}_{K,X}$, i.e.,

$$\lambda_{\min}\left(\mathbf{A}_{K,X}\right) \leqslant \min_{1\leqslant i\leqslant N} P^2_{K,X\setminus\{x_i\}}(x_i) \qquad \text{for all } X = \{x_1, \ldots, x_N\}.$$

For a continuous positive definite kernel K and an arbitrary pairwise distinct point set $X \subset \Omega$, the interpolation matrix $\mathbf{A}_{K,X}$ is positive definite. Hence, its minimal eigenvalue $\lambda_{\min}\left(\mathbf{A}_{K,X}\right)$ is positive. In the case where $\Omega \subset \mathbb{R}^d$ is finite, we have $\mathcal{H}_{K,\Omega} = S_{K,\Omega}$. Furthermore, we see that $\{K(y, \cdot)\}_{y\in\Omega}$ is linearly independent, since

$$0 < \lambda_{\min}\left(\mathbf{A}_{K,\Omega}\right) \leqslant P_{K,\Omega\setminus\{x\}}(x) \qquad \text{for every} x \in \Omega.$$

Showing that in this case $\dim \mathcal{H}_{K,\Omega} = |\Omega|$.

Let $\Omega \subset \mathbb{R}^d$ contain inner points or be infinitely countable. Then for every finitely pairwise distinct point set $X \subset \Omega$ we can find $x_0 \in \Omega\setminus X$. Using the same arguments as above, $K(x_0, \cdot)$ is linearly independent of $\{K(y, \cdot)\}_{y\in X}$. Hence, the dimension of $\mathcal{H}_{K,\Omega}$ must be infinite and because of Theorem 2.26 *(iii)* there is a countable basis. We summarize the above analysis in the following theorem.

Theorem 2.28. *Let K_1, K_2 be continuous, symmetric and positive definite kernels on $\Omega \subseteq \mathbb{R}^d$, then their native spaces are isometric isomorph to another, i.e.,*

$$\mathcal{H}_{K_1,\Omega} \simeq \mathcal{H}_{K_2,\Omega}.$$

Proof. The above analysis shows

$$\dim \mathcal{H}_{K_1,\Omega} = \dim \mathcal{H}_{K_2,\Omega}.$$

Two separable Hilbert spaces H_1 and H_2 are isometric isomorph if they have the same dimension, see [Wer18, Satz V.4.12]. Theorem 2.26 *(iii)* proves the assumption. ■

The following lemma demonstrates that multiplying a kernel by a positive constant reproduces the same set of functions as the original kernel, and the corresponding norms are identical up to the positive constant. Referring to the uniqueness results in Theorem 2.22, this lemma underscores that the uniqueness of the reproducing kernel pertains to the Hilbert space $(H, \langle\cdot,\cdot\rangle)$ – which includes both the space H and its inner product – rather than the space H alone. This topic will be addressed in Section 3.2, where we analyze the native spaces of summation kernels.

Lemma 2.29. *Let $a > 0$ and K be a symmetric positive semi-definite kernel on Ω then $\mathcal{H}_{aK,\Omega} = \mathcal{H}_{K,\Omega}$ and the corresponding norms satisfy the relation $a\|\cdot\|_{aK} = \|\cdot\|_K$.*

Proof. First we note that aK is still a symmetric and positive semi-definite kernel by Lemma 2.8. Hence, it is reasonable to look at its native space. We see that $f \in S_{K,\Omega}$ if and only if

$$f = \sum_{i=1}^{N} c_i \, K(x_i, \cdot) = \sum_{i=1}^{N} \frac{c_i}{a} \, aK(x_i, \cdot) \in S_{aK,\Omega}$$

and the following relation of the norm holds

$$\|f\|_K = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j \, K(x_i, x_j) = a \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{c_i}{a}\frac{c_j}{a} \, aK(x_i, x_j) = a\|f\|_{aK}\,.$$

As the norms $\|\cdot\|_K$ and $\|\cdot\|_{aK}$ are equivalent, the completions $\mathcal{H}_K$ and $\mathcal{H}_{aK}$ of the set $S_{K,\Omega}$ for the respective norms coincide, see [Hac12, Remark 4.2]. Let $f \in \mathcal{H}_{K,\Omega}$ be the normwise limit of $(s_n)_{n\in\mathbb{N}} \subset S_{K,\Omega}$. Then

$$\|f\|_K = \lim_{n\to\infty} \|s_n\|_K = \lim_{n\to\infty} a\|s_n\|_{aK} = a\|f\|_{aK}\,.$$

∎

2.3.4 Sobolev Spaces

We dedicate this section to an eminent native space, the Sobolev space. Theorem 2.30 shows that native spaces of translation-invariant kernels K can be characterized in terms of Fourier transforms, as do Sobolev spaces, see Theorem 2.32. We deduce a reproducing kernel of Sobolev spaces, the Matérn kernel (Ex. 2.33), and introduce another compactly supported kernel, the Wendland kernel (Theorem 2.34). Both kernels reproduce Sobolev spaces with equivalent norms.

For the proof of the subsequent theorem we refer to [Wen05, Theorem 10.12].

Theorem 2.30. *Let K be a symmetric positive semi-definite and translation-invariant kernel on $\mathbb{R}^d$ regarding the univariate function $\Phi \in C(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$. Then*

$$\mathcal{H}_{K,\mathbb{R}^d} = \left\{ f \in L^2(\mathbb{R}^d) \cap C(\mathbb{R}^d) \;:\; \widehat{f} \Big/ \sqrt{\widehat{\Phi}} \in L^2(\mathbb{R}^d) \right\}$$

is its associated native space. The space is a Hilbert space equipped with inner product

$$(f, g)_{\mathcal{H}_{K,\mathbb{R}^d}} = (2\pi)^{-d/2} \left(\widehat{f}/\sqrt{\widehat{\Phi}},\; \widehat{g}/\sqrt{\widehat{\Phi}} \right)_{L^2(\mathbb{R}^d)} = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \frac{\widehat{f}(\omega)\,\widehat{g}(\omega)}{\widehat{\Phi}(\omega)} d\omega.$$

Let us recall some basics of Sobolev spaces. These spaces were originally defined to classify weak solutions of partial differential equations. They consist of equivalence classes of functions whose weak derivatives up to a certain degree lie in a certain L^p space (cf. [McL00, Chapter 3]). Here, we restrict $p = 2$.

Definition 2.31. Let D^α denote the weak derivative for a muli-index $\alpha \in \mathbb{N}^d$. The space

$$W^m(\mathbb{R}^d) := \left\{ f \in L^2(\mathbb{R}^d) \ : \ D^\alpha f \in L^2(\mathbb{R}^d) \text{ for all } |\alpha| \leqslant m \right\}$$

equipped with the inner product

$$\langle f, g \rangle_{W^m(\mathbb{R}^d)} := \left(\sum_{|\alpha| \leqslant m} \langle D^\alpha f, D^\alpha g \rangle_{L^2(\mathbb{R}^d)} \right)^{\frac{1}{2}}$$

is called *Sobolev space* of order $m \in \mathbb{N}$.

These spaces can be generalized to $m \in \mathbb{R}_{>0}$, as discussed, for example, in [DNPV12]. Furthermore, they are Hilbert spaces and equivalent to the so called fractional Sobolev spaces, as $m \in \mathbb{R}$ is allowed, or Bessel potential spaces (cf. [McL00, Theorem 3.18]).

Theorem 2.32. *For $m > d/2$, the Sobolev space $W^m(\mathbb{R}^d)$ is given by the following set of functions*

$$W^m(\mathbb{R}^d) = \left\{ f \in L^2(\mathbb{R}^d) \ : \ \widehat{f}(\cdot) \left(1 + \|\cdot\|_2^2\right)^{m/2} \in L^2(\mathbb{R}^d) \right\}$$

Furthermore, the norm induced by

$$\langle f, g \rangle_{W^m} := (2\pi)^{-d/2} \int_{\mathbb{R}^d} \widehat{f}(\omega) \widehat{g}(\omega) \left(1 + \|\omega\|_2^2\right)^m d\omega \tag{2.13}$$

is equivalent to $\|\cdot\|_{W^m(\mathbb{R}^d)}$.

Note, that Sobolev spaces $W^m(\mathbb{R}^d)$ are spaces of equivalence classes, due to the integral representations. Still, each class contains a C^m representer (cf. [Eva10, Chapter 5.2.1]). Here, we interpret $W^m \subset C^m$. Moreover, we can view the non-negative and non-vanishing weight function in (2.13) as a Fourier transform of a function Φ, i.e.,

$$\widehat{\Phi} = \left(1 + \|\cdot\|_2^2\right)^{-m}.$$

By Bochner's theorem, the function Φ is positive definite on $\mathbb{R}^d$, and it is possible to derive its explicit form up to a constant factor

$$\Phi(x) = \frac{\|x\|_2^{m-d/2} K_{m-d/2}(\|x\|_2)}{2^{m-1}\Gamma(m)} \qquad \text{for } m > d/2,$$

where K_ν denotes the modified Bessel function second kind of order ν (cf. [Fas07, Chapter 4.4]). Hence, the kernel given by $K(x, y) = \Phi(x - y)$ for $x, y \in \mathbb{R}^d$ is the reproducing kernel of $W^m(\mathbb{R}^d)$ and $\|\cdot\|_K$ is equivalent to $\|\cdot\|_{W^m(\mathbb{R}^d)}$ because of Lemma 2.29. This gives rise to the Matérn kernels, first introduced in [Mat86], and often used in statistics.

Example 2.33. The Matérn kernels, also called Sobolev splines because of their connection to Sobolev spaces, are radial kernels given by the RBFs

$$\psi_{\frac{d+1}{2}-m}(r) = \frac{r^{m-d/2}K_{m-d/2}(r)}{2^{m-1}\Gamma(m)} \qquad \text{for } m > d/2,$$

where K_ν denotes the Bessel function of the second kind of order ν. The corresponding radial kernel $K_\ell(x,y) = \psi_\ell(\|x-y\|_2)$ is positive definite on $\mathbb{R}^d$ for $d < 2m$. We list representatives of the Matérn RBFs in Tab. 2.1 and visualize them in Fig. 2.1.

Using Lemma 2.29 and the analysis above we deduce that the native space corresponding to the RBF ψ_ℓ is given by the Sobolev space $W^{(d+1)/2+\ell}(\mathbb{R}^d)$ with equivalent norms. Furthermore, we observe, that the functions ψ_ℓ do not depend on the dimension d. With Schoenberg's characterization, ψ_ℓ is positive definite on every $\mathbb{R}^d$. Additionally, we have $\psi_\ell(\|\cdot\|) = \Phi_\ell \in C^{2\ell}$. This implies $K_\ell(x,\cdot) = \Phi_\ell(x-\cdot) \in C^\ell$ for all $x \in \mathbb{R}^d$, and Theorem 2.26 yields $\mathcal{H}_{K,\mathbb{R}^d} \subset C^\ell(\mathbb{R}^d)$ for every dimension $d \in \mathbb{N}$. The differentiability of kernels gains on importance in Section 2.4.1. There, high differentiability is found to be associated with good approximations.

Definition	Native space	Differentiability of $\psi_\ell(\|\cdot\|)$
$\psi_0(r) = \mathrm{e}^{-r}$	$W^{\frac{d+1}{2}}$	C^0
$\psi_1(r) = (1+r)\,\mathrm{e}^{-r}$	$W^{\frac{d+3}{2}}$	C^2
$\psi_2(r) = (3+3r+r^2)\,\mathrm{e}^{-r}$	$W^{\frac{d+5}{2}}$	C^4

Table 2.1: Representatives of the Matérn RBFs up to a dimension dependent scaling factor

We introduce another family of positive definite kernels reproducing to Sobolev spaces, called the Wendland kernels. The following is based on Chapter 9 of H. Wendland's book [Wen05].

A kernel that satisfies

$$K(x,y) = 0 \qquad \text{if } \|x-y\|_2 \geqslant R$$

for some $R > 0$ is called *locally supported*. Such kernels are beneficial in interpolation theory since their interpolation matrices $\mathbf{A}_{K,X}$ are sparse. For now, we focus on radial positive definite kernels K with one-dimensional function ϕ. Hence, K is locally supported if and only if there exist $R > 0$ such that $\phi(r) = 0$ for all $r \geqslant R$. Using a general formulation of Bochner's theorem it is possible to show that a continuous, univariate and compactly supported function Φ cannot be positive definite on every $\mathbb{R}^d$, see [Wen05, Corollary 9.3]. We point out, that the characteristic kernels and the Matérn kernels, introduced in Ex. 2.16 and Ex. 2.33 respectively, align with this result. The Matérn kernels are positive definite for every dimension and have global support, whereas the characteristic kernels ϕ_ℓ are locally supported, and their positive definiteness is restricted to $\mathbb{R}^d$, with $\ell \geqslant \lfloor d/2 \rfloor + 1$. This provides reason for working with fixed dimensions. H. Wendland constructed piece-wise polynomial functions with compact support using the radial approach

$$\phi(r) = \begin{cases} p(r) & 0 \leqslant r \leqslant 1, \\ 0 & r > 1, \end{cases} \tag{2.14}$$

where p denotes a one-dimensional polynomial. Note, that the truncated power function already is a first example. We state [Wen05, Theorem 9.13] and [Wen05, Theorem 10.35] without proof, and refer to the same reference for a precise definition of the polynomials $p_{d,k}$.

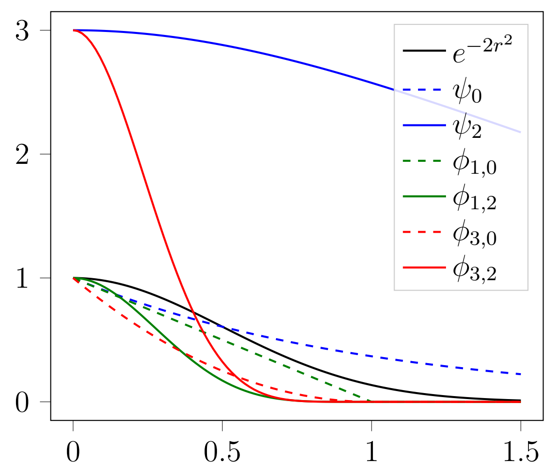

Figure 2.1: Visualization of the RBFs given in Ex. 2.19: Gaussian with $\alpha = 2$ (black), Tab. 2.1: Matérn's ψ_ℓ (blue), and Tab. 2.2: Wendland's $\phi_{1,\ell}$ (green) and $\phi_{3,\ell}$ (red) for $\ell = 0, 2$.

Theorem 2.34. *The RBFs* $\phi_{d,k}$ *of the form*

$$\phi_{d,k}(r) = \begin{cases} p_{d,k}(r) & 0 \leqslant r \leqslant 1, \\ 0 & r > 1, \end{cases}$$

with a one-dimensional polynomial $p_{d,k}$ *of degree* $\lfloor d/2 \rfloor + 3k + 1$ *are*

(i) positive definite on $\mathbb{R}^d$.

(ii) Their symmetric extensions lie in $C^{2k}(\mathbb{R})$.

(iii) They are of minimal degree for given space dimension d.

(iv) If ϕ *is another function of the form* (2.14) *and with* $C^{2k}(\mathbb{R})$ *extension, it is* $\phi = c\,\phi_{d,k}$ *for a constant* $c > 0$.

These RBFs give rise to the Wendland kernels.

Theorem 2.35. *Let* $\phi_{d,k}$ *be a Wendland function and* $d \geqslant 3$ *if* $k = 0$*. Then*

$$\mathcal{H}_{\phi_{d,k}}(\mathbb{R}^d) = W^{(d+1)/2+k}(\mathbb{R}^d)$$

with equivalent norms.

We explicitly state Wendland's RBFs used later in this thesis in Tab. 2.2 and visualize them in Fig. 2.1. Note, that $\phi_{1,0}$ equals the truncated power function out of Ex. 2.16 for $l = 1$.

	$C^0,\ k=0$	$C^4,\ k=2$
$d = 1$	$\phi_{1,0}(r) = (1-r)_+$	$\phi_{1,2}(r) = (1-r)_+^5\ (8r^2 + 5r + 1)$
$d \leqslant 3$	$\phi_{3,0}(r) = (1-r)_+^2$	$\phi_{3,2}(r) = (1-r)_+^6\ (35r^2 + 18r + 3)$

Table 2.2: Wendland's radial basis functions

Remark 2.36. We conclude this section with a brief remark on the omnipresent Gaussian kernel (Ex. 2.19). Since it is C^∞, its native space is very small. Indeed, G.E. Fasshauer and Q. Ye showed in [FY11] that its native space is contained in every $W^m \cap C_b^\infty$, $m \in \mathbb{N}$, where C_b^∞ denotes all bounded C^∞ functions. For a detailed analysis of the Gaussian kernels' native space, we refer to [SC08, Chapter 4.4]. There, among other findings, it is demonstrated in Corollary 4.44 that for an Ω with a non-empty interior, the only constant function in the Gaussian's native space is the zero function.

2.4 Interpolation

This section states results on the numerical stability (Section 2.4.2) and accuracy (Section 2.4.1) of the interpolation method introduced in Section 2.1.

In Section 2.4.1 we have to assume that the target function f lies in the native space. In fact, results on the interpolant without prior knowledge on the space of the target function are few. It is still possible to show, that the chosen interpolant has minimal norm under all functions $s \in \mathcal{H}_K(\Omega)$ interpolating the data f_X. We state [Wen05, Theorem 13.2].

Theorem 2.37. *Let K be a symmetric positive definite kernel on Ω and $X \subset \Omega$ be a finite and pairwise distinct point set. Furthermore, let the function values $f_1, \dots, f_N \in \mathbb{R}$ on X be given. Then*

$$\left\|s_{f,K}\right\|_K = \min\left\{\|s\|_K \ : \ s \in \mathcal{H}_K(\Omega) \text{ fulfilling the interpolation condition } (2.1) \right\}$$

2.4.1 Approximation Error

This section is concerned with the approximation error between the target function $f \in \mathcal{H}_{K,\Omega}$ and its interpolant $s_{f,K,X} \in S_{K,X}$ of (2.7). We take a look at the error measured in the native space norm and state an upper bound for the pointwise error depending on the power function (Theorem 2.40). It turns out that the power function can be bounded in terms of the fill distance (Theorem 2.43). This section finishes with a result regarding the convergence of this interpolation method (Theorem 2.45).

We display an orthogonal decomposition of a native space as done in [Isk18, Corollary 8.28] and [Wen05, Lemma 10.24f.]). As the interpolation space $S_{K,X}$ is a closed linear subspace of the Hilbert space $\mathcal{H}_{K,\Omega}$, there exist an orthogonal complement of $S_{K,X}$ in $\mathcal{H}_{K,\Omega}$ by [Mus14, Theorem 10.12]. This orthogonal complement turns out to be the set of functions which are zero on X.

Theorem 2.38. *Let K be a symmetric and positive definite kernel on Ω and $\mathcal{H}_{K,\Omega}$ its native space. For any finite and pairwise distinct point set $X \subset \Omega$, the space $\mathcal{H}_{K,\Omega}$ can be orthogonally decomposed as*

$$\mathcal{H}_{K,\Omega} = S_{K,X} \oplus \left\{f \in \mathcal{H}_{K,\Omega} \ : \ f_X = 0\right\}.$$

For a target function $f \in \mathcal{H}_{K,\Omega}$ and its unique interpolant $s_{f,K,X} \in S_{K,X}$ on X it is

$$\|f\|_K^2 = \left\|s_{f,K,X}\right\|_K^2 + \left\|f - s_{f,K,X}\right\|_K^2.$$

With this theorem we immediately conclude that the interpolant $s_{f,K,X}$ is the best approximation to a target function $f \in \mathcal{H}_{K,\Omega}$ from the interpolation space $S_{K,X}$, as the interpolant $s_{f,K,X}$ is the orthogonal projection of f onto $S_{K,X}$.

Lemma 2.39. *Let the setting of Theorem 2.38 hold. Then*

$$\left\|f - s_{f,K,X}\right\|_K \leqslant \|f - s\|_K \qquad \text{for all } s \in S_{K,X}.$$

Furthermore, it is

$$\left\|f - s_{f,K,X}\right\|_K \leqslant \|f\|_K .$$

The above provides information about the error in the native space norm. As established in Section 2.3, this norm is closely related to the pointwise error due to the reproducing property. We leverage this relationship to derive a pointwise error bound. We interpret $K(x,\cdot)$ as a function on Ω and use its interpolant/orthogonal projection $s_{K(x,\cdot),K,X}$ onto $S_{K,X}$. Let the interpolants of f and $K(x,\cdot)$ be given by

$$s_{f,K,X} = \sum_{i=1}^{N} c_i\, K(x_i,\cdot) \qquad \text{and} \qquad s_{K(x,\cdot),K,X} = \sum_{i=1}^{N} d_i\, K(x_i,\cdot).$$

We compute

$$\begin{aligned} s_{f,K,X}(x) &= \sum_{i=1}^{N} c_i\, K(x_i,x) = \sum_{i=1}^{N}\sum_{j=1}^{N} c_i d_j\, K(x_i,x_j) = \sum_{j=1}^{N} d_j f(x_j) \\ &= \sum_{j=1}^{N} d_j \left\langle f,\, K(x_j,\cdot)\right\rangle_K = \left\langle f,\, s_{K(x,\cdot),K,X}\right\rangle_K \qquad \text{for all } x \in \Omega. \end{aligned}$$

Together with the Cauchy-Schwartz inequality, this implies the following error bound

$$\begin{aligned} \left|f(x) - s_{f,K,X}(x)\right| &= \left|\left\langle f,\, K(x,\cdot) - s_{K(x,\cdot),K,X}\right\rangle_K\right| \\ &\leqslant \|f\|_K \left\|K(x,\cdot) - s_{K(x,\cdot),K,X}\right\|_K \qquad \text{for all } x \in \Omega. \end{aligned}$$

We note that the right factor is independent of the target function f and corresponds to the power function $P_{K,X}$ of Definition 2.27. The above is summarized in the following result, which is also discussed in various literature, such as [Wen05, Theorem 11.4].

Theorem 2.40. *In the setting of Theorem 2.38 the pointwise error can be bounded by*

$$|f(x) - s_{f,K,X}(x)| \leqslant P_{K,X}(x)\, \|f\|_K \qquad \textit{for all } x \in \Omega.$$

This reveals two adjusting tools to improve the pointwise error. The power function on the one hand side, depending on the kernel K and the point set X, and the native space norm of the target function on the other. Regarding the analysis of native space norms we refer to Section 3.2 and concentrate on the power function, for now.

The power function $P_{K,X}(x)$ measures the minimal distance between the function $K(x,\cdot)$ and the interpolation space $S_{K,X}$ in the native space norm. We directly see that $P_{K,X}(x) = 0$ if and only if $x \in X$. One gets the idea, the further away x from X, the larger $P_{K,X}(x)$. To illustrate this, let $X = \{x_1, \ldots, x_N\}$ and K satisfy

$$\left|K(x,y) - K(x,z)\right| \leqslant L/2\, \|y - z\|_2 \qquad \text{for all } x,y,z \in \Omega \tag{2.15}$$

and a constant $L > 0$. This is for example satisfied by a C^1 RBF. Then

$$\begin{aligned} P_{K,X}(x) &= \min_{s \in S_{K,X}} \|K(x,\cdot) - s\|_K \\ &\leqslant \min_{1 \leqslant i \leqslant N} \|K(x,\cdot) - K(x_i,\cdot)\|_K \\ &\leqslant L \min_{1 \leqslant i \leqslant N} \|x - x_i\|_2 \qquad \text{for all } x \in \Omega. \end{aligned}$$

This justifies the definition of the fill distance as the radius of the largest ball in Ω that contains no element of X, as done in [Wen05, Definition 1.4].

Definition 2.41. The *fill distance* of a finite set of points $X = \{x_1, \dots, x_N\} \subseteq \Omega$ for a bounded domain Ω is defined as

$$h_{X,\Omega} := \sup_{x \in \Omega} \min_{1 \leqslant i \leqslant N} \|x - x_i\|_2 .$$

In the definition, we restrict Ω to a bounded domain. Naturally, we could extend the concept of fill distance to non-bounded domains as well. However, in such cases, we would have $h_{X,\Omega} = \infty$, rendering it impractical since our goal is to find an upper bound.

Even though most standard kernels satisfy the Lipschitz condition (2.15), this is not necessarily the case for the nonstandard kernels introduced in the upcoming sections. Still, it is possible to bound the power function from above in dependence of the fill distance. To do so, we introduce the interior cone condition on Ω, as presented in [Wen05, Definition 3.6].

Definition 2.42. The set $\Omega \subseteq \mathbb{R}^d$ fulfills the *interior cone condition* (ICC) if there exist an angle $\theta \in (0, \pi/2)$ and a radius $r > 0$ such that for every $x \in \Omega$ a unit vector $\xi(x)$ exists such that the cone

$$C(x, \xi(x), \theta, r) := \{x + \lambda y \ : \ y \in \mathbb{R}^d \ , \|y\|_2 = 1 \ , \ y^T \xi(x) \geqslant \cos\theta \ , \ \lambda \in [0, r]\} \tag{2.16}$$

is contained in Ω.

We present the general result of [Wen05, Theorem 3.14 and Theorem 11.9] for bounding the power function in terms of the fill distance.

Theorem 2.43. *Let K be a translation-invariant kernel with univariate function $\Phi \in C(\mathbb{R}^d)$, and $\Omega \subset \mathbb{R}^d$ be a bounded set satisfying an ICC for an angle $\theta \in (0, \pi/2)$ and a radius $r > 0$. For a finite pairwise distinct point set $X \subset \Omega$ satisfying $h_{X,\Omega} \leqslant h_0$, the power function can be bounded by*

$$P_{K,X}^2(\tilde{x}) \leqslant c_1 \sup_{x \in B(0,\, 2c_2 h_{X,\Omega})} |\Phi(x) - p(x)|, \qquad \text{for all } \tilde{x} \in \Omega, \tag{2.17}$$

where p is an arbitrary polynomial from $\pi_m(\mathbb{R}^d)$, $m \in \mathbb{N}$ and

$$c_1 = 9, \quad c_2 = \frac{16\,(1 + \sin\theta)^2\ m^2}{3\,\sin^2\theta}, \quad h_0 = \frac{r}{c_2}.$$

Remark 2.44. The right hand side of (2.17) can be bounded by a function $F_{\Phi,\Omega}$ depending on (the smoothness of) Φ and the interior cone condition of Ω, and acting on the fill distance $h_{X,\Omega}$. Hence, for every translation-invariant kernel with univariate function $\Phi \in C(\mathbb{R}^d)$ and bounded set $\Omega \subset \mathbb{R}^d$ satisfying an ICC, there exist an increasing function $F_{\Phi,\Omega} : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$, so that

$$P_{K,X}^2(x) \leqslant F_{\Phi,\Omega}(h_{X,\Omega}).$$

	$\phi(r)$	$F(h)$
Gaussian, Ex. 2.19	$\mathrm{e}^{-\alpha r^2}, \ \alpha > 0$	$e^{-c\frac{\lvert \log h \rvert}{h}}, \ c > 0$
Matérn, Tab. 2.1	$\psi_\ell(r), \ 2m = d + 2\ell + 1$	h^{2m-d}
Wendland, Tab. 2.2	$\phi_{d,k}(r)$	h^{2k+1}

Table 2.3: Upper bounds on the power function $P_{K,X}^2$ in terms of the fill distance $h_{X,\Omega}$

We provide upper bounds on $P_{K,X}^2$, presented as functions F on the fill distance $h_{X,\Omega}$ in Tab. 2.3. See [Wen05, Chapter 11], [Fas07, Chapter 15.1], and [Sch95] for the corresponding proofs. In Fig. 2.2 (right), it is visible that an increase of smoothness of the Wendland and Matérn kernels improves the bound.

We complete this section with a general statement regarding the convergence of the proposed interpolation method. A sequence of interpolants $(s_n)_{n\in\mathbb{N}}$ converges to the target function $f \in \mathcal{H}_{K,\Omega}$ if the corresponding point sets $(X_n)_{n\in\mathbb{N}}$ fill up the whole domain Ω, see [Isk18, Theorem 8.37].

Theorem 2.45. *Let K be a continuous symmetric positive definite kernel on a bounded set $\Omega \subset \mathbb{R}^d$ and $(X_n)_{n\in\mathbb{N}}$ a sequence of finite pairwise distinct point sets in Ω such that*

$$h_{X_n,\Omega} \searrow 0 \qquad \text{for } n \to \infty.$$

For a target function $f \in \mathcal{H}_{K,\Omega}$ and its unique interpolants $\left(s_{f,X_n}\right)_{n\in\mathbb{N}}$ it is

$$\left\| f - s_{f,X_n} \right\|_K \longrightarrow 0 \qquad \text{for } n \to \infty.$$

2.4.2 Numerical Stability

We are looking into the numerical stability of the interpolation problem stated in Section 2.1. A system is called stable, if an error on the input data does not affect the result too much. The condition number, which is a measure of stability, gives reason to examine the maximal and minimal eigenvalue of $\mathbf{A}_{K,X}$. In Theorem 2.48 we present a lower bound for the minimal eigenvalue and apply the theorem to our example kernels.

Condition Number Let us for now consider a general linear system

$$Ax = b \tag{2.18}$$

for a matrix $A \in \mathbb{R}^{N \times M}$ and a second system with an error Δb in b

$$A(x + \Delta x) = b + \Delta b,$$

where Δx describes the resulting error in x. Let $\sigma_{\max}$ and $\sigma_{\min}$ be the largest and smallest singular values of A. We immediately see, that

$$\|b\|_2 \leqslant \sigma_{\max} \|x\|_2 \qquad \text{and} \qquad \|\Delta b\|_2 \geqslant \sigma_{\min} \|\Delta x\|_2 \,,$$

as $A(\Delta x) = \Delta b$ by linearity. Let $\sigma_{\min} \neq 0$, then

$$\frac{\|\Delta x\|_2}{\|x\|_2} \leqslant \frac{\sigma_{\max}}{\sigma_{\min}} \frac{\|\Delta b\|_2}{\|b\|_2},$$

where the quantity $\|\Delta b\|_2 / \|b\|_2$ denotes the relative change in the right-hand side of (2.18) and $\|\Delta x\|_2 / \|x\|_2$ is the resulting relative change in the solution. We see, that a change in the right-hand side of (2.18) can cause a change in the solution $\sigma_{\max}/\sigma_{\min}$ times as large. This relative error factor is called the condition number of A.

Definition 2.46. Let $A \in \mathbb{R}^{N \times M}$, then its spectral *condition number* for inversion is given by

$$\operatorname{cond}_2(A) := \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)},$$

where $\sigma_{\max}(A)$ and $\sigma_{\min}(A)$ denote the highest and smallest singular value of A. The problem of (2.18) is referred to as *ill-conditioned* if the condition number is high and hence a small error in the input data causes a big change in the solution, and *well-conditioned* if not.

Let us return to the interpolation problem. As we use symmetric and positive definite kernels K, the interpolation matrix $\mathbf{A}_{K,X}$ is symmetric and positive definite for every finite pairwise distinct point set X. Hence, the singular values of $\mathbf{A}_{K,X}$ are equal to its eigenvalues and its condition number is given by

$$\operatorname{cond}_2\left(\mathbf{A}_{K,X}\right) = \frac{\lambda_{\max}\left(\mathbf{A}_{K,X}\right)}{\lambda_{\min}\left(\mathbf{A}_{K,X}\right)},$$

where $\lambda_{\max}\left(\mathbf{A}_{K,X}\right)$ and $\lambda_{\min}\left(\mathbf{A}_{K,X}\right)$ denote the largest and smallest eigenvalue of $\mathbf{A}_{K,X}$. This gives a first reason to study minimal and maximal eigenvalues of interpolation matrices.

Maximal Eigenvalue To find an upper bound on the maximal eigenvalue of the interpolation matrix, we apply Gershgorin's circle theorem onto the interpolation matrix $\mathbf{A}_{K,X}$. It implies that for every index $j = 1, \dots, N$, the following inequality holds

$$|\lambda_{\max} - K(x_j, x_j)| \leqslant \sum_{i=1,\, i\neq j}^{N} K(x_i, x_j).$$

Hence,

$$\lambda_{\max} \leqslant N \, \|K(\cdot,\cdot)\|_{L^\infty(X\times X)}$$

and if K was translation-invariant it is

$$\lambda_{\max} \leqslant N \; \Phi(0)$$

by Lemma 2.13. Even though N can grow very fast for high dimensions, numerical tests show that the maximum eigenvalue causes no problems.

Minimal Eigenvalue Let $\{x_1, \dots, x_N\} = X \subseteq \Omega$ be a finite pairwise distinct point set and the kernel K be continuous, symmetric and positive definite. For the vector $c = (1, -1, 0, \dots, 0)^T \in \mathbb{R}^N$ it is

$$\begin{aligned}
\sqrt{2}\,\lambda_{\min}(\mathbf{A}_{K,X}) &= \lambda_{\min}(\mathbf{A}_{K,X})\|c\|_2^2 \\
&\leqslant c^T \mathbf{A}_{K,X} c \\
&= c_1^2 \big(K(x_1, x_1) + K(x_2, x_2)\big) - 2c_1^2 K(x_1, x_2) \\
&= K(x_1, x_1) + K(x_2, x_2) - 2\, K(x_1, x_2).
\end{aligned}$$

If $x_2 \longrightarrow x_1$, then $K(x_1, x_2), K(x_2, x_2) \longrightarrow K(x_1, x_1)$ since K is continuous and the right-hand side tends to zero. This shows that if X contains points that lie close together the minimal eigenvalue is small. We introduce the separation distance, as done in [Wen05, Definition 4.6], to measure the minimal distance between points in X.

Definition 2.47. The *separation distance* of a finite point set $X = \{x_1, \dots, x_N\}$ is defined by

$$q_X := \frac{1}{2} \min_{i\neq j} \big\|x_i - x_j\big\|_2 .$$

We note, that the separation distance q_X can be seen as the maximal radius $r > 0$ such that all balls $B_r(x_i)$ for $i = 1, \dots, N$ are disjoint.

Next, we state the general result of [Sch95, Theorem 3.1], which provides a lower bound on the minimal eigenvalue of $\mathbf{A}_{K,X}$ in dependence of the separation distance. Thereafter, in Tab. 2.4, we present bounds for the introduced specific positive definite kernels.

Theorem 2.48. *Let K be a translation-invariant kernel with univariate function Φ such that Φ possesses a positive Fourier transform $\widehat{\Phi} \in C(\mathbb{R}^d \backslash \{0\})$. With the function*

$$\varphi(M) := \inf_{\|\omega\|_2 \leqslant 2M} \widehat{\Phi}(\omega)$$

a lower bound on $\lambda_{\min}(\mathbf{A}_{K,X})$ is given by

$$\lambda_{\min}(\mathbf{A}_{K,X}) \geqslant \frac{\varphi(M)}{2\,\Gamma(d/2+1)} \left(\frac{M}{2^{3/2}}\right)^d$$

for any $M > 0$ satisfying

$$M \geqslant \frac{12}{q_X} \left(\frac{\pi\,\Gamma^2(d/2+1)}{9}\right)^{1/(d+1)}.$$

Remark 2.49. We see that for every K satisfying the requirements of Theorem 2.48, there is a monotonously increasing function G_Φ, depending on the separation distance q_X, that bounds the minimal eigenvalue of the interpolation matrix $\mathbf{A}_{K,X}$ from below. This function G_Φ is, up to a constant depending on the dimension, listed in Tab. 2.4 for our example kernels, and visualized in Fig. 2.2 (left). For proofs, we refer to [Wen05, Corollary 12.4 and 12.8] and [Sch95, Section 3].

	$\phi(r)$	$G(q)$
Gaussian, Ex. 2.19	$\mathrm{e}^{-\alpha r^2},\ \alpha > 0$	$(2\alpha)^{-d/2}\,\mathrm{e}^{-\frac{40.71\,d^2}{\alpha\,q^2}}\,q^{-d}$
Matérn, Tab. 2.1	$\psi_\ell(r),\ 2m = d + 2\ell + 1$	q^{2m-d}
Wendland, Tab. 2.2	$\phi_{d,k}(r)$	q^{2k+1}

Table 2.4: Lower bounds on $\lambda_{\min}$ in terms of the separation distance q_X

In Fig. 2.2 (left), we observe that the function G for the Gaussian, with $\alpha = 2$, is vanishingly small. Hence, we can expect the interpolation problem to be ill-conditioned. With regards to the Matérn and Wendland kernel, we see that increasing smoothness deteriorates the bound.

2.4.3 Trade-Off Principle

Let us condense the main statements out of Section 2.4.2 and Section 2.4.1:

1. The bigger the smallest eigenvalue of the interpolation matrix, the more stable the interpolation method.

2. The smaller the power function, the better the approximation error.

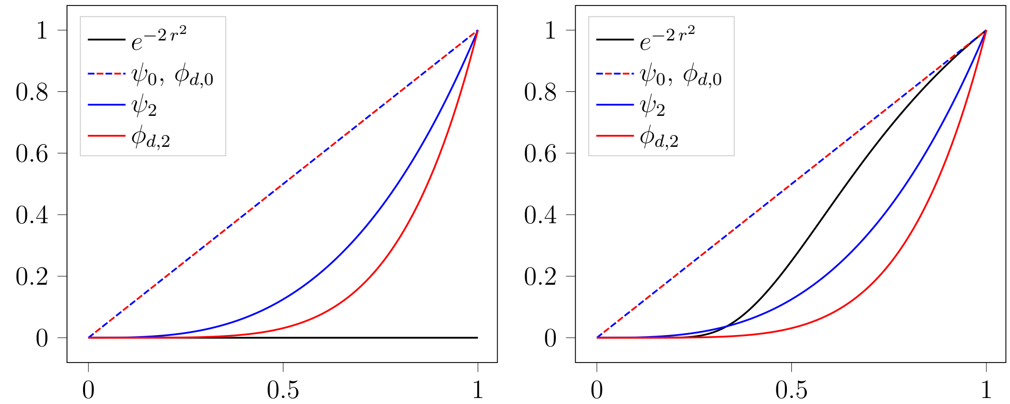

Figure 2.2: Left: Visualization of the function $G(q_X)$ given in Remark 2.49 providing a lower bound on $\lambda_{\min}$ of $\mathbf{A}_{K,X}$. Right: Visualization of the function $F(h_{\Omega,X})$ given in Remark 2.44 providing an upper bound of $P^2_{K,X}$. G and F are presented for the Gaussian with $\alpha = 2$ (black, Ex. 2.19), Matérn's ψ_ℓ (blue, Tab. 2.1), and Wendland's $\phi_{d,\ell}$ (red, Tab. 2.2) for $\ell = 0, 2$.

We presented a lower bound for the smallest eigenvalue regarding our example kernels in terms of the function G depending on the separation distance q_X, and an upper bound on the squared power function in terms of a function F depending on the fill distance $h_{\Omega,X}$. For the Matérn and Wendland kernels the functions G and F coincide (up to an X-independent constant), see Fig. 2.2. In case of $q_X \leqslant h_{\Omega,X}$ we have to decide between good stability and small approximation error, as we cannot combine it.

This is not only the case for Matérn and Wendland kernels. We derive a relation between the smallest eigenvalue and the power function in a more general setting. Let K be a symmetric and positive definite kernel on $\Omega \subseteq \mathbb{R}^d$ and $\{x_1, \ldots, x_N\} = X \subset \Omega$ be a pairwise distinct point set. Furthermore, let w.l.o.g. the minimum of the power function be taken at x_1, i.e.,

$$P^2_{K,X\setminus\{x_1\}}(x_1) = \min_{1\leqslant i\leqslant N} P^2_{K,X\setminus\{x_i\}}(x_i).$$

Let the interpolant $s_{K(x_1,\cdot),K,X\setminus\{x_1\}}$ have the form

$$s_{K(x_1,\cdot),K,X\setminus\{x_1\}} = \sum_{i=2}^{N} c_i\ K(x_i,\cdot)$$

for $c = (-1, c_2, \ldots, c_N)^T \in \mathbb{R}^N$. Then

$$\begin{aligned} \min_{1\leqslant i\leqslant N} P^2_{K,X\setminus\{x_i\}}(x_i) &= P^2_{K,X\setminus\{x_1\}}(x_1) = \left\| s_{K(x_1,\cdot),X\setminus\{x_1\}} - K(x_1,\cdot) \right\|^2 \\ &= \left\| \sum_{i=1}^{N} c_i\ K(x_i,\cdot) \right\|^2 \geqslant \|c\|_2^2\ \lambda_{\min}(\mathbf{A}_{K,X}) \geqslant \lambda_{\min}(\mathbf{A}_{K,X}). \end{aligned} \tag{2.19}$$

This inequality, together with the statements 1 and 2, demonstrates that it is impossible to achieve both arbitrary high accuracy and good stability simultaneously. This dilemma is referred to as the *trade-off principle* (cf. [Wen05, Chapter 12.1]) or Schaback's uncertainty relation (cf. [Sch95]). We address this issue with the approach of problem-adapted non-standard kernels.

Part II

Combinations of Kernels

Chapter 3

Summation Kernels

Understanding orthogonal decompositions and subsets of Hilbert spaces is crucial for several reasons. These decompositions provide a foundational framework for analyzing functions and operators within a Hilbert space. They enable a clearer understanding of the geometric and algebraic properties of these spaces, which is essential for various applications in functional analysis and numerical methods. Moreover, the study of subsets within Hilbert spaces often leads to significant insights into the structure and behavior of functions.

The theory of reproducing kernel Hilbert spaces (RKHS) has been extensively developed with the work [Aro50] from N. Aronszajn in 1950. His contribution to orthogonal decompositions and subsets of RKHSs, as well as on sums and differences of corresponding reproducing kernels laid the groundwork for understanding the nature of RKHS and their norms. While examining *summation kernels*, N. Aronszajn established an order on the space of reproducing kernels. This order elucidates the subset and norm relationships within RKHS, forming a basis for further research such as [Ylv62], [Sch64], and [Dri73]. However, none of these studies explores summation kernels in the context of interpolation, a gap that this chapter aims to address.

In the following, we build on N. Aronszajn's foundational work and extend the theory of summation kernels by focusing on their application in interpolation. The main contributions are:

- Connecting Aronszajn's Findings to Sobolev Spaces: In Section 3.4, we establish a connection between N. Aronszajn's results and Sobolev spaces. This helps in understanding the relationship between Matérn and Wendland kernels (Corollary 3.24 and Corollary 3.18), which are frequently used.

- Examination of Native Space Intersections: We investigate the intersection of two native spaces (Section 3.2.3).

- Linking Mercer's Theorem to Summation Kernels: In Section 3.2.4, we draw a connection between Mercer's theorem and countably infinite summation kernels, offering a deeper theoretical understanding of these kernels.

- Influence of Kernel Equivalence Classes on Interpolants: We explore how the equivalence class of a kernel affects the resulting interpolant, which is crucial for practical interpolations (Section 3.3) and find a trade-off for kernels schematically visualized in Remark 3.42.

- Detailed Analysis of Interpolation with Summation Kernel: We explore the approximation error (Section 3.3.1) and numerical stability (Section 3.3.2) of interpolation using a summation kernel, providing both theoretical analysis and numerical results (Section 3.4).

By addressing these topics, this chapter not only fills a significant gap in the existing literature but also provides a comprehensive framework for understanding and using summation kernels in interpolation problems.

This chapter is structured as follows. We define summation kernels and state initial findings in Section 3.1. Section 3.2 explores the native spaces of summation kernels, extending results of [Aro50]. We provide fundamentals of summing up spaces in Section 3.2.1. As the topic of summation kernels is closely related to the topic of subsets of native spaces, we devote Section 3.2.2 to this. In Section 3.2.3, we show that the intersection of two native spaces is a RKHS and compare norms. Section 3.2.4 links countable summation kernels to Mercer's theorem, demonstrating a decomposition into orthogonal RKHSs. Section 3.3 examines interpolation with summation kernels, emphasizing error minimization (Section 3.3.1) and stability (Section 3.3.2). In Section 3.4, we provide supporting results for the theoretical analysis.

3.1 Definition and Basic Properties

In the following, we provide a precise definition of summation kernels, visualized in Fig. 3.1, and state basic findings regarding positive definiteness (Theorem 3.2), translation-invariance and radial symmetry (Lemma 3.3).

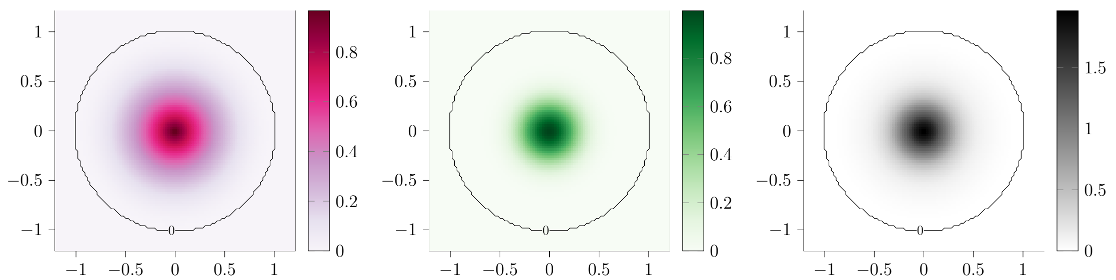

Figure 3.1: Top view of two two-dimensional component kernels resulting from the Wendland RBFs $\phi_{3,0}$ (left) and $\phi_{3,3}$ (middle), and the corresponding two-dimensional summation kernel (right). The black line visualizes supports.

From now on, we solely consider functions and function spaces on the domain $\Omega \subseteq \mathbb{R}^d$. Hence, we shorten the notation for native spaces $\mathcal{H}_{K,\Omega}$ to $\mathcal{H}_K$ equipped with the inner product $\langle \cdot, \cdot \rangle_K$ and the corresponding norm $\|\cdot\|_K$ defined in Section 2.3.2.

Definition 3.1. Let $K_\ell : \Omega \times \Omega \longrightarrow \mathbb{R}$ for $\ell = 1, \dots, M$, then

$$\begin{aligned} K : \Omega \times \Omega &\longrightarrow \mathbb{R}, \\ K(x, y) &= \textstyle\sum_{\ell=1}^{M} K_\ell(x, y) \text{ for } x, y \in \Omega \end{aligned}$$

is called a *summation kernel.*

Section 2.3 gives reason to work with positive (semi-)definite kernels to ensure the existence of a native space and unique solutions to the interpolation problem given in Section 2.1. The observation that the interpolation matrix of the summation kernel K equals the sum of the components' interpolation matrices, i.e.,

$$\mathbf{A}_{K,X} = \sum_{\ell=1}^{M} \mathbf{A}_{K_\ell,X}, \tag{3.1}$$

facilitates the subsequent statement regarding positive (semi-)definiteness.

Theorem 3.2. *The summation kernel $K : \Omega \times \Omega \longrightarrow \mathbb{R}$ of positive semi-definite component kernels $K_\ell : \Omega \times \Omega \longrightarrow \mathbb{R}$ for $\ell = 1, \dots, M$, is again positive semi-definite. If at least one component kernel K_ℓ is positive definite, so is the summation kernel K.*

Proof. Let $X \subset \Omega$ a finite and pairwise distinct set of data points. Since (3.1) holds and each component kernel is positive semi-definite, it is

$$c^T \mathbf{A}_{K,X}\, c = \sum_{\ell=1}^{M} c^T \mathbf{A}_{K_\ell,X}\, c \geqslant 0 \qquad \text{for all } c \in \mathbb{R}^{|X|}.$$

■

We note that the set of positive semi-definite kernels equipped with the sum can be viewed as a commutative semigroup, with $K(x, y) = 0$ for all x, y as the neutral element.

Lemma 3.3. *Let K_ℓ be kernels on Ω for $\ell = 1, \dots, M$ and K be their summation kernel.*

(i) If K_ℓ is translation-invariant with uni-variate function Φ_ℓ for $\ell = 1, \dots, M$, their summation kernel K is translation-invariant with the uni-variate function $\Phi = \sum_{\ell=1}^{M} \Phi_\ell$.

(ii) If K_ℓ is radial with RBF ϕ_ℓ for $\ell = 1, \dots, M$, their summation kernel K is a radial kernel with RBF $\phi = \sum_{\ell=1}^{M} \phi_\ell$.

Proof. The subsequent equations provide the required statements:

(i) $K(x, y) = \sum_{\ell=1}^{M} K_\ell(x, y) = \sum_{\ell=1}^{M} \Phi_\ell(x - y)$

(ii) $K(x, y) = \sum_{\ell=1}^{M} K_\ell(x, y) = \sum_{\ell=1}^{M} \phi_\ell(\|x - y\|_2)$.

■

3.2 Native Spaces

Subsequently, we study the native space of summation kernels. In Section 3.2.1 we first develop an understanding of the sum of sets in general. With foundational concepts of summing up spaces at hand, we examine the native spaces structure of summation kernels. A particular focus is placed on orthonormal sums.

As the topic of summation kernels is closely related to the topic of subsets of native spaces, we devote Section 3.2.2 to this. Here, we draw a connection to Sobolev spaces, elucidating the relationship between Matérn and Wendland kernels. Additionally, we establish the order '$\lesssim$' on the space of native spaces (and equivalently on the space of symmetric positive definite kernels), as this order provides insights into the relationship between component kernels and summation kernels.

In Section 3.2.3, we explore a particular subspace: the intersection of two native spaces. We show that the intersection is also an RKHS and derive results to compare native space norms of a function f within the intersection.

Moreover, we establish a link between countable summation kernels and the famous Mercer theorem out of [Mer09] in Section 3.2.4. We briefly analyze infinitely countable summation kernels, demonstrating that the decomposition of a kernel promoted by Mercer's theorem can be viewed as a countable summation kernel. Indeed, the native space of a reproducing kernel is decomposed into infinitely countable orthogonal RKHSs via Mercer's theorem.

3.2.1 Fundamental Concepts

In order to settle on consistent definitions and develop an understanding of set combinations in general, we provide basic definitions of the Minkowski sum (Definition 3.4) and the orthogonal direct sum (Definition 3.7), first. Equipped with these foundational concepts we examine the native spaces structure of a summation kernel in Theorem 3.9, based on [Aro50]. Since we are solely working with kernels acting on Ω, we omit Ω in the subscript of native spaces and its inner products. It turns out that trivial intersections, i.e.,

$$\mathcal{H}_{K_1} \cap \mathcal{H}_{K_2} = \{0\}, \tag{3.2}$$

of the components' native spaces are of special interest as in this case they form an orthogonal decomposition of the summation kernels native space. We emphasize the benefits of orthogonal sums and derive six equivalence statements of (3.2) in Theorem 3.12.

We begin with the fundamental concept of the Minkowski sum.

Definition 3.4. Let $A_\ell \subseteq V$ for $\ell = 1, \ldots, M$ be subsets of a vector space V then

$$A := \sum_{\ell=1}^{M} A_\ell = \left\{ \sum_{\ell=1}^{M} a_\ell \;:\; a_\ell \in A_\ell \text{ for all } \ell = 1, \ldots, M \right\} \subseteq V$$

is called the *Minkowski sum* of $\{A_\ell\}_{\ell=1}^{M}$.

In Definition 3.4 a unique decomposition of an element in A into elements of A_ℓ for $\ell = 1, \dots, M$ is not ensured. To gain a unique decomposition, the subspaces A_ℓ need to be complemented (cf. [UA22]).

Definition 3.5. Let A_ℓ for $\ell = 1, \dots, M$ be *complemented* subspaces of a vector space V, i.e.,

$$A_m \cap \sum_{\ell=1, \ell \neq m}^{M} A_\ell = \{0\} \qquad \text{for all } m = 1, \dots, M.$$

Then, the *internal direct sum* A of $\{A_\ell\}_{\ell=1}^M$ is given by

$$A := \bigoplus_{\ell=1}^{M} A_\ell = \left\{ \sum_{\ell=1}^{M} a_\ell \ : \ a_\ell \in A_\ell \text{ for all } \ell = 1, \dots, M \right\}.$$

Lemma 3.6. *Let A be the internal direct sum of the vector spaces A_ℓ for $\ell = 1, \dots, M$. Then, every element $a \in A$ has a unique representation*

$$a = \sum_{\ell=1}^{M} a_\ell,$$

where $a_\ell \in A_\ell$ for all $\ell = 1, \dots, M$.

Proof. Let $a \in A$ have the representations $\sum_{\ell=1}^M a_\ell$ and $\sum_{\ell=1}^M a'_\ell$, where $a_\ell, a'_\ell \in A_\ell$ for $\ell = 1, \dots, M$. Then,

$$a_m - a'_m = \sum_{\ell=1, \ell \neq m}^{M} a'_\ell - \sum_{\ell=1, \ell \neq m}^{M} a_\ell \in A_m \cap \sum_{\ell=1, \ell \neq m}^{M} A_\ell = \{0\} \qquad \text{for all } m = 1, \dots, M.$$

This implies $a_m = a'_m$ for all $m = 1, \dots, M$, thereby ensuring the uniqueness of the representation of a.

■

Native spaces of positive semi-definite kernels are Hilbert spaces, as demonstrated in Section 2.3. Hence, we shift our focus to the special case of spaces with an inner product. Here, the internal direct sum can be characterized in the following way.

Definition 3.7. Let H_ℓ for $\ell = 1, \dots, M$ be pairwise orthogonal subspaces of a Hilbert space with inner product $\langle \cdot, \cdot \rangle_H$. In other words, from $\ell \neq m$ follows $H_\ell \perp H_m$, i.e.,

$$\langle f_\ell, f_m \rangle_H = 0 \qquad \text{for all } f_\ell \in H_\ell, f_m \in H_m.$$

Then, the *orthogonal direct sum* H of H_ℓ for $\ell = 1, \dots, M$ is given by

$$H := \bigoplus_{\ell=1}^{M} H_\ell = \left\{ \sum_{\ell=1}^{M} f_\ell \ : \ f_\ell \in H_\ell \text{ for all } \ell = 1, \dots, M \right\}$$

equipped with the inner product

$$\langle f, g\rangle_H = \sum_{\ell=1}^{M} \langle f_\ell, g_\ell\rangle_{H_\ell}, \tag{3.3}$$

where $f = \sum_{\ell=1}^{M} f_\ell$, $g = \sum_{\ell=1}^{M} g_\ell$ with $f_\ell, g_\ell \in H_\ell$ for all $\ell = 1, \ldots, M$.

Remark 3.8. In the context of Definition 3.7, it can be demonstrated that the requirement for H_ℓ to be pairwise orthogonal is equivalent to requiring that H_ℓ are complemented.

Having gained an understanding of summing sets with differing structures, we redirect our focus to summation kernels and their corresponding native spaces. N. Aronszajn analyzed such spaces as early as the 1950s. Subsequently, we repeat [Aro50, Part I, §6] and extend the statement by taking more than two component kernels into account.

Theorem 3.9. *Let K_ℓ be positive semi-definite kernels on Ω for $\ell = 1, \ldots, M$ and $K = \sum_{\ell=1}^{M} K_\ell$ their summation kernel. Then, the native space $\mathcal{H}_K$ of the summation kernel equals the Minkowski sum of its components' native spaces, i.e.,*

$$\mathcal{H}_K = \sum_{\ell=1}^{M} \mathcal{H}_{K_\ell} = \left\{ \sum_{\ell=1}^{M} f_\ell \;:\; f_\ell \in \mathcal{H}_{K_\ell} \text{ for } \ell = 1, \ldots, M \right\}.$$

The native space $\mathcal{H}_K$ is equipped with the norm

$$\|f\|_K^2 = \min \left\{ \sum_{\ell=1}^{M} \|f_\ell\|_{K_\ell}^2 \right\}, \tag{3.4}$$

where the minimum is taken over all decompositions $f = \sum_{\ell=1}^{M} f_\ell$ with $f_\ell \in \mathcal{H}_{K_\ell}$ for $\ell = 1, \ldots, M$.

Additionally, if the spaces $\mathcal{H}_{K_\ell}$ for $\ell = 1, \ldots, M$ are complemented, the Hilbert space $\mathcal{H}_K$ equals the orthogonal direct sum, i.e.,

$$\mathcal{H}_K = \bigoplus_{\ell=1}^{M} \mathcal{H}_{K_\ell},$$

and

$$\langle f, g\rangle_K = \sum_{\ell=1}^{M} \langle f_\ell, g_\ell\rangle_{K_\ell},$$

where $f = \sum_{\ell=1}^{M} f_\ell$ and $g = \sum_{\ell=1}^{M} g_\ell$, with $f_\ell, g_\ell \in \mathcal{H}_{K_\ell}$ for $\ell = 1, \ldots, M$, denote the unique decomposition of f and g.

Proof. The proof is carried out trough induction on M.
$\underline{M=2}$: The base case is divided into tree steps. We introduce the outer sum H of $\mathcal{H}_{K_1}$ and $\mathcal{H}_{K_2}$, and provide an isomorphism between a special subset of H and the Minkowski sum $\mathcal{H}_{K_1}+\mathcal{H}_{K_2}$ in (i). The second step (ii) shows that the summation kernel K is the reproducing kernel of $\mathcal{H}_{K_1}+\mathcal{H}_{K_2}$. At last, step (iii) provides the norm statement.

(i) We define H as the Hilbert space of the outer sum regarding $\mathcal{H}_{K_1}$ and $\mathcal{H}_{K_2}$, i.e.,

$$H := \mathcal{H}_{K_1} \times \mathcal{H}_{K_2} = \{(f_1, f_2) \ : \ f_i \in \mathcal{H}_{K_\ell} \text{ for } \ell = 1, 2\},$$

where the inner product is given by

$$\begin{aligned} \big\langle (f_1,f_2),(g_1,g_2)\big\rangle_H &= \langle f_1, g_1\rangle_{K_1} + \langle f_2, g_2\rangle_{K_2} \text{ and} && (3.5)\\ (f_1+g_1, f_2+g_2) &= (f_1,f_2)+(g_1,g_2) \qquad \text{for } f_\ell, g_\ell \in \mathcal{H}_{K_\ell}, \ell=1,2. && (3.6)\end{aligned}$$

Let $\mathcal{H}_0 := \mathcal{H}_{K_1} \cap \mathcal{H}_{K_2}$ be the intersection of both component's native spaces, which is reduced to $\{0\}$ in the additional assumption, then

$$H_0 := \big\{(f,-f) \ : \ f \in \mathcal{H}_0\big\} \subseteq H.$$

Let us additionally define a mapping φ onto the Minkowski sum $\mathcal{H}_{K_1}+\mathcal{H}_{K_2}$

$$\varphi : H \longrightarrow \mathcal{H}_{K_1}+\mathcal{H}_{K_2}, \qquad (f_1,f_2) \longmapsto f_1+f_2 = f.$$

As the equation (3.6) holds, the mapping φ is linear. On the one hand, it follows that $H_0 = \ker\varphi$ and hence it is a closed subspace of H. Then the space H can be written as the orthogonal direct sum $H = H_0 \oplus H_0^\perp$, where $H_0^\perp \subseteq H$ denotes the orthogonal complement of H_0 in H. On the other hand, we get that φ restricted to $H_0^\perp$ is a bijective mapping. We denote

$$\varphi^{-1}(f) = (f_1', f_2') \in H_0^\perp \qquad \text{for } f \in \mathcal{H}_{K_1}+\mathcal{H}_{K_2}, \tag{3.7}$$

and define the inner product on $\mathcal{H}_{K_1}+\mathcal{H}_{K_2}$ by

$$\langle f, g\rangle_{\mathcal{H}_{K_1}+\mathcal{H}_{K_2}} := \big\langle (f_1',f_2'),(g_1',g_2')\big\rangle_H = \big\langle f_1', g_1'\big\rangle_{K_1} + \big\langle f_2', g_2'\big\rangle_{K_2} \tag{3.8}$$

for $f, g \in \mathcal{H}_{K_1}+\mathcal{H}_{K_2}$, where the last equation is due to (3.5). This makes φ restricted to $H_0^\perp$ an isometric isomorphism.

With the additional assumption $\mathcal{H}_{K_1} \cap \mathcal{H}_{K_2} = \{0\}$, it is $H = H_0^\perp$. This makes φ an isomorphism between H and $\mathcal{H}_{K_1}+\mathcal{H}_{K_2}$. Therefore, (3.8) yields

$$\langle f, g\rangle_{\mathcal{H}_{K_1}+\mathcal{H}_{K_2}} = \langle f_1, g_1\rangle_{K_1} + \langle f_2, g_2\rangle_{K_2}.$$

(ii) We show that the summation kernel, $K(x,y) = K_1(x,y)+K_2(x,y)$ for all $x,y \in \Omega$, is the reproducing kernel of $\mathcal{H}_{K_1} + \mathcal{H}_{K_2}$. Clearly it is

$$K(x,\cdot) = K_1(x,\cdot) + K_2(x,\cdot) \in \mathcal{H}_{K_1} + \mathcal{H}_{K_2} \qquad \text{for all } x \in \Omega.$$

We denote $\varphi^{-1}(K(x,\cdot)) = \big(\kappa_1(x,\cdot), \kappa_2(x,\cdot)\big) \in H_0^\perp$. Since

$$K_1(x,\cdot) + K_2(x,\cdot) = K(x,\cdot) = \kappa_1(x,\cdot) + \kappa_2(x,\cdot),$$

it follows

$$K_1(x,\cdot) - \kappa_1(x,\cdot) = -\big(K_2(x,\cdot) - \kappa_2(x,\cdot)\big)$$

and therefore

$$\big(K_1(x,\cdot) - \kappa_1(x,\cdot),\ K_2(x,\cdot) - \kappa_2(x,\cdot)\big) \in H_0.$$

Let $f \in \mathcal{H}_{K_1} + \mathcal{H}_{K_2}$ then the above yields

$$\begin{aligned} f(x) &\overset{(3.7)}{=} f_1'(x) + f_2'(x) \\ &= \big\langle f_1', K_1(x,\cdot)\big\rangle_{K_1} + \big\langle f_2', K_2(x,\cdot)\big\rangle_{K_2} \\ &\overset{(3.5)}{=} \big\langle (f_1', f_2'), (K_1(x,\cdot), K_2(x,\cdot))\big\rangle_H \\ &= \big\langle (f_1', f_2'), (\kappa_1(x,\cdot), \kappa_2(x,\cdot))\big\rangle_H \\ &\qquad + \big\langle (f_1', f_2'), (K_1(x,\cdot) - \kappa_1(x,\cdot), K_2(x,\cdot) - \kappa_2(x,\cdot))\big\rangle_H \\ &\overset{(3.8)}{=} \big\langle f, K(x,\cdot)\big\rangle_{\mathcal{H}_{K_1}+\mathcal{H}_{K_2}}, \end{aligned}$$

since $(f_1', f_2') \in H_0^\perp$ and $(K_1(x,\cdot) - \kappa_1(x,\cdot), K_2(x,\cdot) - \kappa_2(x,\cdot)) \in H_0$. As a consequence, K is the reproducing kernel of $\mathcal{H}_{K_1} + \mathcal{H}_{K_2}$. By Theorem 2.22, the Hilbert space $\mathcal{H}_{K_1} + \mathcal{H}_{K_2}$ equals the native space $\mathcal{H}_K$ and the inner products are the same, i.e.,

$$\langle \cdot, \cdot \rangle_K = \langle \cdot, \cdot \rangle_{\mathcal{H}_{K_1}+\mathcal{H}_{K_2}}.$$

Taking the additional assumption into account, we obtain $\mathcal{H}_K = \mathcal{H}_{K_1} \oplus \mathcal{H}_{K_2}$ by Remark 3.8.

(iii) We intend a characterization of the norm $\|\cdot\|_{\mathcal{H}_K}$ in $\mathcal{H}_K$ without using the auxiliary space H. Let $f = f_1 + f_2 \in \mathcal{H}_{K_1} + \mathcal{H}_{K_2}$, then

$$\begin{aligned} \|f_1\|_{K_1}^2 + \|f_2\|_{K_2}^2 &= \langle f_1, f_1 \rangle_{K_1} + \langle f_2, f_2 \rangle_{K_2} \\ &\overset{(3.5)}{=} \big\langle (f_1, f_2), (f_1, f_2)\big\rangle_H \\ &= \big\langle (f_1', f_2'), (f_1', f_2')\big\rangle_H \\ &\qquad + \big\langle (f_1 - f_1', f_2 - f_2'), (f_1 - f_1', f_2 - f_2')\big\rangle_H \\ &\overset{(3.8)}{=} \|f\|_K^2 + \big\|f_1 - f_1' + f_2 - f_2'\big\|_K^2. \end{aligned}$$

This shows

$$\|f\|_K^2 = \min\Big\{\|f_1\|_{K_1}^2 + \|f_2\|_{K_2}^2\Big\},$$

where the minimum is taken over all decompositions $f = f_1 + f_2$, where $f_\ell \in \mathcal{H}_{K_\ell}$ for $\ell = 1,2$.

$\underline{M \to M+1}$: Let $\tilde{K} = \sum_{\ell=1}^{M} K_\ell$ and $K = \sum_{\ell=1}^{M+1} K_\ell$. Due to the induction hypothesis, it is

$$\mathcal{H}_{\tilde{K}} = \sum_{\ell=1}^{M} \mathcal{H}_{K_\ell} \quad \text{and} \quad \left\|\tilde{f}\right\|_{\tilde{K}}^2 = \min\left\{\sum_{\ell=1}^{M} \|f_\ell\|_{K_\ell}^2\right\}, \tag{3.9}$$

where the minimum is taken over all decompositions $\tilde{f} = \sum_{\ell=1}^{M} f_\ell \in \mathcal{H}_{\tilde{K}}$, so that $f_\ell \in \mathcal{H}_{K_\ell}$ for $\ell = 1, \ldots, M$. Analogue arguments as in (i) and (ii) provide

$$\mathcal{H}_K = \mathcal{H}_{\tilde{K}} + \mathcal{H}_{K_{M+1}} = \sum_{\ell=1}^{M+1} \mathcal{H}_{K_\ell}$$

and $\|f\|_K^2 = \min\left\{\left\|\tilde{f}\right\|_{\tilde{K}}^2 + \|f_{M+1}\|_{K_{M+1}}^2\right\}$ for all $f \in \mathcal{H}_K$, where the minimum is taken over all representations $f = \tilde{f} + f_{M+1}$, where $\tilde{f} \in \mathcal{H}_{\tilde{K}}$ and $f_{M+1} \in \mathcal{H}_{K_{M+1}}$. In fact, this equals

$$\|f\|_K = \min\left\{\sum_{\ell=1}^{M+1} \|f_\ell\|_{K_\ell}\right\},$$

where the minimum is taken over all decompositions $f = \sum_{\ell=1}^{M+1} f_\ell \in f \in \mathcal{H}_K$, so that $f_\ell \in \mathcal{H}_{K_\ell}$ for $\ell = 1, \ldots, M+1$ by the induction hypothesis (3.9).

If the set of spaces $\{\mathcal{H}_{K_\ell}\}_{\ell=1}^{M+1}$ is complemented, then in particular $\{\mathcal{H}_{K_\ell}\}_{\ell=1}^{M}$ is complemented as well. By the induction hypothesis, it is

$$\mathcal{H}_{\tilde{K}} = \bigoplus_{\ell=1}^{M} \mathcal{H}_{K_\ell} \quad \text{and} \quad \left\langle \tilde{f}, \tilde{g} \right\rangle_{\tilde{K}} = \sum_{\ell=1}^{M} \langle f_\ell, g_\ell \rangle_{K_\ell},$$

where $\tilde{f} = \sum_{\ell=1}^{M} f_\ell$, $\tilde{g} = \sum_{\ell=1}^{M} g_\ell \in \mathcal{H}_{\tilde{K}}$, with $f_\ell, g_\ell \in \mathcal{H}_{K_\ell}$ for $\ell = 1, \ldots, M$, denote the unique decomposition of $\tilde{f}$ and $\tilde{g}$. As $\sum_{\ell=1}^{M} \mathcal{H}_{K_\ell} \cap \mathcal{H}_{K_{M+1}} = \{0\}$ holds by the assumption, analogue arguments as in (i) provide

$$\mathcal{H}_K = \mathcal{H}_{\tilde{K}} \oplus \mathcal{H}_{K_{M+1}} = \bigoplus_{\ell=1}^{M+1} \mathcal{H}_{K_\ell}$$

and

$$\langle f, g \rangle_K = \left\langle \tilde{f}, \tilde{g} \right\rangle_{\tilde{K}} + \langle f_{M+1}, g_{M+1} \rangle_{K_{M+1}} = \sum_{\ell=1}^{M+1} \langle f_\ell, g_\ell \rangle_{K_\ell},$$

where unique representations of f and g are given by $f = \tilde{f} + f_{M+1} = \sum_{\ell=1}^{M+1} f_\ell$ and $g = \tilde{g} + g_{M+1} = \sum_{\ell=1}^{M+1} g_\ell \in \mathcal{H}_K$, with $\tilde{f}, \tilde{g} \in \mathcal{H}_{\tilde{K}}$ and $f_\ell, g_\ell \in \mathcal{H}_{K_\ell}$ for $\ell = 1, \ldots, M+1$. ■

Each component kernel K_ℓ of a summation kernel K can be connected to a mapping from $\mathcal{H}_K$ onto $\mathcal{H}_{K_\ell}$, see [Aro50, Part I, §7, Theorem IV].

Theorem 3.10. *Let K be the reproducing kernel of $\mathcal{H}_K$, such that K can be decomposed into symmetric and positive semi-definite kernels K_1 and K_2, i.e.,*

$$K = K_1 + K_2.$$

Then, this decomposition corresponds to a decomposition of the identity operator $I : \mathcal{H}_K \longrightarrow \mathcal{H}_K$ into positive operators L_1 and L_2

$$I = L_1 + L_2,$$

where

$$L_\ell f(x) = \left\langle f, K_\ell(x,\cdot)\right\rangle_K \qquad \text{for } \ell = 1,2. \tag{3.10}$$

Furthermore, it is $\operatorname{Im}(L_\ell^{1/2}) = \mathcal{H}_{K_\ell}$ with the norm $\left\|L_\ell^{1/2} f\right\|_{K_\ell} = \|f\|_K$ for all $f \in \mathcal{H}_{K_\ell}$ for $\ell = 1,2$. The norm is well-defined since $L_\ell^{1/2}$ establishes a one-to-one correspondence between the quotient space $\mathcal{H}_K \backslash \ker(L_\ell)$ and $\mathcal{H}_{K_\ell}$ for $\ell = 1,2$.

Conversely, to each decomposition $I = L_1 + L_2$ of the identity operator in two positive operators, it is $\operatorname{Im}(L_\ell^{1/2}) = \mathcal{H}_{K_\ell}$ with the norm $\left\|L_\ell^{1/2} f\right\|_{K_\ell} = \|f\|_K$ for all $f \in \mathcal{H}_{K_\ell}$, where $K_\ell = L_\ell K$ for $\ell = 1,2$. Furthermore, $K = K_1 + K_2$ holds.

Remark 3.11. In the setting of Theorem 3.10 it is $\mathcal{H}_{K_\ell} \subseteq \mathcal{H}_K$, by Theorem 3.9. We directly deduce, that

$$\mathcal{H}_K = \mathcal{H}_{K_\ell} \oplus \ker(L_\ell) \qquad \text{for } \ell = 1,2.$$

Hence, $\mathcal{H}_K$ is an orthogonal sum of $\mathcal{H}_{K_1}$ and $\mathcal{H}_{K_2}$ if and only if the following relations hold

$$\ker(L_1) = \mathcal{H}_{K_2} \qquad \Longleftrightarrow \qquad \mathcal{H}_K = \mathcal{H}_{K_1} \oplus \mathcal{H}_{K_2} \qquad \Longleftrightarrow \qquad \ker(L_2) = \mathcal{H}_{K_1}.$$

By the definition of the linear operator in (3.10), this is the case if and only if for $k, \ell \in \{1,2\}$ with $\ell \neq k$,

$$\langle f, K_\ell(x,\cdot)\rangle_K = 0 \qquad \text{for all } f \in \mathcal{H}_{K_k}.$$

This in turn holds true if and only if

$$\langle f, K_\ell(x,\cdot)\rangle_K = \langle f, K_\ell(x,\cdot)\rangle_{K_\ell} \qquad \text{for all } f \in \mathcal{H}_{K_\ell} \text{ and } \ell = 1,2$$

by the proof of Theorem 3.9.

Complemented native spaces are advantageous as they provide a unique decomposition, and hence a straightforward formula for the inner product of the summation kernel's Hilbert space, as demonstrated in Theorem 3.9. Characterizations for complemented native spaces are extracted from the theory above and cumulated in the subsequent theorem.

Theorem 3.12. *Let K be the summation kernel of the component kernels K_1 and K_2 which are symmetric positive semi-definite on $\Omega \subseteq \mathbb{R}^d$. Then, the following statements are equivalent:*

(i) $\left\langle K_1(x,\cdot), K_2(y,\cdot)\right\rangle_K = 0$ *for all* $x, y \in \Omega$.

(ii) $\mathcal{H}_{K_1} \cap \mathcal{H}_{K_2} = \{0\}$.

(iii) $\mathcal{H}_K = \mathcal{H}_{K_1} \oplus \mathcal{H}_{K_2}$.

(iv) *Every* $f \in \mathcal{H}_K$ *can be uniquely decomposed into* $f = f_1 + f_2$*, where* $f_\ell \in \mathcal{H}_{K_\ell}$ *for* $\ell = 1, 2$ *and*

$$\langle f, g\rangle_K = \langle f_1, g_1\rangle_{K_1} + \langle f_2, g_2\rangle_{K_2},$$

with $f = f_1 + f_2$ *and* $g = g_1 + g_2$.

(v) $\langle K_\ell(x,\cdot), f\rangle_K = \langle K_\ell(x,\cdot), f\rangle_{K_\ell}$ *for all* $f \in \mathcal{H}_{K_\ell}$ *and all* $x \in \Omega$, $\ell = 1, 2$.

(vi) *There is a decomposition* $I = L_1 + L_2$ *of the identity operator in two positive operators* L_1 *and* L_2*, that satisfy* $\ker(L_\ell) = \mathcal{H}_{K_k}$ *for* $k, \ell \in \{1, 2\}$ *with* $\ell \neq k$.

Proof.
$(i) \Rightarrow (ii)$ As the inner product is linear, we quickly see $\left\langle f_1, f_2\right\rangle_K = 0$ for all $f_\ell \in S_{K_\ell,\Omega}$, $\ell = 1, 2$. This can be extended to all $f_l \in \mathcal{H}_{K_\ell}$, $\ell = 1, 2$, by the definition of the native space. Then $\mathcal{H}_{K_1} \perp \mathcal{H}_{K_2}$ and (ii) follows by Remark 3.8.
$(ii) \Rightarrow (iii)$ See Theorem 3.9.
$(iii) \Rightarrow (ii)$ For subspaces U, W of a vector space V, $U + W$ is an internal direct sum if and only if $U \cap W = \{0\}$, see [Axl15, Theorem 1.45].
$(iii) \Rightarrow (iv)$ See Definition 3.7 and Remark 3.8.
$(iv) \Rightarrow (i)$ The unique decomposition of $K_1(x,\cdot)$ and $(K_2)(y,\cdot)$ consists of the kernels themselves and the zero function in the respective spaces. This yields

$$\langle K_1(x,\cdot), K_2(y,\cdot)\rangle_K = \langle K_1(x,\cdot), 0\rangle_{K_1} + \langle 0, K_2(y,\cdot)\rangle_{K_2} = 0.$$

$(iv) \Rightarrow (v)$ For any $f \in \mathcal{H}_{K_1} \subseteq \mathcal{H}_K$ its unique decomposition is given by $f + 0 = f \in \mathcal{H}_K$. Let $f \in \mathcal{H}_{K_1}$, then

$$\left\langle K_1(x,\cdot), f\right\rangle_K = \left\langle K_1(x,\cdot), f\right\rangle_{K_1} + \langle 0, 0\rangle_{K_2}.$$

$(v) \Rightarrow (iii)$ See Remark 3.11.
$(iii) \Leftrightarrow (vi)$ See the second statement of Theorem 3.10 and Remark 3.11. ■

Remark 3.13. The above statements can be generalized to arbitrary, but finite sums using induction.

3.2.2 Subsets

The relation between summation kernels and the concept of subsets of RKHS is significant. We introduce results of [Aro50] (Theorem 3.16 and 3.17) and establish the order '$\lesssim$' (Definition 3.19) on the space of native spaces (and equivalently on the space of symmetric positive definite kernels). This order provides insights into the relationship between component kernels and their summation kernels (Theorem 3.25). Furthermore, we draw a link to Sobolev spaces in this section, and illuminate the relation between the Matérn and Wendland kernels in Corollary 3.18 and Corollary 3.24 as they are both reproducing to Sobolev spaces.

We begin this section with a definition from [Aro50, Part I, §7].

Definition 3.14. Let K_1 and K_2 be symmetric positive semi-definite kernels. If $K_2 - K_1$ is a symmetric positive semi-definite kernel, we denote

$$K_1 \ll K_2.$$

With this definition at hand, we deduce the following statements.

Lemma 3.15. *Let K_1 and K_2 be symmetric positive semi-definite kernels.*

(i) $K_1 \ll K_2$ if and only if the inequality

$$0 \leqslant \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j\, K_1(x_i, x_j) \leqslant \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j\, K_2(x_i, x_j) \tag{3.11}$$

holds for any finite pairwise distinct data set $\{x_1, \ldots, x_N\} = X \subseteq \Omega$ and $c \in \mathbb{R}^N$.

(ii) The relation '$\ll$' partially orders the set of symmetric positive semi-definite kernels.

(iii) Let $K_1 \ll K_2$, then $aK_1 \ll K_2$ for every $0 \leqslant a \leqslant 1$ and $K_1 \ll bK_2$ for every $b \geqslant 1$.

Proof. The first statement (i) follows directly from Definition 2.7 and Definition 3.14. We check the requirements for a partial order to show (ii).

1. Reflexivity: $K \ll K$ holds, as $K - K = 0$ is symmetric positive semi-definite.
2. Antisymmetry: Let $K_1 \ll K_2$ and $K_2 \ll K_1$. With $X = \{x\}$, $c = 1$ and (3.11) we deduce
$$K_1(x,x) \leqslant K_2(x,x) \quad (\text{and } \geqslant).$$
Hence, $K_1(x,x) = K_2(x,x)$ for every $x \in \Omega$. With $X = \{x, y\}$, $c = (1,1)^T$ and (3.11) it is
$$K_1(x,x) + 2K_1(x,y) + K_1(y,y) \leqslant K_2(x,x) + 2K_2(x,y) + K_2(y,y) \quad (\text{and } \geqslant).$$
We reduce this inequality to $K_1(x,y) \leqslant K_2(x,y)$ (and $\geqslant$) as $K_1(x,x) = K_2(x,x)$ for all $x \in \Omega$. Which proves $K_1(x,y) = K_2(x,y)$ for all $x, y \in \Omega$.

3. Transitivity: Let $K_1 \ll K_2$ and $K_2 \ll K_3$. By (3.11) it is

$$0 \leqslant \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j\, K_1(x_k, x_l) \leqslant \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j\, K_2(x_k, x_l) \leqslant \sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j\, K_3(x_k, x_l).$$

This proves $K_1 \ll K_3$.

To show the third result (*iii*) we use (3.11) to gain $aK_1 \ll K_2$ and $K_1 \ll bK_2$ for $0 \leqslant a \leqslant 1$ and $b \geqslant 1$. The transitivity of '$\ll$' yields the desired result. ■

The relation '$\ll$' between to kernels provides information about the relation between the corresponding RKHSs. We state [Aro50, Part I, §7 Theorem I].

Theorem 3.16. *Let K_1 and K_2 be reproducing kernels of the spaces $\mathcal{H}_{K_1}$ and $\mathcal{H}_{K_2}$ and let $K_1 \ll K_2$. Then $\mathcal{H}_{K_1} \subseteq \mathcal{H}_{K_2}$ and $\|f\|_{K_1} \geqslant \|f\|_{K_2}$ for every $f \in \mathcal{H}_{K_1}$.*

In Lemma 2.29 we saw, that kernels which only differ by a constant factor reproduce the same set of functions. However, this relation is not sufficient. The Wendland and Matérn kernels, introduced in Section 2.3.4, are both reproducing kernels of Sobolev spaces with equivalent norms. To be more precise, the kernels corresponding to the RBFs ψ_ℓ (Matérn kernel, Tab. 2.1) and $\phi_{d,\ell}$ (Wendland kernel, Tab. 2.2) are both reproducing kernels of the Sobolev space $W^{\frac{d+1}{2}+\ell}(\mathbb{R}^d)$ and their native spaces have equivalent norms. But, as the Wendland kernel has compact support and the Matérn kernel does not, they cannot be multiplicands of one another. Nevertheless, we deduce a relation between the two kernels in Corollary 3.24. To do so, we need [Aro50, Part I, §7, Theorem II] stated below.

Theorem 3.17. *Let the Hilbert space H with the norm $\|\cdot\|$ be a subset of $\mathcal{H}_{K_2}$, such that $\|f\| \geqslant \|f\|_{K_2}$ for every $f \in H$. Then, H possesses a reproducing kernel K_1 satisfying $K_1 \ll K_2$.*

With Theorem 3.17, we can deduce a relation between the Matérn kernels corresponding to ψ_0 and ψ_1. Since the norms $\|\cdot\|_{\psi_\ell}$ and $\|\cdot\|_{W^{\frac{d+1}{2}+\ell}}$ are equivalent for $\ell \in \mathbb{N}_0$, there exist positive constants a, c and C, so that the equations

$$\|\cdot\|_{c\psi_\ell} = \frac{1}{c}\|\cdot\|_{\psi_\ell} \leqslant \|\cdot\|_{W^{\frac{d+1}{2}+\ell}} \leqslant a\|\cdot\|_{W^{\frac{d+1}{2}+m}} \leqslant \frac{a}{C}\|\cdot\|_{\psi_m} = \|\cdot\|_{\frac{C}{a}\psi_m}$$

and

$$\mathcal{H}_{\frac{C}{a}\psi_m} = \mathcal{H}_{\psi_m} = W^{\frac{d+1}{2}+m} \subseteq W^{\frac{d+1}{2}+\ell} = \mathcal{H}_{\psi_\ell} = \mathcal{H}_{c\psi_\ell}$$

hold for all $\ell, m \in \mathbb{N}_0$ with $\ell \leqslant m$. Applying Theorem 3.17 on the special case of the Matérn kernels ψ_ℓ results in the following relation, and holds by analogue arguments for Wendland kernels $\phi_{\ell,d}$ as well.

Corollary 3.18. *For all $\ell, m \in \mathbb{N}_0$ with $\ell \leqslant m$ there exists a constant $C > 0$, so that the Matérn kernels satisfy $\psi_\ell \ll C\psi_m$.*
For all $d \in \mathbb{N}$ and $\ell, m \in \mathbb{N}_0$ with $\ell \leqslant m$ there exists a constant $C > 0$, so that the Wendland kernels satisfy $\phi_{\ell,d} \ll C\phi_{m,d}$.

To simplify notations, we introduce a relation on the set of Hilbert spaces, dating back to [Sch64, Chapter 7], as well as a relation on the set of reproducing kernels.

Definition 3.19. Let H_1 and H_2 be Hilbert spaces. If $H_1 \subseteq H_2$ and $c\,\|f\|_1 \geqslant \|f\|_2$ for a constant $c > 0$ and all $f \in H_1$, we denote

$$H_1 \lesssim H_2.$$

Let K_1 and K_2 be symmetric positive semi-definite kernels. If $cK_2 - K_1$ is a positive semi-definite kernel for a constant $c > 0$, we denote

$$K_1 \lesssim K_2.$$

The relation '$\lesssim$' is reflexive and transitive for both, the set of Hilbert spaces and the set of reproducing kernels. Hence, it defines a preorder on both sets. We derive equivalence statements of these preorders, by summarizing results of [Aro50], [Ylv62] and [Dri73].

Theorem 3.20. *Let the kernels K_1 and K_2 be reproducing kernels to the Hilbert spaces $(H_1, \|\cdot\|_1)$ and $(H_2, \|\cdot\|_2)$ respectively. Then, the following statements are equivalent:*

(i) $H_1 \lesssim H_2$.

(ii) $K_1 \lesssim K_2$.

(iii) *Let* $\left(\sum_{i=1}^{N(n)} c_{i(n)} K_2(\cdot, x_{i(n)})\right)_{n\in\mathbb{N}}$ *be a Cauchy sequence in* H_2*, then* $\left(\sum_{i=1}^{N(n)} c_{i(n)} K_1(\cdot, x_{i(n)})\right)_{n\in\mathbb{N}}$ *is a Cauchy sequence in* H_1.

(iv) *There is a bounded linear operator* $L : H_2 \longrightarrow H_1$ *such that* $LK_2(x,\cdot) = K_1(x,\cdot)$ *for every* x.

Proof. The equivalence of *(i)* and *(ii)* is a direct consequence of Theorem 3.16 and Theorem 3.17, and can be found in [Aro50, Part I, §13, Corollary IV_2]. The equivalence of *(iii)* can be shown using Theorem 3.9 and the reproducing property. For details, we refer to [Ylv62, Theorem 2.4]. The connection to *(iv)* was first drawn in [Dri73, Theorem 1]. This result can be established by applying Theorem 3.10. ∎

Remark 3.21. Every preorder gives rise to a reflexive, symmetric and transitive equivalence relation. We define the equivalence relation '$\sim$' on Hilbert spaces by

$$H_1 \sim H_2 \quad \text{if and only if} \quad H_1 \lesssim H_2 \text{ and } H_2 \lesssim H_1.$$

In particular, the equivalence relation '$\sim$' holds for the RKHSs $\mathcal{H}_{K_1}$ and $\mathcal{H}_{K_2}$ if and only if their corresponding reproducing kernels K_1 and K_2 fulfill the equivalence relation '$\sim$' defined by

$$K_1 \sim K_2 \quad \text{if and only if} \quad K_1 \lesssim K_2 \text{ and } K_2 \lesssim K_1.$$

This is a direct consequence of Theorem 3.20.

Equipped with these definitions we deduce the following equivalence statements from Theorem 3.20.

Lemma 3.22. *Let the kernels K_1 and K_2 be reproducing to the Hilbert spaces $(H_1, \|\cdot\|_1)$ and $(H_2, \|\cdot\|_2)$, respectively. Then, the following statements are equivalent:*

(i) $H_1 \sim H_2$

(ii) $K_1 \sim K_2$

(iii) $\left(\sum_{i=1}^{N(n)} c_{i(n)} K_1(\cdot, x_{i(n)})\right)_{n\in\mathbb{N}}$ *is a Cauchy sequence in* $(H_1, \|\cdot\|_1)$ *if and only if* $\left(\sum_{i=1}^{N(n)} c_{i(n)} K_2(\cdot, x_{i(n)})\right)_{n\in\mathbb{N}}$ *is a Cauchy sequence in* $(H_2, \|\cdot\|_2)$.

Remark 3.23. We want to emphasize the finding of [Aro50, Corollary IV$_1$, §13] relying on the above equivalent statements (i) and (ii). If two kernels K_1 and K_2 are reproducing to the same set of functions $\mathcal{H}_{K_1} = \mathcal{H}_{K_2}$, their corresponding native space norms are equivalent, i.e., there exists constants $c, C > 0$ so that $c\|\cdot\|_{K_1} \leqslant \|\cdot\|_{K_2} \leqslant C\|\cdot\|_{K_1}$.

This remark, applied to the Wendland and Matérn kernel $\phi_{d,\ell}$ and ψ_ℓ, gives the following.

Corollary 3.24. *For $d \in \mathbb{N}$, $m \in \mathbb{N}_0$ and $d \geqslant 3$ if $\ell = 0$ it is $\phi_{d,\ell} \sim \psi_\ell$.*

Proof. The kernels corresponding to the radial functions $\phi_{d,\ell}$ and ψ_ℓ reproduce the same Sobolev space $W^{\frac{d+1}{2}+\ell}(\mathbb{R}^d)$, where $d \in \mathbb{N}$, $\ell \in \mathbb{N}_0$ and $d \geqslant 3$ if $\ell = 0$, with norms equivalent to the Sobolev space norm. Hence, the norms $\|\cdot\|_{\phi_{d,\ell}}$ and $\|\cdot\|_{\psi_\ell}$ induced by the kernel functions are equivalent as well. Lemma 3.22 yields the statement. ∎

The classification of reproducing kernels into equivalence classes is important for assessing whether adding kernels to a summation kernel is reasonable.

Theorem 3.25. *Let K_1 and K_2 be reproducing kernels and $K = K_1 + K_2$ their summation kernel. Then the following statements hold true:*

(i) $K_1 \lesssim K$ *and* $K_2 \lesssim K$.

(ii) From $K_1 \lesssim K_2$ *follows* $K \sim K_2$.

(iii) From $K_1 \sim K_2$ *follows* $K \sim K_1 \sim K_2$.

Proof. The first finding (i) is given by Theorem 3.9 and Definition 3.14. Let $K_1 \lesssim K_2$ then $C\,K_2 - K_1$ is a symmetric positive semi-definite kernel for a constant $C > 0$. Adding and subtracting K_2 gives that $(C+1)K_2 - (K_1 + K_2) = (C+1)K_2 - K$ is a symmetric positive semi definite kernel. This yields

$$K = K_1 + K_2 \ll (C+1)\,K_2$$

and hence $K \lesssim K_2$. Together with $K_2 \lesssim K$ of (i), this implies $K \sim K_2$ and (ii) is shown. If additionally $K_2 \lesssim K_1$ is given, as in (iii), analogue arguments show $K \sim K_1$. ∎

It is worth mentioning that the equivalence class of a kernel is invariant under the summation with a kernel of a smaller class. Consequently, the summation of kernels that do not stand in the '$\lesssim$' relation is of particular interest. Because, in that case the summation kernel belongs to a third larger equivalence class and is hence reproducing to a strict superset of every component kernels' native space.

3.2.3 Intersections

Next, we study the intersection of two native spaces. In Theorem 3.26 we show that the intersection is again a RKHS and examine the relation of its so-called intersection kernel to the initial 'superior' reproducing kernels. In Theorem 3.27 we provide norm estimates for functions lying in the intersection.

Theorem 3.26. *Let K and K' be symmetric positive semi-definite kernels. Then, there exists a unique symmetric positive semi-definite kernel κ_1, called* intersection kernel, *such that*

$$\mathcal{H}_{\kappa_1} = \mathcal{H}_K \cap \mathcal{H}_{K'} \tag{3.12}$$

and

$$\langle f, g\rangle_{\kappa_1} = \langle f, g\rangle_K + \langle f, g\rangle_{K'} \tag{3.13}$$

holds for all $f, g \in \mathcal{H}_{\kappa_1}$. Additionally,

(i) $\kappa_1 \ll K$ and $\kappa_1 \ll K'$.

(ii) The intersection is trivial $\mathcal{H}_K \cap \mathcal{H}_{K'} = \{0\}$ if and only if the intersection kernel $\kappa_1 \equiv 0$.

(iii) There exist positive semi-definite kernels κ_2 and κ_2' such that

$$K = \kappa_1 + \kappa_2 \qquad \text{and} \qquad K' = \kappa_1 + \kappa_2', \tag{3.14}$$

and

$$\mathcal{H}_{\kappa_2} \cap \mathcal{H}_{\kappa_2'} = \{0\}. \tag{3.15}$$

(iv) From $K \lesssim K'$ follows $\kappa_1 \sim K$.

(v) From $K \sim K'$ follows $\kappa_1 \sim K \sim K'$.

Proof. The intersection equipped with the inner product $\langle f, g\rangle := \langle f, g\rangle_K + \langle f, g\rangle_{K'}$ and norm $\|f\|^2 = \langle f, f\rangle$ forms a Hilbert space. Furthermore, $\mathcal{H}_K \cap \mathcal{H}_{K'}$ is a subset of $\mathcal{H}_K$ and $\mathcal{H}_{K'}$. Since $\|f\| \geqslant \|f\|_K$ and $\|f\| \geqslant \|f\|_{K'}$ for every function $f \in \mathcal{H}_K \cap \mathcal{H}_{K'}$, the requirements of Theorem 3.17 are satisfied. Applying it on $\mathcal{H}_K$ and $\mathcal{H}_{K'}$ gives the existence of two reproducing kernels for $\left(\mathcal{H}_K \cap \mathcal{H}_{K'}, \|\cdot\|\right)$. By Theorem 2.22, they must be equal and will be denoted here as κ_1. Hence, (3.12) holds since

$$\langle f, g\rangle_{\kappa_1} = \langle f, g\rangle = \langle f, g\rangle_K + \langle f, g\rangle_{K'}.$$

(i) follows directly from the above construction and Theorem 3.17.

(ii) The statement follows from (3.12). Either the right-hand side equals $\{0\}$, hence $\{0\} = \mathcal{H}_{\kappa_1}$ and $\kappa_1 \equiv 0$ follows, or $\kappa_1 \equiv 0$ holds and therefore

$$\{0\} = \mathcal{H}_{\kappa_1} = \mathcal{H}_K \cap \mathcal{H}_{K'}.$$

(iii) Because of (i), the kernels $\kappa_2 := K - \kappa_1$ and $\kappa_2' := K' - \kappa_1$ are symmetric positive semi-definite. If $\mathcal{H}_{\kappa_2} \cap \mathcal{H}_{\kappa_2'} \neq \{0\}$ then κ_1 was no intersection kernel as its native space did not cover the whole intersection $\mathcal{H}_K \cap \mathcal{H}_{K'}$.

(iv) Since $K \lesssim K'$ we have $\mathcal{H}_K \subseteq \mathcal{H}_{K'}$, by Remark 3.21. This implies the intersection space to be $\mathcal{H}_{\kappa_1} = \mathcal{H}_K$ and by the construction of κ_1 it is $\|\cdot\|_{\kappa_1} \geqslant \|\cdot\|_K$ which gives the relation $\mathcal{H}_{\kappa_1} \sim \mathcal{H}_K$ and hence $\kappa_1 \sim K$.

(v) is a direct consequence of (iv).

■

We remark that the intersection kernel κ_1 of K and K' belongs to third and smaller equivalence class of kernels if the initial kernels K and K' are not ordered by the '$\lesssim$' relation. Examining Theorem 3.26 and its proof, we observe that $\|f\|_{\kappa_1} \geqslant \|f\|_K$ and $\|f\|_{\kappa_1} \geqslant \|f\|_{K'}$ for all $f \in \mathcal{H}_{\kappa_1}$. We aim at developing a relation between $\|f\|_K$ and $\|f\|_{K'}$. Note that the orthogonality requirements of the upcoming theorem can be replaced by any of the equivalent statements of Theorem 3.12.

Theorem 3.27. *Let K and K' be symmetric positive semi-definite kernels on Ω and κ_1 their intersection kernel, so that*

$$K = \kappa_1 + \kappa_2 \qquad \text{and} \qquad K' = \kappa_1 + \kappa_2',$$

with positive semi-definite kernels κ_2, κ_2'.

(i) From $\mathcal{H}_K = \mathcal{H}_{\kappa_1} \oplus \mathcal{H}_{\kappa_2}$ follows $\|f\|_K \geqslant \|f\|_{K'}$ for all $f \in \mathcal{H}_{\kappa_1}$.

(ii) Let (i) and additionally $\mathcal{H}_{K'} = \mathcal{H}_{\kappa_1} \oplus \mathcal{H}_{\kappa_2'}$ hold. Then, $\|f\|_K = \|f\|_{K'}$ for all $f \in \mathcal{H}_{\kappa_1}$.

Proof. By Theorem 3.9 it is $\|f\|_K = \|f\|_{\kappa_1}$ for all $f \in \mathcal{H}_{\kappa_1}$ and from the same theorem we deduce $\|f\|_{\kappa_1} \geqslant \|f\|_{K'}$ for all $f \in \mathcal{H}_{\kappa_1}$. This yields the first statement. The second follows by the additional assumption $\|f\|_{K'} = \|f\|_{\kappa_1}$ for all $f \in \mathcal{H}_{\kappa_1}$ holds.

■

3.2.4 Infinite Sums and Mercer's Theorem

In this section, we establish a link between the famous Mercer theorem and summation kernels. To do so, we begin by examining infinitely countable orthogonal decompositions of native spaces in Theorem 3.28, finding that the corresponding reproducing kernel can be expressed as an infinitely countable sum of kernels. We then present the main result of J. Mercer from [Mer09] in Theorem 3.30, which promotes the infinitely countable decomposition of a kernel. Finally, Lemma 3.34 demonstrates that this kernel decomposition corresponds to an orthogonal decomposition of the kernel's native space.

N. Aronszajn demonstrates that if a RKHS with reproducing kernel K can be decomposed into complemented subspaces, then there exist reproducing kernels K_1 and K_2 for those subspaces so that $K = K_1 + K_2$ (cf. [Aro50, Part I, §12] and Theorem 3.10). Subsequently, this statement is extended to countable infinite sums.

Theorem 3.28. *Let K be bounded symmetric positive semi-definite on Ω, H_ℓ subspaces of $\mathcal{H}_K$, and*

$$\mathcal{H}_K = \bigoplus_{\ell=1}^{\infty} H_\ell = \left\{ \sum_{\ell=1}^{\infty} f_\ell \ : \ f_\ell \in H_\ell \text{ for } \ell \in \mathbb{N} \right\}.$$

Then there exist symmetric positive semi-definite kernels K_ℓ for $\ell \in \mathbb{N}$, so that

$$K(x,y) = \sum_{\ell=1}^{\infty} K_\ell(x,y) \qquad \text{and} \qquad \big(H_\ell, \langle\cdot,\cdot\rangle_K\big) = \big(\mathcal{H}_{K_\ell}, \langle\cdot,\cdot\rangle_{K_\ell}\big)$$

Furthermore, $\sum_{\ell=1}^{\infty} K_\ell^2(x,y) < \infty$ for all $x, y \in \Omega$.

Proof. Let $f = \sum_{\ell=1}^{\infty} f_\ell, g = \sum_{\ell=1}^{\infty} g_\ell$ be the unique representations of $f, g \in \mathcal{H}_K$ with $f_\ell, g_\ell \in H_\ell$ for $\ell \in \mathbb{N}$. Then,

$$\langle f, g\rangle_K = \left\langle \sum_{\ell=1}^{\infty} f_\ell, \sum_{\ell=1}^{\infty} g_\ell \right\rangle_K = \sum_{\ell=1}^{\infty} \langle f_\ell, g_\ell\rangle_K\,,$$

by the orthogonality assumption. This implies convergence

$$\sum_{\ell=1}^{\infty} \|f_\ell\|_K^2 = \|f\|_K^2 < \infty. \tag{3.16}$$

Let $x \in \Omega$ be arbitrarily chosen and $k_\ell^x \in H_\ell$ for $\ell \in \mathbb{N}$ be the unique functions representig $K(x,\cdot) \in \mathcal{H}_K$, so that

$$K(x,\cdot) = \sum_{\ell=1}^{\infty} k_\ell^x. \tag{3.17}$$

It follows that

$$f_\ell(x) = \langle f_\ell, K(x,\cdot)\rangle_K = \langle f_\ell, k_\ell^x\rangle_K + \sum_{m=1, m\neq \ell}^{\infty} \langle 0, k_m^x\rangle_K = \langle f_\ell, k_\ell^x\rangle_K \qquad \text{for all } f_\ell \in H_\ell.$$

We define $K_\ell(x,\cdot) := k_\ell^x$ for all $x \in \Omega$ and $\ell \in \mathbb{N}$. Hence, K_ℓ is the reproducing kernel of the Hilbert space $(H_\ell, \langle\cdot,\cdot\rangle_K)$. By the uniqueness properties of reproducing kernels of Theorem 2.22, it is $(H_\ell, \langle\cdot,\cdot\rangle_K) = (\mathcal{H}_{K_\ell}, \langle\cdot,\cdot\rangle_{K_\ell})$. Consequently, (3.16) and (3.17) turn into

$$\|f\|_K^2 = \sum_{\ell=1}^{\infty} \|f_\ell\|_{K_\ell}^2 \qquad \text{and} \qquad K(x,y) = \sum_{\ell=1}^{\infty} K_\ell(x,y).$$

This finding and the Cauchy-Schwarz inequality yield

$$\begin{aligned}
\sum_{\ell=1}^{\infty} K_\ell^2(x,y) &= \sum_{\ell=1}^{\infty} |\langle K_\ell(x,\cdot), K_\ell(y,\cdot)\rangle_{K_\ell}|^2 \\
&\leqslant \sum_{\ell=1}^{\infty} \|K_\ell(x,\cdot)\|_{K_\ell}^2 \|K_\ell(y,\cdot)\|_{K_\ell}^2 \\
&\leqslant \sum_{\ell=1}^{\infty} \|K_\ell(x,\cdot)\|_{K_\ell}^2 \sum_{\ell=1}^{\infty} \|K_\ell(y,\cdot)\|_{K_\ell}^2 \\
&= \|K(x,\cdot)\|_K \|K(y,\cdot)\|_K \\
&= K(x,x)K(y,y) < \infty \qquad \text{for all } x,y \in \Omega.
\end{aligned}$$

■

We obtain a formula for the component kernels in the case of separable native spaces, see [BTA11, Theorem 14].

Lemma 3.29. *Let $\mathcal{H}_K$ be a separable native space with countable orthonormal basis $\{e_\ell\}_{\ell\in\mathbb{N}}$ and bounded reproducing kernel K, then there exist an orthogonal decomposition so that*

$$\mathcal{H}_K = \bigoplus_{\ell=1}^{\infty} \mathcal{H}_{K_\ell},$$

where $K_\ell(x,y) = e_\ell(x)e_\ell(y)$ for a countable orthonormal basis $\{e_\ell\}_{\ell\in\mathbb{N}}$ of $\mathcal{H}_K$. Furthermore,

$$\begin{aligned}
K(x,y) &= \sum_{\ell=1}^{\infty} K_\ell(x,y) = \sum_{\ell=1}^{\infty} e_\ell(x)e_\ell(y) && \text{and} \\
\sum_{\ell=1}^{\infty} e_\ell^2(x)e_\ell^2(y) &< \infty && \text{for all } x,y \in \Omega.
\end{aligned}$$

Proof. As $\mathcal{H}_K$ is a separable Hilbert space it has a countable orthonormal basis $\{e_\ell\}_{\ell\in\mathbb{N}}$. Let $\ell \in \mathbb{N}$, then the subset $\overline{\operatorname{span}\{e_\ell\}}^K \subset \mathcal{H}_K$, equipped with the same inner product $\langle\cdot,\cdot\rangle_K$, forms a Hilbert space. By Theorem 3.17, $\overline{\operatorname{span}\{e_\ell\}}^K$ is a reproducing kernel Hilbert space, and we denote its reproducing kernel with K_ℓ. Hence,

$$\mathcal{H}_K = \bigoplus_{\ell=1}^{\infty} \overline{\operatorname{span}\{e_\ell\}}^K = \bigoplus_{\ell=1}^{\infty} \mathcal{H}_{K_\ell}$$

and as $f = \sum_{\ell=1}^{\infty} \langle f, e_\ell\rangle_K e_\ell$ with absolute convergence (basic functional analysis) for all $f \in \mathcal{H}_K$ it is

$$\langle f, g\rangle_K = \left\langle \sum_{\ell=1}^{\infty} \langle f, e_\ell\rangle_K e_\ell, g \right\rangle_K = \sum_{\ell=1}^{\infty} \langle f_\ell, g\rangle_K = \sum_{\ell=1}^{\infty}\sum_{m=1}^{\infty} \langle f_\ell, g_m\rangle_K = \sum_{\ell=1}^{\infty} \langle f_\ell, g_\ell\rangle_K ,$$

where $f_\ell = \langle f, e_\ell\rangle_K e_\ell$ and $g_\ell = \langle g, e_\ell\rangle_K e_\ell \in \mathcal{H}_{K_\ell}$ for all $\ell \in \mathbb{N}$ and the last equation holds because of the orthogonality.

Furthermore, it is

$$K_\ell(x,y) = \left\langle K_\ell(x,\cdot)_K, e_\ell \right\rangle_K e_\ell(y) = e_\ell(x)e_\ell(y),$$

where the first equality holds by Definition 2.20 as $K_\ell(x,\cdot) \in \overline{\operatorname{span}\{e_\ell\}}^K$, and basic properties of orthonormal bases in Hilbert spaces. The second equality is due to the fact that the kernel K_ℓ is reproducing regarding the inner product $\langle\cdot,\cdot\rangle_K$. Together with Theorem 3.28, this implies the last required statement. ∎

The lemma above directs our attention to Mercer's Theorem first published by J. Mercer in 1909 (cf. [Mer09]). We state the version of [Wen05, p. 154f]. Let $\mathcal{T}$ denote the compact Hilbert-Schmidt integral operator

$$\mathcal{T}f(x) = \int_\Omega K(x,y)f(y)dy, \tag{3.18}$$

that maps $L^2(\Omega)$ continuously to $\mathcal{H}_{K,\Omega}$, if K was a symmetric positive semi-definite kernel on the compact set Ω.

Theorem 3.30 (Mercer's Theorem)**.** *Let K be a continuous symmetric positive semi-definite kernel on the compact set Ω. Then*

$$K(x,y) = \sum_{\ell=1}^{\infty} \lambda_\ell \, e_\ell(x)e_\ell(y), \tag{3.19}$$

where $\lambda_\ell \geqslant 0$ are the eigenvalues of the continuous eigenfunctions e_ℓ regarding $\mathcal{T}$ and the set of functions $\{e_\ell\}_{\ell=1}^\infty$ forms an orthonormal basis of $L^2(\Omega)$. The convergence is absolute and uniform.

So far, we have shown that an infinite orthogonal decomposition of a RKHS correspond to a kernel, which is the infinite sum of the corresponding component reproducing kernels. Now, we ask whether from an infinite kernel representation, as Mercer's theorem suggests, it follows that the associated RKHS can be decomposed into an infinite orthogonal sum of RKHSs. To answer this, we need Lemma 3.31 stating that the infinite sum of positive semi-definite kernels remains positive semi-definite. Then, it is possible to deduce Theorem 3.32 as the infinite version of Theorem 3.9. At the end of this section, we analyze the orthogonality of the component native spaces corresponding to the representation of Mercer's theorem.

Lemma 3.31. *Let K_ℓ be positive semi-definite kernels on Ω for $\ell \in \mathbb{N}$ such that $K(x,y) = \sum_{\ell=1}^\infty K_\ell(x,y)$ convergences absolutely for all $x,y \in \Omega$. Then the kernel K is positive semi-definite.*
Additionally, if at least one K_ℓ is positive definite, the kernel K positive definite.

Proof. Let $X = \{x_1,\dots,x_N\} \subset \Omega$ be a pairwise distinct point set. The positive semi-definiteness requirement yields

$$\sum_{i,j=1}^{N} c_i c_j K_\ell(x_i,x_j) \geqslant 0 \qquad \text{for all } c = (c_1,\dots,c_N)^T \in \mathbb{R}^N \text{ and } \ell = 1,\dots,M.$$

Because of the absolute convergence, we can interchange the sums to obtain

$$0 \leqslant \sum_{\ell=1}^{\infty} \sum_{i,j=1}^{N} c_i c_j K_\ell(x_i, x_j) = \sum_{i,j=1}^{N} c_i c_j \sum_{\ell=1}^{\infty} K_\ell(x_i, x_j) = \sum_{i,j=1}^{N} c_i c_j K(x_i, x_j) < \infty.$$

The additional statement directly follows from the findings above, as '$\leqslant$' can be replaced by '$<$'. ■

Now that we understand the circumstances under which countable summation kernels are positive semi-definite, we can turn our attention to the structure of their native spaces. Next we show, that the native space of a countable summation kernel is given by the countable sum of its components' native spaces. This is the same result as Theorem 3.9 states, but now for a countably infinite many component kernels. The proof is similar to the one in the finite setting. Therefore, we solely focus on the differing details in the proof given below.

Theorem 3.32. *Let K_ℓ be symmetric positive semi-definite kernels on Ω for $\ell \in \mathbb{N}$, such that $K(x,y) = \sum_{\ell=1}^{\infty} K_\ell(x,y)$ convergences absolutely for all $x, y \in \Omega$. Then*

$$\mathcal{H}_K = \left\{ \sum_{\ell \in \mathbb{N}} f_\ell \ : \ f_\ell \in \mathcal{H}_{K_\ell} \text{ for } \ell \in \mathbb{N} \text{ and } \sum_{\ell \in \mathbb{N}} \|f_\ell\|_{K_\ell}^2 < \infty \right\}$$

and the norm is given by

$$\|f\|_K^2 = \min \left\{ \sum_{\ell \in \mathbb{N}} \|f_\ell\|_{K_\ell}^2 \right\},$$

where the minimum is taken over all decomposition $\sum_{\ell \in \mathbb{N}} f_\ell$ of $f \in \mathcal{H}_K$, with $f_\ell \in \mathcal{H}_{K_\ell}$ for $\ell \in \mathbb{N}$.

Additionally, let $K_I := \sum_{\ell \in I} K_\ell$ for an index set $I \subset \mathbb{N}$ and

$$\mathcal{H}_{K_I} \cap \mathcal{H}_{K_{\mathbb{N} \setminus I}} = \{0\} \text{ for all index sets } I \subset \mathbb{N}.$$

Then

$$\mathcal{H}_K = \left\{ f = \sum_{\ell \in \mathbb{N}} f_\ell \ : \ f_\ell \in \mathcal{H}_{K_\ell} \text{ for } \ell \in \mathbb{N} \text{ and } \sum_{\ell \in \mathbb{N}} \|f_\ell\|_{K_\ell}^2 < \infty \right\},$$

so that the representation $\sum_{\ell \in \mathbb{N}} f_\ell$ of $f \in \mathcal{H}_K$ is uniquely determined, and the inner product is given by

$$\langle f, g \rangle_K = \sum_{\ell \in \mathbb{N}} \langle f_\ell, g_\ell \rangle_{K_\ell},$$

where $\sum_{\ell \in \mathbb{N}} f_\ell$ and $\sum_{\ell \in \mathbb{N}} g_\ell$ are the unique representations of f and $g \in \mathcal{H}_K$.

Proof. Just as in the proof of Theorem 3.9 that considers the final case, we divide the proof into three steps.

(i) We define the set of series

$$H := \left\{ (f_\ell)_{\ell\in\mathbb{N}} \ : \ f_\ell \in \mathcal{H}_{K_\ell} \text{ and } \sum_{\ell\in\mathbb{N}} \|f_\ell\|_{K_\ell}^2 < \infty \right\}.$$

Equipped with the inner product

$$\langle (f_\ell)_{\ell\in\mathbb{N}}, (g_\ell)_{\ell\in\mathbb{N}} \rangle_H := \sum_{\ell=1}^{\infty} \langle f_\ell, g_\ell \rangle_{K_\ell},$$

H forms a Hilbert space. Furthermore, we define the mapping φ

$$\begin{aligned} \varphi : H &\longrightarrow G := \left\{ \sum_{\ell\in\mathbb{N}} f_\ell \ : \ f_\ell \in \mathcal{H}_{K_\ell} \text{ for } \ell \in \mathbb{N} \text{ and } \sum_{\ell\in\mathbb{N}} \|f_\ell\|_{K_\ell}^2 < \infty \right\}, \\ (f_\ell)_{\ell\in\mathbb{N}} &\longmapsto \sum_{\ell\in\mathbb{N}} f_\ell. \end{aligned}$$

The kernel space $\ker(\varphi)$ consists of the series $(f_\ell)_{\ell\in\mathbb{N}}$ such that $\sum_{\ell\in\mathbb{N}} f_\ell = 0$. This is the case, if and only if there exists an index set $I \subseteq \mathbb{N}$ so that

$$\sum_{\ell\in I} f_\ell = - \sum_{\ell\in\mathbb{N}\setminus I} f_\ell.$$

We denote $\ker\varphi = H_0 \subseteq H$, then $H_0 \oplus H_0^\perp = H$ where $H_0^\perp$ denotes the orthogonal complement of H_0 in H. With this $\varphi|_{H_0^\perp}$ is bijective and we denote

$$\varphi^{-1}(f) = (f'_\ell)_{\ell\in\mathbb{N}} \in H_0^\perp.$$

We equip the sum G with the inner product

$$\langle f, g \rangle_G := \langle (f'_\ell)_{\ell\in\mathbb{N}}, (g'_\ell)_{\ell\in\mathbb{N}} \rangle_H = \sum_{\ell=1}^{\infty} \langle f'_\ell, g'_\ell \rangle_{K_\ell},$$

which makes $\varphi|_{H_0^\perp}$ an isomorphism between $H_0^\perp$ and G.

(ii) Since $\sum_{\ell\in\mathbb{N}} K_\ell = K$ converges absolutely and $K_\ell(x,x) = \|K_\ell(x,\cdot)\|_{K_\ell}^2$ for $\ell \in \mathbb{N}$, we deduce $K \in G$. Furthermore, we denote $\varphi^{-1}(K(x,\cdot)) = (\kappa_\ell)_{\ell\in\mathbb{N}} \in H_0^\perp$. Then

$$\sum_{\ell=1}^{\infty} K_\ell(x,\cdot) = \sum_{\ell=1}^{\infty} \kappa_\ell(x,\cdot)$$

and since

$$K_1(x,\cdot) - \kappa_1(x,\cdot) = -\left(\sum_{\ell=2}^{\infty} K_\ell(x,\cdot) - \kappa_\ell(x,\cdot) \right),$$

we deduce $(K_\ell - \kappa_\ell)_{\ell\in\mathbb{N}} \in H_0$. Analogously to the proof of Theorem 3.9 we obtain that

$$f(x) = \langle f, K(x,\cdot) \rangle_G \qquad \text{for all } f \in G.$$

This makes K the reproducing kernel of G equipped with the inner product $\langle \cdot,\cdot \rangle_K = \langle \cdot,\cdot \rangle_G$.

(iii) Again an analogue computation as in the proof of Theorem 3.9 yields

$$\sum_{\ell\in\mathbb{N}} \|f_\ell\|_{K_\ell}^2 = \|f\|_K^2 + \left\| \sum_{\ell\in\mathbb{N}} (f_\ell - f'_\ell) \right\|_K^2 .$$

Therefore, the required relation of the norms hold.

Taking the additional assumption into account, we obtain

$$\ker\varphi = \left\{ (f_\ell)_{\ell\in\mathbb{N}} \ : \ \sum_{\ell\in I} f_\ell = -\sum_{\ell\in\mathbb{N}\setminus I} f_\ell \right\} = \{0\} \subset H,$$

as $\sum_{\ell\in I} f_\ell = -\sum_{\ell\in\mathbb{N}\setminus I} f_\ell \in \mathcal{H}_{K_I} \cap \mathcal{H}_{K_{\mathbb{N}\setminus I}} = \{0\}$. Then, $H = H_0^\perp$ and therefore the required statements hold true. ■

Let us return to the representation of K given in Mercer's theorem (Theorem 3.30). In order to apply Theorem 3.32, positive semi-definiteness of the component kernels $K_\ell(x,y) := \lambda_\ell e_\ell(x) e_\ell(y)$ is required for every $\ell \in \mathbb{N}$.

Lemma 3.33. *Let $\psi : \Omega \longrightarrow \mathbb{R}$ be a function on $\Omega \subseteq \mathbb{R}^d$. Then, the composition $K(x,y) = \psi(x)\psi(y)$ is a positive semi-definite kernel on Ω.*

Proof. Let $\{x_1, \dots, x_N\} = X$ be a set of points. The interpolation matrix $\mathbf{A}_{K,X}$ is the dyadic product of $\big(\psi(x_1), \dots, \psi(x_N)\big) \in \mathbb{R}^N$ and itself with entries $\big(\psi(x_i)\psi(x_j)\big)_{i,j=1}^N$. Therefore, we can deduce positive semi-definiteness as

$$\sum_{i,j=1}^N c_i c_j K(x_i, x_j) = \sum_{i,j=1}^N c_i c_j \psi(x_i)\psi(x_j) = \left\| \big(c_i \psi(x_i)\big)_{i=1}^N \right\|_2^2 \geqslant 0 \qquad \text{for all } c \in \mathbb{R}^N.$$

■

Consequently, Mercer's theorem provides an orthogonal decomposition of the initial kernel's native space, as proven below.

Lemma 3.34. *Let K be a continuous symmetric positive semi-definite kernel on a compact set Ω and*

$$K(x,y) = \sum_{\ell=1}^\infty \lambda_\ell e_\ell(x) e_\ell(y)$$

its representation given in Theorem 3.30. If we denote $K_\ell(x,y) = \lambda_\ell e_\ell(x) e_\ell(y)$, then

$$\mathcal{H}_K = \left\{ f = \sum_{\ell\in\mathbb{N}} f_\ell \ : \ f_\ell \in \mathcal{H}_{K_\ell} \text{ for } \ell \in \mathbb{N} \text{ and } \sum_{\ell\in\mathbb{N}} \|f_\ell\|_{K_\ell}^2 < \infty \right\},$$

so that the representation $\sum_{\ell\in\mathbb{N}} f_\ell$ of $f \in \mathcal{H}_K$ is uniquely determined, and the inner product is given by

$$\langle f, g\rangle_K = \sum_{\ell\in\mathbb{N}} \langle f_\ell, g_\ell\rangle_{K_\ell},$$

where $\sum_{\ell\in\mathbb{N}} f_\ell$ and $\sum_{\ell\in\mathbb{N}} g_\ell$ are the unique representations of f and $g \in \mathcal{H}_K$.

Proof. We deduce that the components $K_\ell(x,y) = \lambda_\ell e_\ell(x)e_\ell(y)$ given by Mercer's theorem are positive semi-definite kernels on Ω for $\ell \in \mathbb{N}$. This is due to $\lambda_\ell \geqslant 0$, Lemma 2.8, and Lemma 3.33. It enables us to apply Theorem 3.32 and hence decompose the native space $\mathcal{H}_K$ of $K = \sum_{\ell\in\mathbb{N}} K_\ell$ into the infinite sum of $\mathcal{H}_{K_\ell}$ for $\ell \in \mathbb{N}$.

Let us now consider the additional orthogonality requirement of Theorem 3.32. The question is whether

$$\mathcal{H}_{K_I} \cap \mathcal{H}_{K_{\mathbb{N}\setminus I}} = \{0\} \text{ for all index sets } I \subset \mathbb{N},$$

where $K_I := \sum_{\ell\in I} K_\ell$. We recall the definition of a native space for a given kernel K provided in Theorem 2.24:

$$\mathcal{H}_K = \left\{ f \text{ is the pointwise limit of a Chauchy sequence } (s_n)_{N\in\mathbb{N}} \subset S_K \right\}.$$

In our special situation, it is

$$\begin{aligned} S_{K_I} &= \operatorname{span}\left\{K_I(x,\cdot) \ : \ x \in \Omega\right\} \\ &= \operatorname{span}\left\{\sum_{\ell\in I} \lambda_\ell e_\ell(x) e_\ell(\cdot) \ : \ x \in \Omega\right\} \\ &\subset \operatorname{span}\left\{e_\ell \ : \ \ell \in I\right\}. \end{aligned}$$

From pointwise convergence follows L^2 convergence, and we obtain

$$\mathcal{H}_{K_I} \subset \overline{\operatorname{span}\left\{e_\ell \ : \ \ell \in I\right\}}^{L^2} \qquad \text{for all intex sets } I \subset \mathbb{N}.$$

We conclude,

$$\mathcal{H}_{K_I} \cap \mathcal{H}_{K_{\mathbb{N}\setminus I}} \subset \overline{\operatorname{span}\left\{e_\ell \ : \ \ell \in I\right\}}^{L^2} \cap \overline{\operatorname{span}\left\{e_\ell \ : \ \ell \in \mathbb{N}\setminus I\right\}}^{L^2} = \{0\},$$

where the last equation holds by the fact that the set of functions $\{e_\ell\}_{\ell\in\mathbb{N}}$ forms an orthonormal basis of $L^2(\Omega)$.

■

3.3 Interpolation

In the following, we offer a theoretical examination of interpolation using summation kernels. At this juncture, we emphasize that the summation kernel can be viewed as a conventional positive definite kernel (maybe even translation-invariant or radially symmetric), allowing the application of the results from Section 2.4.

Initially, we investigate the relationship between the interpolants corresponding to component kernels and the one corresponding to their summation kernel. We discover that, in terms of approximation error minimization, it is most advantageous to interpolate with a kernel whose native space is just large enough to encompass the target function, as discussed in Section 3.3.1. All interpolants associated with kernels that reproduce a larger native space result in inferior approximations. Additionally, Section 3.3.2 is dedicated to the numerical stability of the interpolation process with summation kernels. We find that the summation kernel exhibits the same level of stability as its most stable component kernel in Theorem 3.40. The analyses in both chapters lead to a trade-off principle between error and stability, which is summarized in Remark 3.42. A smaller kernel yields a better approximation but possesses poorer stability, whereas a larger kernel improves numerical stability at the expense of approximation quality.

From now on we assume the component kernels to be positive definite so that the summation kernel K is also positive definite by Theorem 3.2. Hence, it can uniquely be interpolated as done in Section 2.1. In this section, we omit the subscript X in the notation of the interpolant.

Let the function values $f_X \in \mathbb{R}^N$ be known for a pairwise distinct point set $X = \{x_1, \dots, x_N\} \subseteq \Omega$. Furthermore, let $K = K_1 + K_2$ be the summation kernel of positive definite components K_1 and K_2, then the linear system

$$f_X = \mathbf{A}_{K,X} c = \mathbf{A}_{K_1,X} c + \mathbf{A}_{K_2,X} c$$

has to be solved for $c = (c_1, \dots, c_N) \in \mathbb{R}^N$ to derive the interpolant

$$s_{f,K} = \sum_{i=1}^{N} c_i K(x_i, \cdot) = \sum_{i=1}^{N} c_i K_1(x_i, \cdot) + \sum_{i=1}^{N} c_i K_2(x_i, \cdot),$$

where the right-hand side is a decomposition of $s_{f,K} \in \mathcal{H}_K$ into

$$s^1 := \sum_{i=1}^{N} c_i K_1(x_i, \cdot) \in \mathcal{H}_{K_1} \qquad \text{and} \qquad s^2 := \sum_{i=1}^{N} c_i K_2(x_i, \cdot) \in \mathcal{H}_{K_2}.$$

In the following we are focusing on the case where the target function f lies in $\mathcal{H}_K$ and the kernels K_1 and K_2 correspond to complemented native spaces. This implies that the target function can uniquely be decomposed into $f_1 \in \mathcal{H}_{K_1}$ and $f_2 \in \mathcal{H}_{K_2}$ so that $f = f_1 + f_2$, by Theorem 3.9. In general however, the decomposition of the interpolant $s_{f,K} \in \mathcal{H}_K$ does not correspond to the interpolants $s_{f_1,K_1} \in \mathcal{H}_{K_1}$ and $s_{f_2,K_2} \in \mathcal{H}_{K_2}$ of f_1 and f_2, i.e.,

$$s_{f,K} \neq s_{f_1,K_1} + s_{f_2,K_2}. \tag{3.20}$$

Put differently, the equation

$$\mathbf{A}_{K_1,X} a + \mathbf{A}_{K_2,X} a = \mathbf{A}_{K,X} a = f_X = f_{1X} + f_{2X} = \mathbf{A}_{K_1,X} b + \mathbf{A}_{K_2,X} c \tag{3.21}$$

holds for possibly different coefficient vectors a, b and c. An example is given below.

Example 3.35. Let K_1 and K_2 be positive definite kernels, corresponding to complemented native spaces. Furthermore, let the target function $f = K_1(x_0, \cdot) \in \mathcal{H}_K$ and $K = K_1 + K_2$ be the summation kernel of K_1 and K_2. Additionally, let the data point set be given by $X = \{x_0\}$. Then the interpolant corresponding to the function values f_X and kernel K is given by

$$s_{f,K} = \frac{K_1(x_0, x_0)}{K(x_0, x_0)} K(x_0, \cdot).$$

As $f = K_1(x_0, \cdot)$ we can deduce that the unique decomposition $f_1 + f_2$ of f, where $f_\ell \in \mathcal{H}_{K_\ell}$ for $\ell = 1, 2$, is given by $f_1 = K_1(x_0, \cdot) \in \mathcal{H}_{K_1}$ and $f_2 = 0 \in \mathcal{H}_{K_2}$. Hence,

$$s_{f_1,K_1} = K_1(x_0, \cdot) \quad \text{and} \quad s_{f_2,K_2} = 0$$

are the uniquely defined interpolants in $\mathcal{H}_{K_1}$ and $\mathcal{H}_{K_2}$. Consequently, the coefficient vectors a, b and c of (3.21) are given by $a = \frac{K_1(x_0,x_0)}{K(x_0,x_0)}$, $b = 1$ and $c = 0$. This results in (3.20), i.e.,

$$s_{f,K} = \frac{K_1(x_0, x_0)}{K(x_0, x_0)} K(x_0, \cdot) \neq K_1(x_0, \cdot) = s_{f_1,K_1} + s_{f_2,K_2}.$$

Nevertheless, we can derive the following two relations between the interpolant regarding the summation kernel and the one corresponding to its components.

Lemma 3.36. *Let K_1, K be symmetric positive definite kernels on $\Omega \subseteq \mathbb{R}^d$, such that $K_1 \ll K$. Furthermore, let $X \subseteq \Omega$ be a set of pairwise distinct data points and f_X corresponding function values. Then,*

$$\left\| s_{f,K} \right\|_K \leqslant \left\| s_{f,K_1} \right\|_{K_1}.$$

Proof. Since $s_{f,K}$ has minimal $\mathcal{H}_K$-norm of all functions of $\mathcal{H}_K$ that interpolate f_X by Theorem 2.37, the interpolant s_{f,K_1} lies in $\mathcal{H}_{K_1} \subseteq \mathcal{H}_K$, and it interpolates the function values f_X as well, the first inequality of

$$\left\| s_{f,K} \right\|_K \leqslant \left\| s_{f,K_1} \right\|_K \leqslant \left\| s_{f,K_1} \right\|_{K_1}$$

holds. The second is given by Theorem 3.16. ■

Lemma 3.37. *Let K_1, K_2 be symmetric positive definite kernels on Ω, $K = K_1 + K_2$ their summation kernel and $f_1 + f_2 = f \in \mathcal{H}_K$ the target function, where $f_\ell \in \mathcal{H}_{K_\ell}$ for $\ell = 1, 2$. Then*

$$\left\| s_{f,K} \right\|_K^2 \leqslant \left\| s_{f_1,K_1} \right\|_{K_1}^2 + \left\| s_{f_2,K_2} \right\|_{K_2}^2.$$

Proof. The sum of the interpolants of $s_{f_\ell,K_\ell} \in \mathcal{H}_{K_\ell}$ for $\ell = 1, 2$ is a function in $\mathcal{H}_K$ by Theorem 3.9, and fulfills the interpolation condition for f as

$$\left(s_{f_1,K_1} + s_{f_2,K_2} \right) |_X = s_{f_1,K_1}|_X + s_{f_2,K_2}|_X = f_1|_X + f_2|_X = f|_X.$$

Since $s_{f,K}$ has minimal norm of all functions in $\mathcal{H}_K$ that satisfy the interpolation condition by Theorem 2.37, it is

$$\left\| s_{f,K} \right\|_K^2 \leqslant \left\| s_{f_1,K_1} + s_{f_2,K_2} \right\|_K^2 = \min \left\{ \left\| s^1 \right\|_{K_1}^2 + \left\| s^2 \right\|_{K_2}^2 \right\} \leqslant \left\| s_{f_1,K_1} \right\|_{K_1}^2 + \left\| s_{f_2,K_2} \right\|_{K_2}^2,$$

where the minimum is taken over all representations $s^1 + s^2$, with $s^\ell \in \mathcal{H}_{K_\ell}$ for $\ell = 1, 2$, so that $s_{f_1,K_1} + s_{f_2,K_2} = s^1 + s^2$, see Theorem 3.9. ■

3.3.1 Approximation Error

In Section 2.4.1, we observed that it is advantageous for error estimations if the target function is contained within the native space associated with the kernel used for interpolation. However, it is not sensible, in terms of approximation error, to excessively expand the kernel used for interpolation solely to ensure that the target function is included in the corresponding native space. Theorem 3.38 and the following analysis demonstrates this point. Additionally, we derive an upper bound on the summation kernel's power function in Lemma 3.39.

Example 3.35 given in the above section, hints at a better approximation error if the target function f can be decomposed into f_1 and f_2 and the interpolation is done separately and summed up at the end. In fact, we demonstrate in Theorem 3.38 that a component-wise interpolation leads to a smaller error. Additionally, the theorem shows that an interpolation carried out with a kernel whose native space just contains the target function, but not more, provides the best approximation.

Theorem 3.38. *Let K_1 and K_2 be complemented positive definite component kernels and $K = K_1 + K_2$ their summation kernel. Let $f \in \mathcal{H}_K$ be given by the unique decomposition $f = f_1 + f_2$, where $f_\ell \in \mathcal{H}_{K_\ell}$ for $\ell = 1, 2$. Then,*

$$\left\|f_1 - s_{f_1,K_1}\right\|_K^2 + \left\|f_2 - s_{f_2,K_2}\right\|_K^2 = \left\|f_1 - s_{f_1,K_1}\right\|_{K_1}^2 + \left\|f_2 - s_{f_2,K_2}\right\|_{K_2}^2 \leqslant \left\|f - s_{f,K}\right\|_K^2 .$$

Proof. As $s_{f,K} = \sum c_i K(x_i, \cdot) \in \mathcal{H}_K$, its unique decomposition is given by

$$\sum c_i K_1(x_i, \cdot) + \sum c_i K_2(x_i, \cdot),$$

where $\sum c_i K_\ell(x_i, \cdot) \in \mathcal{H}_{K_\ell}$ for $\ell = 1, 2$. We compute

$$\begin{aligned}
\left\|f_1 - s_{f_1,K_1}\right\|_K^2 + \left\|f_2 - s_{f_2,K_2}\right\|_K^2 &= \left\|f_1 - s_{f_1,K_1}\right\|_{K_1}^2 + \left\|f_2 - s_{f_2,K_2}\right\|_{K_2}^2 \\
&\leqslant \left\|f_1 - \sum c_i K_1(x_i, \cdot)\right\|_{K_1}^2 + \left\|f_2 - \sum c_i K_2(x_i, \cdot)\right\|_{K_2}^2 \\
&= \left\|f_1 + f_2 - \sum c_i K(x_i, \cdot)\right\|_K^2 \\
&= \left\|f - s_{f,K}\right\|_K^2 ,
\end{aligned}$$

where the fist inequality holds since the component kernels are complemented. The second inequality holds as s_{f_ℓ,K_ℓ} is the orthogonal projection of f_ℓ onto $S_{K_\ell,X}$ by Lemma 2.39, and $\sum c_i K_\ell(x_i, \cdot) \in S_{K_\ell,X}$ for $\ell = 1, 2$. The third equality holds by Theorem 3.9. ■

The theorem states, that it is disadvantageous to interpolate in the large space $\mathcal{H}_K$ if the target function f is known to belong to the subspace $\mathcal{H}_{K_1}$. In this case, the unique decomposition of f is given by $f_1 = f \in \mathcal{H}_{K_1}$ and $f_2 = 0 \in \mathcal{H}_{K_2}$. Theorem 3.38 implies

$$\left\|f - s_{f,K_1}\right\|_K = \left\|f - s_{f,K_1}\right\|_{K_1} \leqslant \left\|f - s_{f,K}\right\|_K .$$

This shows, the error can be expected to be smaller when interpolation is carried out with K_1 instead of K. Even more, Theorem 3.38 and the characterization of the power function, given in Definition 2.27, imply

$$\begin{aligned} P_{K_1,X}(x)^2 + P_{K_2,X}(x)^2 &= \left\| K_1(\cdot,x) - s_{K_1(x,\cdot),K_1} \right\|_{K_1}^2 + \left\| K_2(\cdot,x) - s_{K_2(x,\cdot),K_2} \right\|_{K_2}^2 \\ &\leqslant \left\| K(\cdot,x) - s_{K(x,\cdot),K} \right\|_{K}^2 \\ &= P_{K,X}(x)^2. \end{aligned}$$

Let a sequence of point sets $\{X_n\}_{n\in\mathbb{N}}$ be such that $P_{K,X_n} \overset{n\to\infty}{\longrightarrow} 0$, then $P_{K_1,X_n}(x)^2 \longrightarrow 0$ and $P_{K_2,X_n}(x)^2 \longrightarrow 0$. This hits at a faster convergence in smaller spaces, which is well-known in the case of Sobolev spaces but not in the general setting presented here.

Lemma 3.39. *Let K_ℓ be translation-invariant kernels on a bounded set $\Omega \subseteq \mathbb{R}^d$ satisfying an ICC, with univariate functions $\Phi_\ell \in C(\mathbb{R}^d)$ for $\ell = 1, \ldots, M$. Let K be their summation kernel on Ω and $X \subset \Omega$ be a finite pairwise distinct point set satisfying $h_{X,\Omega} \leqslant h_0$. Then,*

$$P_{K,X}^2(x) \leqslant F_{\Phi,\Omega}(h_{X,\Omega}) = \sum_{\ell=1}^{M} F_{\Phi_\ell,\Omega}(h_{X,\Omega}) \qquad \text{for all } x \in \Omega,$$

where the functions $F_{\Phi,\Omega}$ and $F_{\Phi_\ell,\Omega}$ come from Remark 2.44.

Proof. By Lemma 3.3, the summation Kernel K is translation-invariant with univariate function $\Phi = \sum_{\ell=1}^{M} \Phi_\ell$. We apply Theorem 2.43 and Remark 2.44, to obtain

$$P_{K,X}^2(\tilde{x}) \leqslant F_{\Phi,\Omega}(h_{X,\Omega}) = c_1 \sup_{x \in B(0,\, 2c_2 h_{X,\Omega})} |\Phi(x) - p(x)| \qquad \text{for all } \tilde{x} \in \Omega,$$

where p is an arbitrary polynomial from $\pi_m(\mathbb{R}^d)$. We split the right-hand side, so that the following equation holds

$$P_{K,X}^2(\tilde{x}) \leqslant \sum_{\ell=1}^{M} c_1 \sup_{x \in B(0,\, 2c_2 h_{X,\Omega})} |\Phi_\ell(x) - p_\ell(x)| \qquad \text{for all } \tilde{x} \in \Omega,$$

where p_ℓ are arbitrary polynomials from $\pi_m(\mathbb{R}^d)$ and $p = \sum_{\ell=1}^{M} p_\ell$. Since the constants h_0, c_1 and c_2 only depend on the ICC of Ω and not on the kernel, the bound of the power function of the component kernels K_ℓ is given by

$$F_{\Phi_\ell,\Omega}(h_{X,\Omega}) = c_1 \sup_{x \in B(0,\, 2c_2 h_{X,\Omega})} |\Phi_\ell(x) - p_\ell(x)| \qquad \text{for all } \ell = 1, \ldots, M,$$

where p_ℓ is an arbitrary polynomial from $\pi_m(\mathbb{R}^d)$. Combining the two preceding equations, results in the required bound. ∎

3.3.2 Numerical Stability

The stability of an interpolation problem is mirrored by the condition number of the interpolation matrix, as demonstrated in Section 2.4.2. Therefore, we study the smallest and largest eigenvalues of sums of symmetric matrices to establish a bound on the condition number corresponding to a summation kernel.

Theorem 3.40. *Let K_ℓ be positive definite kernels on Ω for $\ell = 1, \dots, M$, K be their summation kernel and $X \subset \Omega$ a pairwise distinct point set. Then*

$$\lambda_{\min}\left(\mathbf{A}_{K,X}\right) \geqslant \max_{\ell \in \{1,\dots,M\}} \left\{\lambda_{\min}\left(\mathbf{A}_{K_\ell,X}\right)\right\} \tag{3.22}$$

and

$$\operatorname{cond}_2(\mathbf{A}_{K,X}) \leqslant \frac{M \max\limits_{l \in \{1,\dots,M\}} \{\lambda_{\max}(\mathbf{A}_{K_\ell,X})\}}{\max\limits_{\ell \in \{1,\dots,M\}} \{\lambda_{\min}(\mathbf{A}_{K_\ell,X})\}}.$$

Additionally, let $\ell_0 \in \{1, \dots, M\}$ satisfy

$$\lambda_{\min}\left(\mathbf{A}_{K_{\ell_0},X}\right) = \max_{\ell \in \{1,\dots,M\}} \left\{\lambda_{\min}\left(\mathbf{A}_{K_\ell,X}\right)\right\}$$

and K_{ℓ_0} be a translation-invariant kernel with univariate function Φ_{ℓ_0} such that $\widehat{\Phi_{\ell_0}} \in C(\mathbb{R}^d \backslash \{0\})$. Then

$$\lambda_{\min}\left(\mathbf{A}_{K,X}\right) \geqslant G_{\Phi_{\ell_0}}(q_X),$$

where the function $G_{\Phi_{\ell_0}}$ comes from Remark 2.49.

Proof. Let $c \in \mathbb{R}^{|X|}$. As $\mathbf{A}_{K,X} = \sum_{\ell=1}^M \mathbf{A}_{K_\ell,X}$, we compute

$$\left\langle c, \mathbf{A}_{K,X} c\right\rangle = \sum_{\ell=1}^{M} \left\langle c, \mathbf{A}_{K_\ell,X} c\right\rangle \geqslant \sum_{\ell=1}^{M} \lambda_{\min}\left(\mathbf{A}_{K_\ell,X}\right) \|c\|^2 .$$

Because the matrix $\mathbf{A}_{K_\ell,X}$ is positive definite for all $\ell = 1, \dots, M$, we can deduce (3.22) as a lower bound on the minimal eigenvalue

$$\lambda_{\min}\left(\mathbf{A}_{K,X}\right) \geqslant \sum_{\ell=1}^{M} \lambda_{\min}\left(\mathbf{A}_{K_\ell,X}\right) \geqslant \max_{\ell \in \{1,\dots,M\}} \left\{\lambda_{\min}\left(\mathbf{A}_{K_\ell,X}\right)\right\}.$$

To obtain an upper bound on the maximal eigenvalue $\lambda_{\max}\left(\mathbf{A}_{K,X}\right)$, we compute

$$\left\langle c, \mathbf{A}_{K,X} c\right\rangle = \sum_{\ell=1}^{M} \left\langle c, \mathbf{A}_{K_\ell,X} c\right\rangle \leqslant \sum_{\ell=1}^{M} \lambda_{\max}\left(\mathbf{A}_{K_\ell,X}\right) \|c\|^2 .$$

This implies

$$\lambda_{\max}\left(\mathbf{A}_{K,X}\right) \leqslant \sum_{\ell=1}^{M} \lambda_{\max}\left(\mathbf{A}_{K_\ell,X}\right) \leqslant M \max_{\ell \in \{1,\dots,M\}} \{\lambda_{\max}\left(\mathbf{A}_{K_\ell,X}\right)\}. \tag{3.23}$$

The equations (3.22) and (3.23) yield

$$\operatorname{cond}_2(\mathbf{A}_{K,X}) = \frac{\lambda_{\max}(\mathbf{A}_{K,X})}{\lambda_{\min}(\mathbf{A}_{K,X})} \leqslant \frac{M \max\limits_{\ell \in \{1,\ldots,M\}} \left\{\lambda_{\max}(\mathbf{A}_{K_\ell,X})\right\}}{\max\limits_{\ell \in \{1,\ldots,M\}} \left\{\lambda_{\min}(\mathbf{A}_{K_\ell,X})\right\}}.$$

Let us consider the additional statement. As the component kernel K_{ℓ_0} satisfies the requirements out of Theorem 2.48, there exist, by Remark 2.49, a function $G_{\Phi_{\ell_0}}$, so that $\lambda_{\min}(\mathbf{A}_{K_{\ell_0},X}) \geqslant G_{\Phi_{\ell_0}}(q_X)$. With equation (3.22) and the fact that $\mathbf{A}_{K,X} = \sum_{\ell=1}^{M} \mathbf{A}_{K_\ell,X}$, it is

$$\lambda_{\min}(\mathbf{A}_{K,X}) \geqslant \max_{\ell \in \{1,\ldots,M\}} \left\{\lambda_{\min}(\mathbf{A}_{K_\ell,X})\right\} = \lambda_{\min}\left(\mathbf{A}_{K_{\ell_0},X}\right) \geqslant G_{\Phi_{\ell_0}}(q_X).$$

■

Numerical tests indicate that the maximum eigenvalue does not pose problems. Rather, it is the minimal eigenvalue that causes issues. The above theorem asserts that the stability of a summation kernel aligns with that of its most stable component kernel. Numerical examples in Section 3.4 confirm this finding.

The subsequent theorem demonstrates that the interpolation method, employing a kernel with a larger native space, is expected to exhibit greater stability compared to interpolation with a kernel reproducing to a smaller space.

Theorem 3.41. *Let $K_1 \lesssim K$, then there exist a constant $c > 0$ such that*

$$\lambda_{\min}(\mathbf{A}_{K_1,X}) \leqslant c\lambda_{\min}(\mathbf{A}_{K,X})$$

for every finite and pairwise distinct point set X.

Proof. Since $K_1 \lesssim K$, there is a constant $c > 0$ so that $cK - K_1$ is a symmetric positive semi-definite kernel. Hence, cK is the summation kernel of K_1 and $cK - K_1$. Theorem 3.40 yields

$$\begin{aligned} c\lambda_{\min}(\mathbf{A}_{K,X}) &= \lambda_{\min}(c\mathbf{A}_{K,X}) = \lambda_{\min}(\mathbf{A}_{cK,X}) \\ &\geqslant \max\left\{\lambda_{\min}(\mathbf{A}_{K_1,X}), \lambda_{\min}(\mathbf{A}_{cK-K_1,X})\right\} \geqslant \lambda_{\min}(\mathbf{A}_{K_1,X}). \end{aligned}$$

■

The preceding theorem hints at another trade-off principle between numerical stability and approximation error, additionally to the one discussed in Section 2.4.3. For the best numerical stability, the above Theorem 3.41 suggests to use a kernel reproducing a large native space, whereas the analysis below Theorem 3.38 emphasizes a better approximation quality using a small kernel. This trade-off is simplified in Remark 3.42 below.

Remark 3.42. If $K_1 \lesssim K_2$ and the target function lies in $\mathcal{H}_{K_1}$, then

	K_1	K_2
Stability	bad	good
Approximation	good	bad

3.4 Numerical Tests

This section aims to numerically underline the theoretical results from the previous sections. Theorem 3.25 shows, if the kernels K_1 and K_2 satisfy the relation $K_2 \lesssim K_1$, their summation kernel K lies in one equivalence class with K_1. As a consequence, the native spaces $\mathcal{H}_{K_1}$ and $\mathcal{H}_K$ encompass the same space, equipped with equivalent norms. The subsequent provides an example of how the equivalence class of a kernel affects the interpolation method.

Here, we use

- the Wendland kernel K_1 with RBF $\phi_{3,0}$,
- the Wendland kernel K_2 with RBF $\phi_{3,3}$, and
- the summation kernel K of K_1 and K_2

for interpolation, where $\phi_{d,k}$ is defined in Theorem 2.34. The three kernels K_1, K_2 and K are visualized in Fig. 3.1. By Corollary 3.18, the kernels K_1 and K_2 satisfy $K_2 \lesssim K_1$. This implies $K \sim K_1$, by Theorem 3.25. We expect similar behavior from the interpolants and interpolation matrices of these two kernels. To determine this, we perform the interpolation on

- the domain $\Omega = [0, 1]^2$ and
- the developing point sets X_n consisting of 2^n, $n = 6, \ldots, 11$, random points in Ω, satisfying $X_m \subseteq X_n$ for $m \leqslant n$.

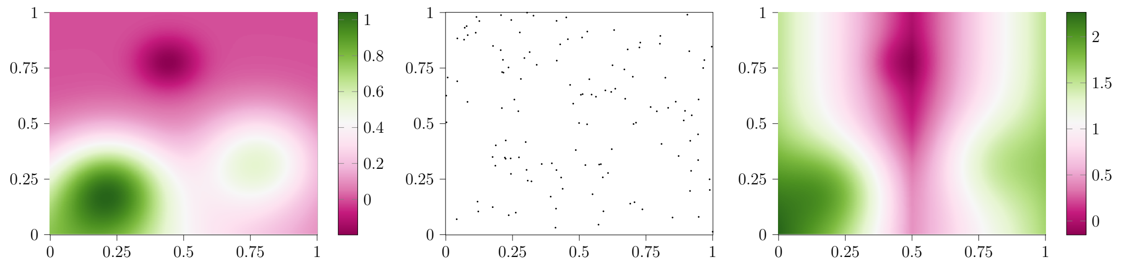

Figure 3.2: Top view of the target functions f_{franke} (left) and f_{kink} (right), and the scattered point set X_7 in Ω (middle).

In this setting, we compare two target functions

1. a C^∞ target function. We have chosen the well-known Franke function

$$
\begin{aligned}
f_{\text{franke}}(x, y) &:= 0.75 \exp\left(-\frac{(9x-2)^2 + (9y-2)^2}{4}\right) \\
&+ 0.75 \exp\left(-\frac{(9x+1)^2}{49} - \frac{(9y+1)^2}{10}\right) \\
&+ 0.5 \exp\left(-\frac{(9x-7)^2 + (9y-3)^2}{4}\right) \\
&- 0.2 \exp\left(-(9x-4)^2 - (9y-7)^2\right),
\end{aligned}
\tag{3.24}
$$

   which was first used by R. Franke in [Fra79] and is visualized in Fig. 3.2 (left). Since then, it has been widely used for the analysis of reconstruction methods (cf. [Fra82], [Mül09], [BLRS15])

2. a C^0 target function. Here, we chose the function

$$f_{\text{kink}}(x, y) := f_{\text{franke}}(x, y) + 3\|x - 0.5\|$$

   that exhibits a kink along $x = 0.5$, visualized in Fig. 3.2 (right).

We test with two target functions to illustrate the independence of the results on the underlying target. The target functions f_{franke} and f_{kink} together with the interpolation point set X_7 are visualized in Fig. 3.2.

In Fig. 3.3 (left) and (right), we observe the interpolant of the summation kernel K to behave as the interpolant of K_1 regarding the approximation error development for both target functions f_{franke} and f_{kink}. Furthermore, the numerical condition number regarding K develops with the same rate as the one of K_1, see Fig. 3.3 (middle). This can be explained with (3.22) of Theorem 3.40.

Fig. 3.3 (left) and (middle) support the trade-off principal of Remark 3.42. As the target function f_{franke} lies in the small space C^∞, a smaller error is achieved using the kernel K_2 that spans the smaller native space $\mathcal{H}_{K_2}$ (compared to $\mathcal{H}_{K_1}$ and $\mathcal{H}_K$), while still containing C^∞. Whereas the condition number corresponding to K_2 is poor compared to the one of K_1 and K. However, Fig. 3.3 (right) exemplifies, that the approximation error cannot be improved using the kernel K_2 corresponding to a small native space $\mathcal{H}_{K_2}$, if the target function $f_{\text{kink}} \in C^0$ is not contained in that native space.

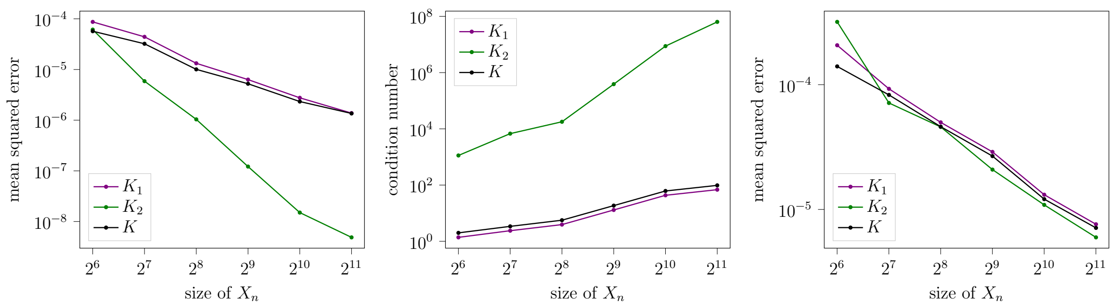

Figure 3.3: Comparison of the component kernels K_j for $j = 1, 2$, and their summation kernel K regarding the condition number of the interpolation matrix (middle) and the mean squared error corresponding to the target functions f_{franke} (left) and f_{kink} (right) on developing point sets X_n for $n = 6, \ldots, 11$.

We summarize two core statements:

1. The summation kernel of two kernels, whose native spaces are subspaces of another, behaves like the kernel of the larger native space.

2. The trade-off principle of kernels: Provided, the target function lies in the native space of a kernel K, this kernel yields better approximation, but worse stability compared to a kernel whose native space is containing the one of K.

Chapter 4

Product Kernels

To the best of our knowledge, explicit investigations into *product kernels* have been sparse. The foundational work [Aro50] by N. Aronszajn, frequently cited in this thesis already, stands as a notable exception. Typically, the fact that the product of two positive (semi-)definite kernels results in a kernel that is again positive (semi-)definite is mentioned only briefly, see for example [Wen05, Theorem 6.2] or [SC08, Lemma 4.6]. This cursory treatment in research is not without reason. Product kernels are a special case of tensor product kernels and inherit the underlying structure of these more general kernels. We devote Chapter 7 to the detailed exploration of these interesting kernels.

Despite this, we introduce product kernels here for completeness and to align with the structure of this thesis. Specifically, we

- connect the interpolation matrix of a product kernel with the Hadamard product.
- analyze the interpolation method using product kernels.

By doing so, we provide the groundwork for discovering possible previously unexplored advantages of product kernels.

This chapter is organized as follows: In Section 4.1, we provide a precise definition of the product kernel and derive its basic properties. In the subsequent sections on native space (Section 4.2) and interpolation (Section 4.3), we demonstrate that we are essentially dealing with a restriction of tensor product kernels.

4.1 Definition and Basic Properties

This section provides a precise definition of the product kernel, and states basics findings regarding positive definiteness, translation-invariance, and radial symmetry.

Definition 4.1. Let $K_\ell : \Omega \times \Omega \longrightarrow \mathbb{R}$, $\ell = 1, \dots, M$, then

$$\begin{aligned} K : \Omega \times \Omega &\longrightarrow \mathbb{R}, \\ K(x, y) &= \prod_{\ell=1}^{M} K_\ell(x, y) \ \text{ for } x, y \in \Omega \end{aligned}$$

is called a *product kernel*.

A product kernel with two component kernels is visualized in Fig. 4.1.

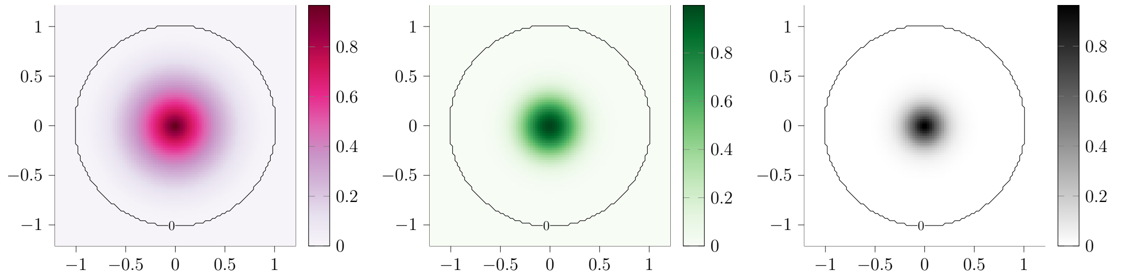

Figure 4.1: Top view of two two-dimensional component kernels resulting from the Wendland RBFs $\phi_{3,0}$ (left) and $\phi_{3,3}$ (middle), and the corresponding two-dimensional product kernel (right). The black line visualizes supports.

Let us consider the interpolation matrix of a product kernel. To do so, let $X = \{x_1, ..., x_N\} \subset \Omega$ be a set of points and K a product kernel with components K_ℓ for $\ell = 1, \ldots, M$. Then, the entries of the interpolation matrix $\mathbf{A}_{K,X}$ can be written as

$$\left(\mathbf{A}_{K,X}\right)_{j,k} = K(x_j, x_k) = \prod_{\ell=1}^{M} K_\ell(x_j, x_k) = \prod_{\ell=1}^{M} \left(\mathbf{A}_{K_\ell,X}\right)_{j,k}.$$

This representation leads us to the Hadamard product, also called Schur product, see [HJ91, Definition 5.0.1].

Definition 4.2. Let $A, B \in \mathbb{R}^{m\times n}$ be two matrices of the same size. The *Hadamard product* $A \odot B$ of A and B is given by

$$(A \odot B)_{j,k} = (A)_{j,k} \cdot (B)_{j,k},$$

where $j = 1, \ldots, m$ and $k = 1, \ldots, n$.

With this definition at hand, the interpolation matrix $\mathbf{A}_{K,X}$ of a product kernel K equals the Hadamard product of the interpolation matrices $\mathbf{A}_{K_\ell,X}$ corresponding to its components K_ℓ, i.e.,

$$\mathbf{A}_{K,X} = \bigodot_{\ell=1}^{M} \mathbf{A}_{K_\ell,X}. \tag{4.1}$$

This representation implies the following result, that can also be found in [Aro50].

Theorem 4.3. *Let K_ℓ be kernels on Ω for $\ell = 1, \ldots, M$.*

(i) If K_ℓ is positive semi-definite for all $\ell = 1, \ldots, M$, their product kernel K is positive semi-definite.

(ii) If K_ℓ is positive definite for all $\ell = 1, \ldots, M$, their product kernel K is positive definite.

We shortly remark that Theorem 4.3 *(ii)* relies on the Schur product theorem, and refer to [HJ91, Chapter 5.2] for details and proofs.

We note, that the set of positive semi-definite kernels equipped with the product can be viewed as a commutative semigroup. Here, the neutral element is given by the constant kernel $K(x,y) = 1$ for all x, y. The set of $m \times n$ matrices with nonzero entries, however, form a commutative group under the Hadamard product. We cannot deduce the same structure for kernels, as the inverse of the Hadamard product regarding a positive semi-definite matrix is not proven to be positive semi-definite again. We refer to [Rea99] for more information.

We close this section with considering translation-invariance and radial symmetry of product kernels.

Lemma 4.4. *Let K_ℓ be kernels on Ω for $\ell = 1, \dots, M$.*

(i) If K_ℓ are translation-invariant kernels with uni-variate functions Φ_ℓ for $\ell = 1, \dots, M$, their product kernel K is translation invariant with the uni-variate function $\Phi = \prod_{\ell=1}^{M} \Phi_\ell$.

(ii) If K_ℓ are radially symmetric kernels with RBF ϕ_ℓ for $\ell = 1, \dots, M$, their product kernel K is a radial kernel with RBF $\phi = \prod_{\ell=1}^{M} \phi_\ell$.

Proof. The subsequent equations prove the required statements:

(i) $K(x,y) = \prod_{\ell=1}^{M} K_\ell(x,y) = \prod_{\ell=1}^{M} \Phi_\ell(x-y)$ for all $x, y \in \Omega$.

(ii) $K(x,y) = \prod_{\ell=1}^{M} K_\ell(x,y) = \prod_{\ell=1}^{M} \phi_\ell(\|x-y\|_2)$ for all $x, y \in \Omega$.

■

4.2 Native Spaces

As we consider functions and kernels on the domain $\Omega \subseteq \mathbb{R}^d$ only, we omit the notation Ω in native spaces and norm notations for simplicity.

In this section, we acknowledge the contributions of N. Aronszajn, who was the first to study RKHSs of product kernels with two component kernels K_1 and K_2 in [Aro50, Part 1, §8]. To do so, he constructed the space

$$\mathcal{H}_{K_1} \otimes \mathcal{H}_{K_2} := \left\{ f'(x_1, x_2) = \sum_{i=1}^{N} f_1^i(x_1) f_2^i(x_2) \ : \ f_\ell^i \in \mathcal{H}_{K_\ell,\Omega},\ \ell = 1,2 \text{ and } j = 1, \dots, N \right\}$$

of functions on the Cartesian product $\Omega \times \Omega$ with the inner product

$$\left\langle f', g' \right\rangle_{\mathcal{H}_{K_1} \otimes \mathcal{H}_{K_2}} = \sum_{i=1}^{N} \sum_{j=1}^{M} \left\langle f_1^i, g_1^j \right\rangle_{K_1} \left\langle f_2^i, g_2^j \right\rangle_{K_2},$$

where M is the number of terms in the representation of g'. He needed to demonstrate the completeness of $\mathcal{H}_{K_1} \otimes \mathcal{H}_{K_2}$ elaborately, to gain that $\mathcal{H}_{K_1} \otimes \mathcal{H}_{K_2}$ possesses the reproducing kernel $K_1(x_1, y_1)K_2(x_2, y_2)$. In [Aro50, Part 1, §8, Theorem II], he concluded that the RKHS of a product kernel is given by the restriction of $\mathcal{H}_{K_1} \otimes \mathcal{H}_{K_2}$ to the diagonal $D = \{(x, x) : x \in \Omega\} \subset \Omega \times \Omega$. We state this result.

Theorem 4.5. *Let K be the product kernel of K_1 and K_2. Then K is the reproducing kernel of*

$$\mathcal{H}_K = \left\{ f|_D \ : \ f \in \mathcal{H}_{K_1} \otimes \mathcal{H}_{K_2} \right\},$$

where $D = \{(x, x) : x \in \Omega\} \subset \Omega \times \Omega$. For any $f \in \mathcal{H}_K$,

$$\|f\|_K = \min \|g\|_{\mathcal{H}_{K_1} \otimes \mathcal{H}_{K_2}},$$

where the minimum is taken over all $g \in \mathcal{H}_{K_1} \otimes \mathcal{H}_{K_2}$ such that $g|_D = f$.

Today, we recognize the space $\mathcal{H}_{K_1} \otimes \mathcal{H}_{K_2}$ as a Hilbert tensor product. We elaborate on this specific tensor product in Section 7.2, highlight N. Aronszajn's results in this context, and describe the relationship between it and tensor product kernels. Indeed, the product kernel can be considered a special case of the tensor product kernel.

4.3 Interpolation

In this subsection, we briefly examine the native space norm of an interpolant corresponding to a product kernel and provide specific bounds on the minimal eigenvalue and condition number. This analysis further demonstrates that the product kernel should be understood as a restriction of the tensor product kernels. Therefore, we occasionally reference concepts from the forthcoming Chapter 7 in this section. Nevertheless, we want to emphasize at this point that the product kernel can be considered as a 'normal' positive definite (translation-invariant or radially symmetric) kernel, and thus, results from Section 2.4 can be applied.

Lemma 4.6. *Let K be the product kernel of positive definite component kernels K_ℓ on Ω for $\ell = 1, \dots, M$ and $X \subset \Omega$ a pairwise distinct point set. Then, for any function values f_X, there exists a unique interpolant $s_{f,K} \in S_{K,X}$.*
Additionally, if the target function has the form $f = \prod_{\ell=1}^M f_\ell$, it is

$$\left\| s_{f,K} \right\|_{K,\Omega} \leqslant \prod_{l=1}^{M} \left\| s_{f_\ell,K_\ell} \right\|_{K_\ell,\Omega}.$$

Proof. The first part is a direct consequence of Theorem 4.3. Regarding the second part, the interpolants s_{f_ℓ,K_ℓ} are uniquely defined. Furthermore,

$$f(x_i) = \prod_{\ell=1}^{M} f_\ell(x_i) = \prod_{\ell=1}^{M} s_{f_\ell,X,K_\ell}(x_i), \qquad \text{for all } x_i \in X.$$

Consequently, $\prod_{\ell=1}^M s_{f_\ell,K_\ell}$ satisfies the interpolation condition and lies in $\mathcal{H}_{K,\Omega}$ by Theorem 4.5. The optimality statement given in Theorem 2.38 implies the required. ∎

We proceed with deriving stability estimates for the product kernel. Even though these estimates are quite poor, we include them here for the sake of completeness. The eigenvalues of principal submatrices (Definition 2.6) of hermitian matrices can be estimated by the Cauchy interlacing theorem, see [HJ91, Corollary 3.1.3].

Lemma 4.7. *Suppose $A \in \mathbb{R}^{N\times N}$ is a hermitian matrix with eigenvalues of increasing order $\lambda_1(A), \ldots, \lambda_N(A)$. Let $A_r \in \mathbb{R}^{N-r\times N-r}$ denote a principal submatrix of A obtained by deleting a total of r rows and columns from A. Then*

$$\lambda_{\min}(A) = \lambda_1(A) \leqslant \lambda_{\min}(A_r) \leqslant \lambda_{r+1}(A)$$

and

$$\lambda_{N-r}(A) \leqslant \lambda_{\max}(A_r) \leqslant \lambda_N(A) = \lambda_{\max}(A).$$

The above relations enable us to bound the minimal eigenvalue of the product kernel's interpolation matrix and its numerical condition number. To achieve this, we exploit the fact that the Hadamard product (Definition 4.2) of two matrices is a principal submatrix of the Kronecker product (Definition 7.13) of these matrices. At this point, we want to anticipate, as already done in Section 4.2 for the native space, that we are leveraging the overarching structure of tensor product kernels of Chapter 7, whose interpolation matrices can be expressed as Kronecker products. The inherent structure of tensor product kernels allows to derive stronger results. We state a weakened form of Theorem 7.27 here.

Theorem 4.8. *Let K be a product kernel with positive definite component kernels K_ℓ on Ω for $\ell = 1, \ldots, M$ and $X \subseteq \Omega$ a pairwise distinct data set. Then*

$$\lambda_{\min}\left(\mathbf{A}_{K,X}\right) \geqslant \prod_{\ell=1}^{M} \lambda_{\min}\left(\mathbf{A}_{K_\ell,X}\right)$$

and

$$\operatorname{cond}_2\left(\mathbf{A}_{K,X}\right) \leqslant \prod_{\ell=1}^{M} \operatorname{cond}_2\left(\mathbf{A}_{K_\ell,X}\right).$$

Additionally, let K_ℓ be translation-invariant with a univariate function Φ_ℓ satisfying $\widehat{\Phi_\ell} \in C(\mathbb{R}^d\backslash\{0\})$ for $\ell = 1, \ldots, M$, then

$$\lambda_{\min}\left(\mathbf{A}_{K,X}\right) \geqslant \prod_{\ell=1}^{M} G_{\Phi_\ell}(q_X),$$

where the functions G_{Φ_ℓ} come from Remark 2.49.

Proof. We recall (4.1) to see that the interpolation matrix of the product kernel K is given by the Hadamard product

$$\mathbf{A}_{K,X} = \bigodot_{\ell=1}^{M} \mathbf{A}_{K_\ell,X}.$$

The fact that the Hadamard product $\bigodot_{\ell=1}^{M} \mathbf{A}_{K_\ell,X} \in \mathbb{R}^{N\times N}$ is a principal submatrix of the Kronecker product $\bigotimes_{\ell=1}^{M} \mathbf{A}_{K_\ell,X} \in \mathbb{R}^{MN\times MN}$ combined with Lemma 4.7 and Lemma 7.14 *(iv)* implies

$$\prod_{\ell=1}^{M} \lambda_{\min}\left(\mathbf{A}_{K_\ell,X}\right) = \lambda_{\min}\left(\bigotimes_{\ell=1}^{M} \mathbf{A}_{K_\ell,X}\right) \leqslant \lambda_{\min}\left(\mathbf{A}_{K,X}\right),$$

which provides the lower bound of $\lambda_{\min}\left(\mathbf{A}_{K,X}\right)$. Furthermore,

$$\lambda_{\max}\left(\mathbf{A}_{K,X}\right) \leqslant \lambda_{\max}\left(\bigotimes_{\ell=1}^{M} \mathbf{A}_{K_\ell,X}\right) = \prod_{\ell=1}^{M} \lambda_{\max}\left(\mathbf{A}_{K_\ell,X}\right).$$

Regarding the numerical condition number we deduce

$$\operatorname{cond}_2\left(\mathbf{A}_{K,X}\right) = \frac{\lambda_{\max}\left(\mathbf{A}_{K,X}\right)}{\lambda_{\min}\left(\mathbf{A}_{K,X}\right)} \leqslant \frac{\prod_{\ell=1}^{M} \lambda_{\max}\left(\mathbf{A}_{K_\ell,X}\right)}{\prod_{\ell=1}^{M} \lambda\mathrm{min}\left(\mathbf{A}_{K_\ell,X}\right)} = \prod_{\ell=1}^{M} \operatorname{cond}_2\left(\mathbf{A}_{K_\ell,X}\right).$$

For the additional statement, we recall Theorem 2.48 and Remark 2.49 to obtain the existence of a function G_{Φ_ℓ} so that

$$\lambda_{\min}(\mathbf{A}_{K_\ell,X}) \geqslant G_{\Phi_\ell}(q_X) \qquad \text{for every } \ell = 1, \ldots, M.$$

This combined with the lower bound on the minimal eigenvalue of $\mathbf{A}_{K,X}$ yields the required result. ■

Part III

Anisotropic Kernels

Chapter 5

Transformation Kernels

In the realm of kernel-based interpolations, the concept of shape parameters holds pivotal significance, serving as a cornerstone in refining interpolation techniques. Here, kernels on $\mathbb{R}^d$ are scaled by a shape parameter $\alpha > 0$ resulting in a new kernel

$$K_\alpha(x, y) = K(\alpha x, \alpha y).$$

The choice of the parameter α is a critical issue as it affects the concentration of the basis functions around the respective interpolation point. While a small parameter increases the condition number of the interpolation matrix ([Fas07, Chapter 16.2]), a large parameter turns the basis functions into sharp peaks, that approximate functions badly, if interpolation points are widely scattered. Due to this significant impact of the parameter α on the interpolation process, numerous optimization and search strategies have been investigated for its fine-tuning over the last 30 years, see e.g. [KC92], [LF05], [MVHÖ23] and the references therein. A review of different techniques can be found in [FM15, Chapter 14].

Building upon this groundwork, a notable progression emerges with the advent of anisotropic kernels. Here, a diagonal matrix D combined with a rotation matrix U comes into play, building the kernel

$$K_{DU}(x, y) = K(DUx, DUy).$$

These kernels, named after their directional sensitivity, mark a significant departure from their isotropic versions, the kernels with shape parameter, where $D = \operatorname{diag}(\alpha, \ldots, \alpha)$ and $U = Id$. They offer a tailored approach to interpolation that aligns with the inherent anisotropy present either within the distribution of interpolation points or the underlying target function. The research of [CLMM06], [AD14] and [LMZ+24] presents methodologies for attaining a problem-adapted diagonal matrix. It is demonstrated that for anisotropic datasets or target functions exhibiting anisotropic behavior, anisotropic kernels can be used to improve the numerical stability and accuracy of the interpolant (cf. [BDL10]). In the context of radial basis functions, this introduces an alternative metric distinct from the Euclidean norm. Specifically, we define the norm as $\|x\|_B = x^T Bx$, where B is a symmetric positive definite matrix given by $B = U^T D^2 U$.

Moreover, the recent research of [WMP24] has transcended traditional paradigms by transforming the full-rank matrix DU associated with anisotropic kernels into a low-rank matrix. The suggested approach facilitates a seamless transition from higher-dimensional spaces to lower-dimensional ones, offering a novel approach for dimensionality reduction without compromising interpolation quality. This is particularly effective, when the target function inherits anisotropic behavior, that is different directions are unequally relevant. The authors present a way of finding such a low-rank matrix to transform the kernel by using machine learning methods. Such innovative methodologies challenge conventional notions and highlight the inherent flexibility and adaptability embedded within kernel-based interpolation frameworks.

Furthermore, the exploration of variably-scaled kernels, as exemplified in [BLRS15], introduces yet another dimension to the discourse on transforming kernels. Departing from the confines of conventional matrix representations, variably-scaled kernels offer an alternative, further expanding the repertoire of tools available for crafting bespoke interpolation strategies tailored to specific problems.

We combine these diverse approaches under the umbrella term of 'transformation kernels'. These are kernels given by the composition of a kernel K and a transformation T, i.e.

$$\Omega \quad \xrightarrow{T} \quad T(\Omega) \quad \xrightarrow{K} \quad \mathbb{R}$$

resulting in the *transformation kernel*

$$K_T(x, y) := K(Tx, Ty).$$

While preceding research has laid a solid groundwork, a comprehensive and nuanced analysis of the overarching transformation kernel remains absent. Thus, the following study intends to bridge this gap, where the main contributions are:

- Examination of Transformation Kernels Native Space: We investigate the transformation kernels native space in Theorem 5.6. This gains importance in Chapter 6 and Chapter 7, where anisotropic versions of product and summation kernels are considered.

- Detailed Analysis of Interpolation with Transformation Kernel: We systematically explore the interpolation conducted with a transformation kernel in Section 5.3, where we reveal underlying principles governing transformation kernels and shed light on their potential applications improving approximation error and/or numerical stability.

By providing a structured framework, we aim at a comprehensive understanding, and by that a suitable usage of transformation kernels in interpolation methods. We thereby contribute to the ongoing discourse on kernel-based interpolations and pave the ground for future advancements in the field, such as the upcoming anisotropic kernels presented in Chapter 6 and Chapter 7.

The subsequent is structured as follows: First, we offer a definition of transformation kernels, along with a presentation of basic findings (Section 5.1). Following this, we delve into an exploration of the native space of transformation kernels (Section 5.2), shedding light on the intricate relationship between these spaces and those of the underlying initial kernels. Subsequently, we turn our attention to the interpolation using transformation kernels, with a particular focus on approximation error and numerical stability (Section 5.3). Finally, we present numerical examples of an adaptation to interpolation points and target function, demonstrating the efficacy of transformation kernels in improving approximation error and/or numerical stability (Section 5.4).

5.1 Definition and Basic Properties

In the following, we present a precise definition of transformation kernels, derive requirements for positive (semi-)definiteness and translation-invariance, and allocate the final part of this section to the special case of radial kernels in conjunction with a linear transformation.

Definition 5.1. Let $T : \Omega \longrightarrow T(\Omega) \subseteq \mathbb{R}^d$ and $K : T(\Omega) \times T(\Omega) \longrightarrow \mathbb{R}$. Then

$$\begin{aligned} K_T &: \Omega \times \Omega \longrightarrow \mathbb{R}, \\ K_T(x, y) &:= K(Tx, Ty) \;\text{ for } x, y \in \Omega \end{aligned}$$

is called a *transformation kernel* with transformation T.

The interpolation matrix $\mathbf{A}_{K_T,X}$ of the transformation kernel K_T equals the interpolation matrix $\mathbf{A}_{K,T(X)}$ of the initial kernel K evaluated at the transformed point set $T(X)$, i.e.,

$$\mathbf{A}_{K_T,X} = \left(K_T(x_i, x_j)\right)_{i,j} = \Big(K\left(T(x_i), T(x_j)\right)\Big)_{i,j} = \mathbf{A}_{K,T(X)}. \tag{5.1}$$

This yields the subsequent characterizations of positive (semi-)definite transformation kernels.

Theorem 5.2. *Let $T : \Omega \longrightarrow T(\Omega)$ and K be a kernel acting on $T(\Omega)$.*

(i) K_T is positive semi-definite on Ω if and only if K is positive semi-definite on $T(\Omega)$.

(ii) K_T is positive definite on Ω if and only if K is positive definite on $T(\Omega)$ and T is injective.

Proof. The first statement (i) is a direct consequence of (5.1). In (ii), the injectivity of T ensures that $T(X)$ is pairwise distinct if and only if X is pairwise distinct. Then, the equality of (5.1) yields the required statement. ∎

In Remark 2.12, we introduced the important set of translation-invariant kernels.

Lemma 5.3. *Let $T : \mathbb{R}^d \longrightarrow T(\Omega)$ and K be a kernel acting on $T(\Omega)$.*

(i) If the transformation T is translation-invariant, the transformation kernel K_T is translation-invariant.

(ii) If the transformation T is linear and the kernel K is translation-invariant, the transformation kernel K_T is translation-invariant. The corresponding univariate functions Φ and Φ_T satisfy the relation $\Phi_T = \Phi \circ T$.

Proof. The first statement (i) holds as

$$\begin{aligned} K_T(x-\xi, y-\xi) &= K\big(T(x-\xi), T(y-\xi)\big) \\ &= K\big(T(x), T(y)\big) = K_T(x,y) \qquad \text{for all } x, y, \xi \in \mathbb{R}^d. \end{aligned}$$

To show the second statement (ii), we compute

$$\begin{aligned} K_T(x-\xi, y-\xi) &= K\big(T(x-\xi), T(y-\xi)\big) \\ &= K\big(T(x)-T(\xi), T(y)-T(\xi)\big) \\ &= K\big(T(x), T(y)\big) = K_T(x,y) \qquad \text{for all } x, y, \xi \in \mathbb{R}^d. \end{aligned}$$

Furthermore,

$$\begin{aligned} \Phi_T(x-y) = K_T(x,y) &= K\big(T(x), T(y)\big) \\ &= \Phi\big(T(x)-T(y)\big) = \Phi \circ T(x-y) \qquad \text{for all } x, y \in \mathbb{R}^d. \end{aligned}$$

■

The subsequent is concerned with an even smaller subset of kernels, namely radially symmetric kernels. As pointed out in Remark 2.14, such kernels K are build by a radial basis function (RBF) $\phi : \mathbb{R}_+ \longrightarrow \mathbb{R}$, i.e.,

$$K(x,y) = \phi(\|x-y\|_2) \qquad \text{for all } x, y \in \Omega.$$

Lemma 5.4. *Let the transformation T be linear and radially symmetric, i.e., $T(x) = T(\|x\|_2\, e_1)$ for all $x \in \Omega$, and the kernel K be translation-invariant with univariate function Φ. Then, the transformation kernel K_T is a radial kernel with the RBF ϕ_T given by $\phi_T(\|\cdot\|_2) = \Phi \circ T$.*

Proof. The statement holds as

$$K_T(x,y) = K\big(T(x), T(y)\big) = \Phi\big(T\|x-y\|_2\, e_1\big) := \phi_T(\|x-y\|_2) \qquad \text{for all } x, y \in \Omega.$$

■

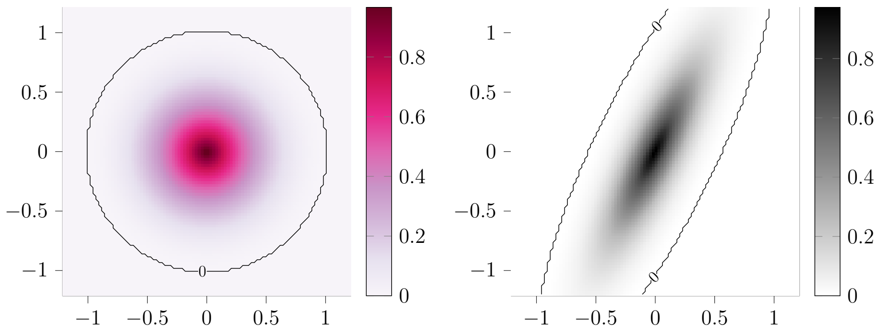

Figure 5.1: Top view of the two-dimensional kernel resulting from the Wendland RBF $\phi_{3,0}$ (left) and a transformed version of it (right). The transformation is given by the full-rank stretching and rotation matrix A_1. The black line visualizes supports.

The following aligns with ideas of [CLMM06]. Let K be a radially symmetric kernel on $T(\Omega)$ with RBF ϕ. Applying a transformation T on K yields

$$K_T(x,y) = K(Tx,Ty) = \phi(\|Tx - Ty\|_2) \qquad \text{for all } x,y \in \Omega.$$

For now, we assume the transformation $T : \Omega \longrightarrow T(\Omega)$ to be linear and to have the matrix representation A, i.e.,

$$T(x) = Ax, \qquad \text{for all } x \in \Omega.$$

Furthermore, we demand the matrix $B := A^T A$ to be positive definite. Consequently, the mapping

$$x \longmapsto \sqrt{x^T B x} =: \|x\|_B$$

defines a norm on Ω, here denoted as $\|\cdot\|_B$, and K_T is given by

$$\begin{aligned} K_T(x,y) = \phi\left(\|Tx - Ty\|_2\right) &= \phi\left(\|A(x-y)\|_2\right) \\ &= \phi\left(\sqrt{(x-y)^T\, A^T A\,(x-y)}\right) = \phi\left(\|x-y\|_B\right). \end{aligned}$$

The unit spheres $\mathbb{S}_{B_i} = \{x \,:\, \|x\|_{B_i} = 1\}$ of the norms $\|\cdot\|_{B_i}$ for $i = 1,2,3$ build by the matrices

$$\begin{aligned} B_1 &= A_1^T A_1, & A_1 &= \begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix} \begin{pmatrix} \cos(\pi/7) & -\sin(\pi/7) \\ \sin(\pi/7) & \cos(\pi/7) \end{pmatrix}, \\ B_2 &= A_2^T A_2, & A_2 &= \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}, \text{ and} \\ B_3 &= A_3^T A_3, & A_3 &= \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \end{aligned}$$

and the Euclidean norm are visualized in Fig. 5.2. Additionally, Fig. 5.1 shows the impact of the transformation A_1 on a kernel.

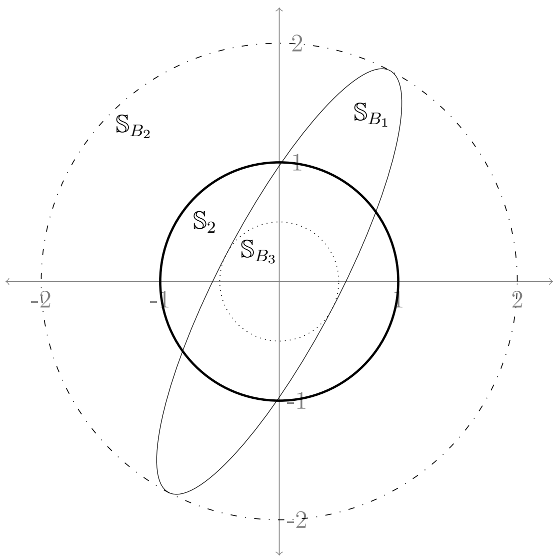

Figure 5.2: Comparison of the Euclidean unit sphere $\mathbb{S}_2$ (thick line), $\mathbb{S}_{B_1}$ (thin line) corresponding to the volume preserving transformation A_1, $\mathbb{S}_{B_2}$ (dash dotted line) corresponding to the enlarging transformation A_2, and $\mathbb{S}_{B_3}$ (dotted line) corresponding to the squeezing transformation A_3.

To sum this up, if K is a radially symmetric kernel that can be represented by the RBF ϕ applied to the Euclidean norm, its (linear) transformation kernel

$$K_T(x, y) = \phi(\|x - y\|_B)$$

can be represented by the same RBF ϕ after a different norm. Every symmetric positive semi definite matrix B can be factorized as $B = A^T A$, where A has full rank, e.g., by Cholesky decomposition. Applying Theorem 5.2, which provides positive (semi-) definiteness of the transformation kernel K_T if the initial kernel K was positive (semi-) definite, yields the subsequent statement.

Theorem 5.5. *Let K be a radially symmetric positive (semi-)definite kernel with RBF ϕ, then*

$$\phi(\|x - y\|)$$

defines a positive (semi-)definite kernel for any norm $\|\cdot\|$ induced by a symmetric positive definite matrix.

We observe that the positive (semi-)definiteness of a RBF is not depending on the Euclidean norm. However, Theorem 5.5 does not hold for any norm in $\mathbb{R}^d$. For example, [Kol92] demonstrates that the interpolation matrix of the Gaussian RBF (Ex. 2.19) with $\alpha = 1$ acting on the $\|\cdot\|_q$-norm is not positive semi-definite for $q > 2$ and more than two interpolation points, i.e., $|X| \geqslant 3$. For further details, we refer to [Kol09] and the references therein.

5.2 Native Spaces

Several works have previously investigated the native spaces of certain transformation kernels. In [BDL10], native spaces of transformation kernels are examined under the assumption that the initial kernel is translation-invariant and the transformation can be represented by a matrix. The authors demonstrate that the transformation kernel's native spaces can be characterized using the Fourier transform, aligned to the finding of Theorem 2.30. In [BLRS15, Theorem 2] isometric isometry between the initial kernel's RKHS and the RKHS of its variably scaled version, a specific transformation kernel, is shown. In [SC08, Prop. 4.37] the same result is stated for the transformation $x \longmapsto cx$, where $c > 0$.

We, however, provide a result covering all kinds of transformations by minimizing requirements and generalizing results. Roughly speaking, the proceeding theorem states that transforming the kernel has the same effect on the RKHS as transforming the input space.

Theorem 5.6. *Let $T : \Omega \longrightarrow T(\Omega)$ and the kernel K be positive semi-definite on $T(\Omega)$. Then,*

$$\mathcal{H}_{K_T,\Omega} = \mathcal{H}_{K,T(\Omega)} \circ T := \left\{ f \circ T \;:\; f \in \mathcal{H}_{K,T(\Omega)} \right\} \tag{5.2}$$

and

$$\langle f \circ T,\ g \circ T \rangle_{K_T,\Omega} = \langle f,\ g \rangle_{K,T(\Omega)} \qquad \text{for all } f, g \in \mathcal{H}_{K,T(\Omega)}.$$

The mapping

$$\mathcal{T} : \mathcal{H}_{K,T(\Omega)} \longrightarrow \mathcal{H}_{K_T,\Omega}, \quad f \longmapsto f \circ T$$

is an isometric isomorphism.

Proof. By Theorem 5.2 and Section 2.3, K_T is positive semi-definite and its native space $\mathcal{H}_{K_T,\Omega}$ exists. We consider the dense subspaces $S_{K,T(\Omega)}$ and $S_{K_T,\Omega}$ first. Let $f \in S_{K,T(\Omega)}$ have the form

$$f = \sum_{i=1}^{N} \alpha_i \, K(\cdot, y_i) \tag{5.3}$$

for $\{y_1, \ldots, y_N\} \subset T(\Omega)$. For every $y \in T(\Omega)$ there exists (possibly more than one) $x \in \Omega$ such that $T(x) = y$. Let $\{x_1, \ldots, x_N\} \subset \Omega$ be such that $T(x_i) = y_i$ for $i = 1, \ldots, N$. Then,

$$\begin{aligned} f \circ T\,(x) &= \sum_{i=1}^{N} \alpha_i \, K(T(x), y_i) \\ &= \sum_{i=1}^{N} \alpha_i \, K\big(T(x), T(x_i)\big) \\ &= \sum_{i=1}^{N} \alpha_i \, K_T(x, x_i) \in S_{K_T,\Omega} \qquad \text{for all } x \in \Omega. \end{aligned}$$

This implies the relation $S_{K,T(\Omega)} \circ T \subseteq S_{K_T,\Omega}$. The opposite relation follows by the same equation. Let $\{\tilde{x}_1, \dots, \tilde{x}_N\} \subset \Omega$ be another point set satisfying $T(\tilde{x}_i) = y_i$ for $i = 1, \dots, N$, then

$$\sum_{i=1}^{N} \alpha_i \; K_T(\cdot, x_i) = f \circ T = \sum_{i=1}^{N} \alpha_i \; K_T(\cdot, \tilde{x}_i).$$

Consequently, the mapping $\mathcal{T}$ is well-defined on the dense subset $S_{K_T,\Omega}$.

In order to show isometric isomorphy of $\mathcal{T}$, we examine bijectivity, linearity and isometry. In fact, $\mathcal{T}$ is bijective with the inverse mapping

$$\mathcal{T}^{-1} : S_{K_T,\Omega} \longrightarrow S_{K,T(\Omega)}, \quad f \longmapsto f \circ T^{-1},$$

since

$$\mathcal{T}(\mathcal{T}^{-1}(f)) = \mathcal{T}(f \circ T^{-1}) = f \circ T^{-1} \circ T = f.$$

We emphasize

$$f \circ T^{-1}(y) = \sum_{i=1}^{N} \alpha_i \; K_T\left(T^{-1}(y), x_i\right) = \sum_{i=1}^{N} \alpha_i \; K\left(y, T(x_i)\right) = \sum_{i=1}^{N} \alpha_i \; K\left(y, y_i\right) \in S_{K,T(\Omega)},$$

where $T^{-1}(y)$ denotes the pre-image of y and may consist of more than one element and $f \in S_{K_T,\Omega}$ has the form $f = \sum_{i=1}^{N} \alpha_i \; K_T\left(\cdot, x_i\right)$. Furthermore, $\mathcal{T}$ is linear as

$$\mathcal{T}\left(af + g\right) = a\left(f \circ T\right) + g \circ T = \left(a\mathcal{T}(f) + \mathcal{T}(g)\right) \qquad \text{for all } f, g \in S_{K,T(\Omega)}.$$

Regarding isometry, let $f, g \in S_{K,T(\Omega)}$, where f is defined as in (5.3) and $g = \sum_{j=1}^{M} \beta_j K(\cdot, \tilde{y}_j)$ for $\{\tilde{y}_1, \dots, \tilde{y}_N\} \subset T(\Omega)$, then

$$\begin{aligned}
\langle f \circ T, \; g \circ T\rangle_{K_T,\Omega} &= \left\langle \sum_{i=1}^{N} \alpha_i \; K_T\left(\cdot, x_i\right), \; \sum_{j=1}^{M} \beta_j \; K_T\left(\cdot, \tilde{x}_j\right) \right\rangle_{K_T,\Omega} \\
&= \sum_{i=1}^{N} \sum_{j=1}^{M} \alpha_i \beta_j \; K_T(x_i, \tilde{x}_j) \\
&= \sum_{i=1}^{N} \sum_{j=1}^{M} \alpha_i \beta_j \; K\left(T(x_i), T(\tilde{x}_j)\right) \\
&= \left\langle \sum_{i=1}^{N} \alpha_i \; K(\cdot, y_i), \; \sum_{j=1}^{M} \beta_j \; K(\cdot, \tilde{y}_j) \right\rangle_{K,T(\Omega)} \\
&= \langle f, \; g\rangle_{K,T(\Omega)}
\end{aligned} \tag{5.4}$$

Consequently, $\|\mathcal{T}(f)\|_{K_T,\Omega} = \|f \circ T\|_{K_T,\Omega} = \|f\|_{K,T(\Omega)}$ for all $f \in S_{K,T(\Omega)}$.

To generalize the results to native spaces, we extend the mapping $\mathcal{T}$ to $\mathcal{H}_{K,T(\Omega)}$ by

$$\mathcal{T}(f)(x) := \lim_{n\to\infty} \mathcal{T}(f_n)(x) = \lim_{n\to\infty} (f_n \circ T)(x) = f \circ T(x),$$

where $f \in \mathcal{H}_{K,T(\Omega)}$ is the pointwise limit of the Cauchy sequence $\{f_n\}_{n\in\mathbb{N}} \subset S_{K,T(\Omega)}$. Because of (5.4), $(f_n \circ T)_{n\in\mathbb{N}}$ is a Cauchy sequence in $S_{K_T,\Omega}$ and by Theorem 2.24 it is $f \circ T \in \mathcal{H}_{K_T,\Omega}$. With this definition at hand, the required statements follow. ∎

We recall Theorem 2.28 to emphasize the common occurrence of isometric isomorphic native spaces $\mathcal{H}_{K_1,\Omega}$ and $\mathcal{H}_{K,\Omega}$. Such relation arises for any continuous kernels K_1 and K acting on a domain $\Omega \subseteq \mathbb{R}^d$. However, the distinctive aspect of the aforementioned theorem lies in the equality (5.2). This knowledge is valuable for the examination of the anisotropic kernels considered in Chapter 6 and Chapter 7.

We conclude with an observation that will be needed later in Section 6.2.

Lemma 5.7. *Let K' and K be positive semi-definite kernels on Ω and $T(\Omega)$, respectively, and $T : \Omega \longrightarrow T(\Omega)$, such that*

$$\mathcal{H}_{K',\Omega} \subseteq \mathcal{H}_{K_T,\Omega}.$$

Then there exist a kernel κ' on $T(\Omega)$, so that $K'(x,y) = \kappa'_T(x,y)$ for all $x, y \in \Omega$.

Proof. For all $f \in \mathcal{H}_{K',\Omega} \subseteq \mathcal{H}_{K_T,\Omega} = \mathcal{H}_{K,T(\Omega)} \circ T$ there exist a function $g \in \mathcal{H}_{K,T(\Omega)}$ so that $f = g \circ T$. This is also the case for $K'(x,\cdot) \in \mathcal{H}_{K',\Omega}$ for all $x \in \Omega$. The symmetry of K' yields that there exist a kernel κ' on $T(\Omega)$, so that

$$K'(x,y) = \kappa'(T(x),T(y)) = \kappa'_T(x,y) \qquad \text{for all } x, y \in \Omega.$$

∎

5.3 Interpolation

This section is concerned with the interpolation method out of Section 2.1 using transformation kernels. We apply results of Chapter 2, regarding the approximation error (Section 5.3.1) and numerical stability (Section 5.3.2), to the transformation kernel. Additionally, we deduce conditions under which the results for the transformation kernel improve compared to its initial kernel.

In the following the interpolants $s_{f,K,X}$ are defined as in (2.7). We generalize the result from [BLRS15] of variably scaled kernels to more general transformations in (5.5) and extend it.

Lemma 5.8. *Let $T : \Omega \longrightarrow T(\Omega)$ be injective and K a positive definite kernel on $T(\Omega)$. Furthermore, let $X = \{x_1, \ldots, x_N\} \subset \Omega$ be a pairwise distinct point set and $(f \circ T)_X = f_{T(X)} \in \mathbb{R}^N$ be function values for a function f acting on $T(\Omega)$. Then,*

$$s_{f\circ T,K_T,X} = s_{f,K,T(X)} \circ T \tag{5.5}$$

and

$$\left\|s_{f,K,T(X)}\right\|_{K,T(\Omega)} = \left\|s_{f\circ T,K_T,X}\right\|_{K_T,\Omega}.$$

Let additionally $f \in \mathcal{H}_{K,T(\Omega)}$, then

$$\left\|f - s_{f,K,T(X)}\right\|_{K,T(\Omega)} = \left\|f \circ T - s_{f\circ T,K_T,X}\right\|_{K_T,\Omega}.$$

Proof. By Theorem 5.2 and Section 2.1 there exist unique interpolants $s_{f\circ T,K_T,X} \in S_{K_T,X}$ and $s_{f,K,T(X)} \in S_{K,T(X)}$. Because of (5.1), the following relation holds

$$\mathbf{A}_{K_T,X}\, c = \mathbf{A}_{K,T(X)}\, c = f_{T(X)} = (f \circ T)_X,$$

for a unique $c \in \mathbb{R}^N$. This shows, in order to fulfill the interpolation condition, the coefficients of $s_{f\circ T,K_T,X}$ and $s_{f,K,T(X)}$ have to equal $c = (c_1, \dots, c_N)^T \in \mathbb{R}^N$. Consequently,

$$s_{f\circ T,K_T,X} = \sum_{i=1}^{N} c_i K_T(\cdot, x_i) = \sum_{i=1}^{N} c_i K(T\cdot, T(x_i)) = s_{f,K,T(X)} \circ T.$$

The two norm equalities follow directly from Theorem 5.6. ∎

5.3.1 Approximation Error

The paper [BDL10] and [WMP24] provide error bounds for transformations that can be represented as matrices. We, however, intend to minimize the requirements on the transformation T.

In line with the standard case, the approximation error between the target function f and interpolant $s_{f,K_T,X}$ can be bounded by the power function multiplied with the norm of the target function (Lemma 5.9). We describe the transformation kernel's power function $P_{K_T,X}$ by the power function $P_{K,T(X)}$ of the initial kernel K (Lemma 5.10) to find an upper bound for $P_{K_T,X}$ depending on the fill distance $h_{T(X),T(\Omega)}$ in Theorem 5.11. Subsequently, we compare the bounds on $P_{K_T,X}$ and $P_{K,X}$.

Lemma 5.9. *Let $T : \Omega \longrightarrow T(\Omega)$ be injective and K a positive definite kernel on $T(\Omega)$, $X \subset \Omega$ a finite pairwise distinct point set, and $f \in \mathcal{H}_{K_T,\Omega}$. Then*

$$|f(x) - s_{f,K_T,X}(x)| \leqslant P_{K_T,X}(x) \|f\|_{K_T,\Omega} \qquad \text{for all } f \in \mathcal{H}_{K_T,\Omega}.$$

Proof. Theorem 5.2 assures positive definiteness of the transformation kernel K_T. The required bound directly follows from Theorem 2.40. ∎

Our focus shifts to an exploration of the power function associated with the transformation kernel. Let $\Omega \subset \mathbb{R}^d$ be a bounded set satisfying an ICC (Definition 2.42) for an angle $\theta \in (0, \pi/2)$ and a radius $r > 0$. If K_T is a translation-invariant kernel with univariate function $\Phi_T \in C(\mathbb{R}^d)$ and $X \subset \Omega$ a finite pairwise distinct point set satisfying $h_{X,\Omega} \leqslant h_0$, we can apply Theorem 2.43 and Remark 2.44 on the transformation kernel to obtain the following bound on the power function

$$P^2_{K_T,X}(\tilde{x}) \leqslant c_1 \sup_{x \in B(0,\, 2c_2 h_{X,\Omega})} |\Phi_T(x) - p(x)| = F_{\Phi_T,\Omega}(h_{X,\Omega}), \qquad \text{for all } \tilde{x} \in \Omega,$$

where p is an arbitrary polynomial from $\pi_m(\mathbb{R}^d)$, $m \in \mathbb{N}$, and h_0, c_1 and c_2 come from Theorem 2.43.

However, the requirement for K_T to be translation-invariant imposes constraints that can be omitted.

Lemma 5.10. *Let $T : \Omega \longrightarrow T(\Omega)$ be injective and K be a positive definite kernel on $T(\Omega)$, then*

$$P^2_{K_T,X} = P^2_{K,T(X)} \circ T.$$

Proof. Definition 2.27 and Lemma 5.8 yield

$$\begin{aligned} P^2_{K_T,X}(x) &= \left\| K_T(\cdot,x) - s_{K_T(\cdot,x),K_T,X} \right\|_{K_T,\Omega} \\ &= \left\| K\left(\cdot, T(x)\right) \circ T - s_{K\left(\cdot,T(x)\right),K,T(X)} \circ T \right\|_{K_T,\Omega}. \end{aligned}$$

With the inner product statement out of Theorem 5.6, we obtain

$$P^2_{K_T,X}(x) = \left\| K\left(\cdot, T(x)\right) - s_{K\left(\cdot,T(x)\right),K,T(X)} \right\|_{K,T(\Omega)} = P^2_{K,T(X)} \circ T\,(x).$$

■

We bound the power function $P_{K_T,X}$ in dependence of the fill distance $h_{T(X),T(\Omega)}$.

Theorem 5.11. *Let $T : \Omega \longrightarrow T(\Omega)$ such that $T(\Omega) \subseteq \mathbb{R}^d$ is bounded and satisfies an ICC. Furthermore, let K be a positive definite translation-invariant kernel on $T(\Omega)$, with a univariate function $\Phi \in C(\mathbb{R}^d)$, $T(\Omega)$ fulfill an ICC, and $X \subset \Omega$ be a finite set such that $T(X)$ is pairwise distinct satisfying $h_{T(X),T(\Omega)} \leqslant h_0$. Then,*

$$P^2_{K_T,X}(x) \leqslant F_{\Phi,T(\Omega)}(h_{T(X),T(\Omega)}) \tag{5.6}$$

where $F_{\Phi,T(\Omega)}$ comes from Remark 2.44.

Proof. Since $\Phi \in C(\mathbb{R}^d)$ is positive definite on $T(\Omega)$, $T(\Omega)$ is a bounded set satisfying an ICC, $T(X)$ is pairwise distinct and $h_{T(X),T(\Omega)} \leqslant h_0$ holds, we are in the position to apply Theorem 2.43 on $P^2_{K,T(X)}$. Together with Remark 2.44, this yields the bound

$$P^2_{K,T(X)}(T(x)) \leqslant F_{\Phi,T(\Omega)}(h_{T(X),T(\Omega)})$$

The relation $P^2_{K_T,X}(x) = P^2_{K,T(X)}(Tx)$ for all $x \in \Omega$ of Lemma 5.10 finishes the proof. ■

We seek to understand the criteria for selecting the transformation T to decrease the expected error. From Theorem 2.43 and Theorem 5.11 we collect the upper bounds

$$\begin{aligned} P^2_{K,X}(\tilde{x}) &\leqslant F_{\Phi,\Omega}(h_{X,\Omega}) \text{ and} \\ P^2_{K_T,X}(\tilde{x}) &\leqslant F_{\Phi,T(\Omega)}(h_{T(X),T(\Omega)}) \qquad \text{for all } \tilde{x} \in \Omega. \end{aligned}$$

Even though the function F is increasing we cannot directly expect a better approximation, when using a transformation satisfying $h_{T(X),T(\Omega)} \leqslant h_{X,\Omega}$, since the function F does as well depend on the angle and radius of the ICC of $T(\Omega)$ and Ω respectively. Upon closer examination of Theorem 2.43 and Theorem 5.11, it is

$$\begin{aligned} P^2_{K,X}(\tilde{x}) &\leqslant 9 \sup_{x \in B(0,\, 2ch_{X,\Omega})} |\Phi(x) - p(x)| &&= F_{\Phi,\Omega}(h_{X,\Omega}) \text{ and} \\ P^2_{K_T,X}(\tilde{x}) &\leqslant 9 \sup_{x \in B(0,\, 2c_T h_{T(X),T(\Omega)})} |\Phi(x) - p(x)| &&= F_{\Phi,T(\Omega)}(h_{T(X),T(\Omega)}) \quad \text{for all } \tilde{x} \in \Omega. \end{aligned}$$

The bounds hold for constants c and c_T depending on the angles θ and $\theta_T \in (0, \pi/2)$ given by the ICC of Ω and $T(\Omega)$ and fill distances satisfying $h_{X,\Omega} \leqslant r/c$ and $h_{T(X),T(\Omega)} \leqslant r_T/c_T$, where r and $r_T > 0$ are the radii defined by the ICC of Ω and $T(\Omega)$. In order to obtain a smaller error bound for the transformation kernel, we need

$$c_T h_{T(X),T(\Omega)} \leqslant c h_{X,\Omega}, \text{ while } h_{X,\Omega} \leqslant \frac{r}{c} = h_0 \text{ and } h_{T(X),T(\Omega)} \leqslant \frac{r_T}{c_T} = h_{T,0}. \tag{5.7}$$

Example 5.12. Let $\Omega = [0, 0.5] \times [0, 2] \subset \mathbb{R}^2$ then it satisfies an ICC with angle $\theta = \pi/2$ and radius $r = 0.5$. Let the transformation T be given by the matrix

$$T(x) = \begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix} x,$$

then $T(\Omega) = [0, 1]^2 \in \mathbb{R}^2$ satisfying an ICC with angle $\theta_T = \pi/2$ and radius $r_T = 1$. With the formulas of Theorem 2.43, we compute

$$\begin{aligned} c = c_T &= \frac{64}{3} m^2 \sim 21.\overline{3} m^2, \\ h_0 &= \frac{3}{128} m^2 \sim 0.023 m^2 \text{ and} \\ h_{T,0} &= \frac{3}{64} m^2 \sim 0.047 m^2, \end{aligned}$$

where $m \in \mathbb{N}$ is determined by the choice of the polynomial. In the setting visualized in Fig. 5.3, we have $h_{T(X),T(\Omega)} \leqslant h_{X,\Omega}$. Therefore, the requirements of (5.7) are satisfied if $h_{T(X),T(\Omega)} \leqslant h_0$, and the bound on the transformed power function $P_{K_T,\Omega}$ is smaller than the bound of the initial kernel's power function $P_{K,\Omega}$.

At this juncture, it is important to note two key points. Firstly, the requirements outlined in (5.7) solely guarantee an enhanced bound of the power function, without directly addressing the actual approximation error. Secondly, the power function constitutes just one component of the error bound, see Lemma 5.9. The additional factor is given by the native spaces norm of the target function. For a comparison between $\|f\|_{K_T,\Omega}$ and $\|f\|_{K,\Omega}$, we refer to Section 3.2, where we examined native spaces associated with summation kernels.

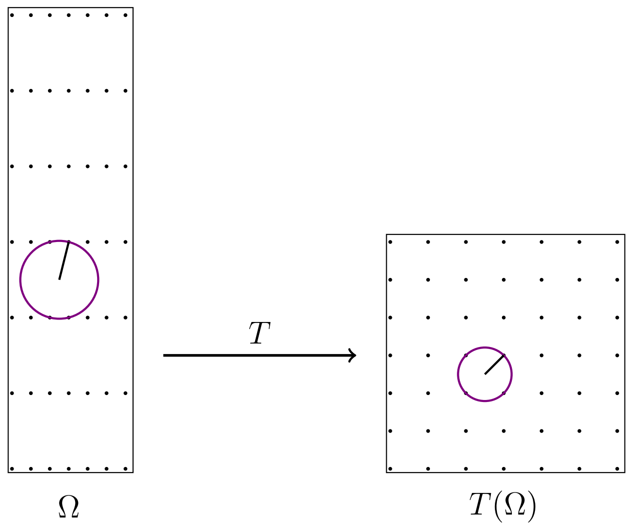

Figure 5.3: Visualization of the transformation given in Ex. 5.12, together with the fill distances (black line), which is defined as the radius of the largest hole (violet circle) in the data set (black dots).

In Section 5.1 we discussed transformations T having a matrix representation A, such that $B := A^T A$ is a symmetric positive definite matrix. In this setting, the fill distance $h_{T(X),T(\Omega)}$ is given by the fill distance of X and Ω measured in the norm $\|\cdot\|_B$ instead of the Euclidean norm, i.e.,

$$\begin{aligned} h_{T(X),T(\Omega)} &= \sup_{x\in\Omega} \min_{1\leqslant i\leqslant N} \|T(x) - T(x_i)\|_2 \\ &= \sup_{x\in\Omega} \min_{1\leqslant i\leqslant N} \|A(x - x_i)\|_2 \\ &= \sup_{x\in\Omega} \min_{1\leqslant i\leqslant N} (x - x_i)^T B(x - x_i) \\ &= \sup_{x\in\Omega} \min_{1\leqslant i\leqslant N} \|x - x_i\|_B \,. \end{aligned}$$

Since $\lambda_{\min}(A)\|x\|_2 \leqslant \|Ax\|_2 \leqslant \lambda_{\max}(A)\|x\|_2$, the above calculation gives

$$\lambda_{\min}(A)\, h_{X,\Omega} \leqslant h_{T(X),T(\Omega)} \leqslant \lambda_{\max}(A)\, h_{X,\Omega},$$

see [CLMM06, Theorem 2]. Theorem 3.2 of [WMP24] concludes further that in the special case of a transformed Matérn kernel, where the transformation is given by a full-rank matrix, the error evolves at the same rate when examining the fill distance. It shows that the error can be bounded by the same function F of Remark 2.44, as its initial kernel up to a non-data-point-dependent constant factor.

5.3.2 Numerical Stability

This section is devoted to examining whether and how the stability of the interpolation process changes when employing a transformation kernel compared to its initial version. Given that Section 2.4.2 emphasizes the importance of investigating the minimal eigenvalue of the interpolation matrix for insights, we adapt Theorem 2.48 to the case of transformation kernels.

Theorem 5.13. *Let $T : \Omega \longrightarrow T(\Omega)$ and K be translation-invariant kernel on $T(\Omega)$ and its univariate function Φ be such that Φ possesses a positive Fourier transform $\widehat{\Phi} \in C(\mathbb{R}^d \backslash \{0\})$. Then*

$$\lambda_{\min}(\mathbf{A}_{K_T,X}) \geqslant G_\Phi(q_{T(X)}),$$

where G_Φ comes from Remark 2.49.

Proof. Since Φ satisfies the requirements of Theorem 2.48 there exists the function G_Φ of Remark 2.49, such that $\lambda_{\min}(\mathbf{A}_{K,T(X)}) \geqslant G_\Phi(q_{T(X)})$ holds. Since $\mathbf{A}_{K_T,X} = \mathbf{A}_{K,T(X)}$ by (5.1), we obtain the required results. ■

The aforementioned theorem proves that the function G_Φ serves as a lower bound for both the smallest eigenvalue of the interpolation matrix $\mathbf{A}_{K,X}$ and its transformed counterpart $\mathbf{A}_{K_T,X}$. In Tab. 2.4, the function G_Φ is provided for specific example kernels. However, to attain the corresponding bound, it is essential to evaluate the increasing function G_Φ at the separation distance q_X in the standard case, while for the transformed scenario G_Φ is assessed at the separation distance of the transformed point set $q_{T(X)}$. Consequently, for enhanced stability, it is advisable to employ a transformation T that enlarges the separation distance of X, ensuring that

$$q_{T(X)} \geqslant q_X.$$

Again, we refer to Ex. 5.12 to illustrate a setting where the aforementioned relation of separation distances holds true. A visualization is provided in Fig. 5.4. We consider the scenario of Section 5.1, where the transformation T is defined by a matrix A such that $B := A^T A$ is positive definite. Here, the separation distance $q_{T(X)}$, as defined in Definition 2.47, aligns with the separation distance of X, though measured by the norm $\|x\|_B$ rather than the Euclidean norm, i.e.,

$$\begin{aligned} q_{T(X)} &= \frac{1}{2} \min_{i \neq j} \big\|T(x_i) - T(x_j)\big\|_2 \\ &= \frac{1}{2} \min_{i \neq j} \big\|A(x_i - x_j)\big\|_2 \\ &= \frac{1}{2} \min_{i \neq j} (x_i - x_j)^T B (x_i - x_j) \\ &= \frac{1}{2} \min_{i \neq j} \big\|x_i - x_j\big\|_B \,. \end{aligned}$$

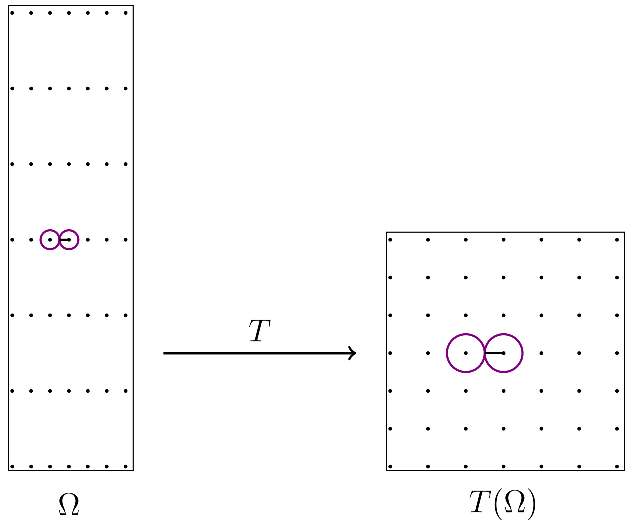

Figure 5.4: Visualization of the transformation given in Ex. 5.12, together with the separation distances (black line), which is defined to be the largest possible radius for two balls (violet circles) centered at different points (black dots) to be essentially disjoint.

Remark 5.14. We can expect improved stability using the transformation kernel if $q_X \leqslant q_{T(X)}$. Since the above calculation yields

$$\lambda_{\min}(A)\, q_X \leqslant q_{T(X)} \leqslant \lambda_{\max}(A)\, q_X,$$

see [CLMM06, Theorem 2], a better bound on the minimal eigenvalue of the interpolation matrix $\mathbf{A}_{K_T,X}$ in comparison to $\mathbf{A}_{K,X}$ is gained if $\lambda_{\min}(A) \geqslant 1$.

5.4 Numerical Tests

In the following sections, we examine two scenarios where interpolation with transformation kernels improves results. Firstly, in Section 5.4.1, we consider an anisotropic target function. Secondly, in Section 5.4.2, we explore a setting, where closely spaced points occur in one direction, while in the other direction, the points are widely spaced apart. Such a distribution of points can be observed in line measurements, for example.

We have already conducted comparisons of differing kernels K_1 and K_2 in Chapter 3. The focus of this section lies on transformation kernels, aiming to illustrate how these kernels alter their behavior under different transformations. Hence, we compare the kernels with their respective transformations and present two kernels, solely to underscore that the change in behavior is not depending upon the selected kernel but the chosen transformation.

5.4.1 Adaptation to Target Function

In this example, we consider an elongated region Ω such as an ideal tube, a blood vessel, or a river. These structures have a distinct geometry, characterized by a much greater length compared to their width. We assume that the target function we are working with also exhibits anisotropic behavior. Anisotropic behavior means that the function changes at different rates in different directions. Specifically, in this context, the function changes more slowly in the longitudinal direction (along the length of Ω, here represented as top-bottom) than in the transverse direction (across the width of Ω, here represented as right-left). This kind of anisotropic behavior can be observed in scientific and engineering fields. For instance, in fluid dynamics, the flow properties of a fluid within a narrow and elongated channel, like a pipe or blood vessel, typically vary more gradually along the length of the channel than across its width. In a river, the flow speed and other characteristics may change slowly as you move downstream (longitudinally) but can vary significantly across the river's width (transversely).

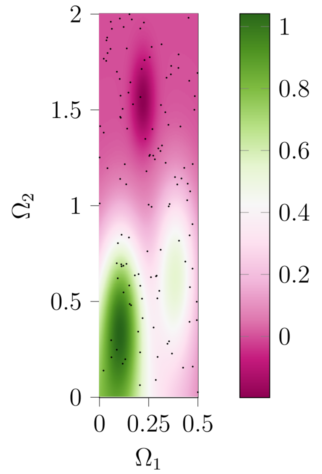

Figure 5.5: Top view of the target function $f_{\text{aniso-t}}$, and the scattered point set X_7 on Ω.

We consider the setting visualized in Fig. 5.5, in particular:

- The set $\Omega = \Omega_1 \times \Omega_2$, where $\Omega_1 = [0, 0.5]$ and $\Omega_2 = [0, 2]$.
- The point sets X_n, that consist of $|X_n| = 2^n$ randomly chosen points in Ω for $n = 6, \dots, 11$, so that $X_n \subset X_m$ for $n < m$. This results in the sizes $64, 128, 256, 512, 1024$ and 2048.
- The target function
$$f_{\text{aniso-t}} := f_{\text{franke}} \circ \begin{pmatrix} 0.5 & 0 \\ 0 & 2 \end{pmatrix},$$
which is given by a stretched version of the Franke function defined in (3.24).

For this setting, we compare the following kernels:

- The initial kernels K_1 and K_2, which are given by Wendland's RBFs $\phi_{3,0}$ and $\phi_{3,3}$ (Tab. 2.2).
- The transformation kernels $K_{1,T}$ and $K_{2,T}$. These kernels are transformation kernels of the corresponding initial kernels K_1 and K_2 with the transformation
$$T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \qquad x \longmapsto \begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix} x.$$
- The transformation kernels $K_{1,\Theta}$ and $K_{2,\Theta}$. These kernels are transformation kernels of the corresponding initial kernels K_1 and K_2 with the transformation
$$\Theta : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \qquad x \longmapsto \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} x.$$

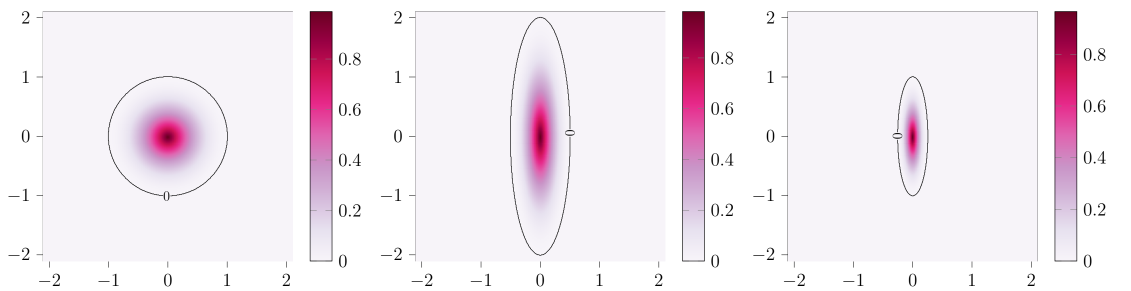

Figure 5.6: Top view of two-dimensional kernels and their support (black line). The initial kernel K_1 (left) and its transformed version $K_{1,T}$ (middle) and $K_{1,\Theta}$ (left).

The transformation kernels used are exemplarily represented for K_1 in Fig. 5.6. Looking at Fig. 5.7 (left), the transformation kernels $K_{1,T}$ and $K_{2,T}$ exhibit stability comparable to that of the initial kernels K_1 and K_2, yet their error is significantly improved, see Fig. 5.7 (right). Although we have chosen the optimal transformation T, any transformations that narrow the kernel along the Ω_1 direction and elongate it along the Ω_2 direction contribute to improved interpolations, provided they do not excessively downsize the kernel. Thus, there is no necessity to find the ideal transformation for improving the results.

The transformation Θ exemplifies the impact of scaling the transformation T by the factor 2, i.e., $2T = \Theta$. It is noted that the condition number improves by a constant factor with using the transformation kernels $K_{1,\Theta}$ and $K_{2,\Theta}$ compared to its initial versions K_1 and K_2, see Fig. 5.7 (left). Such enhancement can be explained by Remark 5.14. Nonetheless, the comparison of mean squared errors for these kernel yields inconclusive results, see Fig. 5.7 (right). The threshold for downsizing, mentioned before, has been reached here.

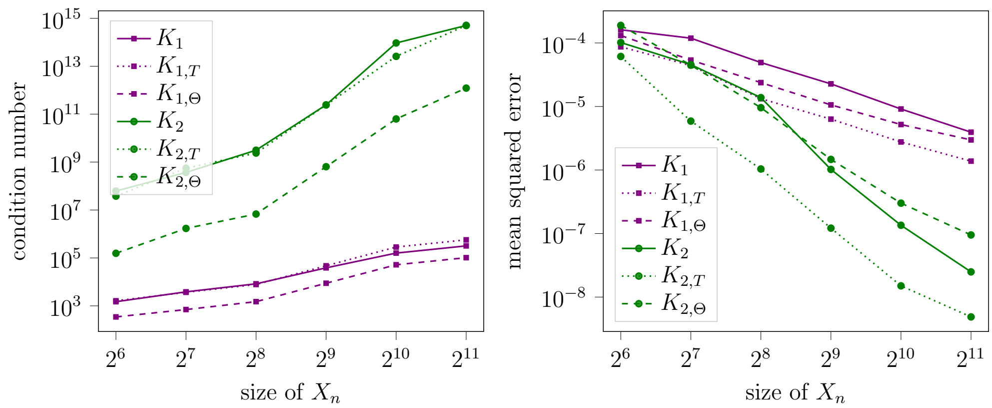

Figure 5.7: Comparison of the initial kernels K_j, their transformed versions $K_{j,T}$ and $K_{j,\Theta}$ for $j = 1, 2$ regarding their condition number (left) and mean squared error (right).

We conclude that transformation kernels offer a convenient approach to enhance results in an anisotropic setting. A transformation kernel is able to improve stability or approximation quality without significantly affecting the other. Hence, they serve as means to counteract the trade-off principle out of Section 2.4.3. The choice of transformation depends on the desired effect. Given an initial kernel with poor stability it is advisable to choose a stability improving transformation and vice versa for the approximation error.

5.4.2 Adaptation to Domain and Point Sets

Subsequently, we examine a similar scenario as Ex. 5.12, whose fill distance and separation distance is depicted in Fig. 5.3 and Fig. 5.4, respectively.

To conduct the interpolation and error analysis, it is essential to define three key components: the domain, the interpolation points, and the target function. In this example, we consider the setting visualized in Fig. 5.8, in particular:

- The domain $\Omega = \Omega_1 \times \Omega_2$, where $\Omega_1 = [0, 0.5]$ and $\Omega_2 = [0, 2]$.

- The point sets

$$\mathbf{X_n} = X_n^1 \times X^2 \subset \Omega, \qquad \text{where } X_n^1 = \left\{ \frac{0.5i}{2^n} \ : \ i = 0, 1, \ldots, 2^n \right\} \subset \Omega_1$$

$$\text{and } X^2 = \left\{ \frac{2i}{9} \ : \ i = 0, 1, \ldots, 9 \right\} \subset \Omega_2$$

  We use the point sets $\mathbf{X_n}$ for $n = 3, \ldots, 8$. This results in the amounts of points $|\mathbf{X_n}| = (2^n + 1) \cdot 10$ for $n = 3, \ldots, 8$, namely $90, 170, 330, 650, 1290$ and 2570.

- The isotropic non-differentiable target function

$$f_{\text{kink-t}}(x^1, x^2) := 4\left\|\cos\left(\frac{\pi}{12}\right)x^1 - \sin\left(\frac{\pi}{12}\right)x^2\right\|_2 + \sin\left(2\pi\left(\sin\left(\frac{\pi}{12}\right)x^1 + \cos\left(\frac{\pi}{12}\right)x^2\right)\right).$$

For this setting, we compare the following kernels:

- The standard kernels K_1 and K_2, which are given by Wendland's RBFs $\phi_{3,0}$ and $\phi_{3,3}$ (Tab. 2.2).

- The domain-adapted kernels $K_{1,T}$ and $K_{2,T}$. These kernels are transformation kernels of the corresponding initial kernels K_1 and K_2 with the transformation

$$T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \qquad x \longmapsto \begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix} x.$$

  Here, the components of the matrix were chosen to transform the set Ω into a square.

- The point-adapted kernels $K_{1,T_{\text{flex}}}$ and $K_{2,T_{\text{flex}}}$. The point-adapted kernels are again transformation kernels, but adapted to the developing point sets $\mathbf{X_n} = X_n^1 \times X^2$. We use a flexible transformation defined in the following manner

$$T_{\text{flex}} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \qquad x \longmapsto \begin{pmatrix} \frac{q_{X_n^1}+q_{X^2}}{q_{X_n^1}} & 0 \\ 0 & \frac{q_{X_n^1}+q_{X^2}}{q_{X^2}} \end{pmatrix} x,$$

  where q_X denotes the separation distance of X. This transforms the point sets $\mathbf{X_n}$ in such a manner that the distances between the points of $T_{\text{flex}}(\mathbf{X_n})$ coincide in each direction. Here, $q_{X_n^1} = \frac{1}{2^n+1}$ and $q_{X^2} = 2/9 = 0.\overline{2}$. Consequently, the matrix that defines T_{flex} is given by

$$\begin{pmatrix} 1 + \frac{2^{n+1}+2}{9} & 0 \\ 0 & \frac{9}{2^{n+1}+2} + 1 \end{pmatrix}. \tag{5.8}$$

  The distance between the points of the transformed point set $T_{\text{flex}}(\mathbf{X_n})$ along the Ω_1 direction aligns with the distance along the Ω_2 direction for every $n = 3, \ldots, 8$.

The transformation kernels used are exemplarily represented for K_1 in Fig. 5.6 (K_1 and $K_{1,T}$) and Fig. 5.9 ($K_{1,T_{\text{flex}}}$). The results regarding approximation error and numerical stability, visualized in Fig. 5.10, underline the analysis conducted in the preceding sections, namely Section 5.3.1 and Section 5.3.2.

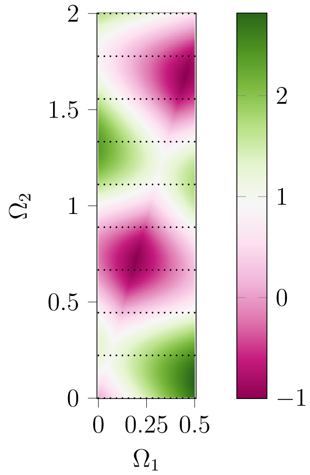

Figure 5.8: Top view of the target function $f_{\text{kink-t}}$ of Section 5.4.2, and the point set $\mathbf{X_4}$ on Ω.

In Fig. 5.10 (left), we see that the transformed versions $K_{1,T}$ and $K_{2,T}$ exhibit a smaller numerical condition number but still have similar behavior to their initial kernels; the condition number increases at the same rate. This phenomenon arises from the transformation's lack of adaptation to the point set. As the separation distance of $\mathbf{X_n}$ decreases, the separation distance $q_{T(\mathbf{X_n})}$ decreases at the same rate. Upon examining the numerical condition number of the point-adapted kernels $K_{1,T_{\text{flex}}}$ and $K_{2,T_{\text{flex}}}$, a significant improvement is observed, as the condition number remains nearly constant. The flexible transformation effectively counteracts the rapidly decreasing separation distance. At this juncture we recall our findings from Remark 5.14, which assures better stability if the smallest eigenvalue of the matrix in defining the transformation is greater or equal to 1. This criterion is met by the matrix (5.8) defining T_{flex}, as it primarily elongates the domain Ω along the Ω_1 direction.

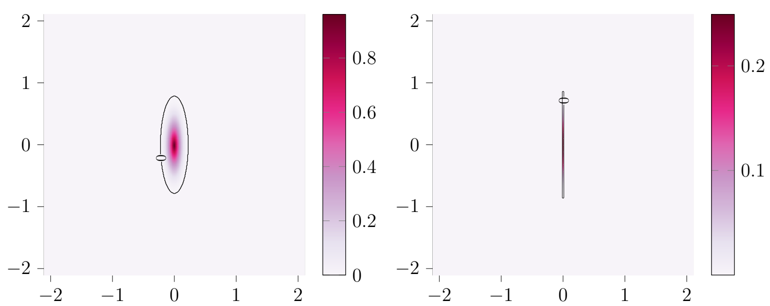

Figure 5.9: Top view of $K_{1,T_{\text{flex}}}$ adapted to $\mathbf{X_3}$ (left) and $\mathbf{X_8}$ (right), and its support (black line).

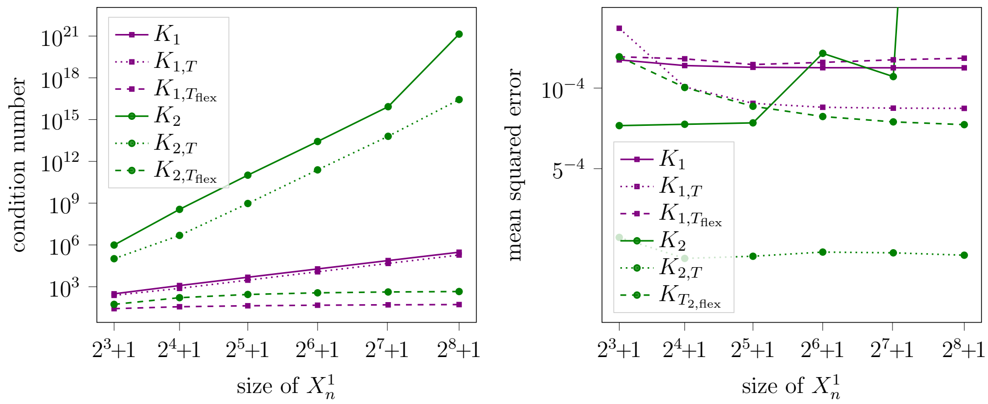

Figure 5.10: Comparison of the initial kernels K_j, their transformed versions $K_{j,T}$ adapted to the domain Ω, and the kernels $K_{j,T_{\text{flex}}}$, which are flexibly adapted to the point sets $\mathbf{X_n}$ for $j = 1, 2$ and $n = 3, \ldots, 8$ regarding the condition number (left) and mean squared error (right).

With regard to the approximation error in Fig. 5.10 (right), we observe a significant reduction in mean squared error for the domain-adapted kernels $K_{1,T}$ and $K_{2,T}$. This can be explained by the decrease in fill distance, as depicted in Fig. 5.3. For the point-adapted kernel $K_{1,T_{\text{flex}}}$, the approximation rate remains unchanged compared to its initial kernel K_1. We attribute this to the fact, that the matrix of (5.8) does not compress the set Ω in the Ω_2 direction. Consequently, the fill distance $h_{\mathbf{X_n},\Omega}$ remains approximately the same as $h_{T_{\text{flex}}(\mathbf{X_n}),T_{\text{flex}}(\Omega)}$. This behavior could also be observed for K_2 if its high condition number did not impair the interpolation. We must emphasize here that the target is not $C^1(\Omega)$. Consequently, based on the results of the preceding Chapter 3, the kernel K_2 should exhibit an error comparable to that of K_1. However, the two transformations depicted not only improve the error compared to their initial kernel K_2 but even surpass K_1 in terms of error reduction. We resume the following:

1. The domain-adapted kernel enhances the approximation quality significantly, while changing the condition number by a factor independent of the interpolation point set. This is particularly advantageous when the initial condition number is so poor that it undermines the interpolation result, as observed for K_2.

2. The flexible transformation should be employed to improve the stability while demanding that the approximation quality remains comparable to that of the initial kernel.

Hence, transformation kernels improve results effectively when the transformation is customized to fit the circumstances and requirements of the application.

Chapter 6

Orthogonal Summation Kernels

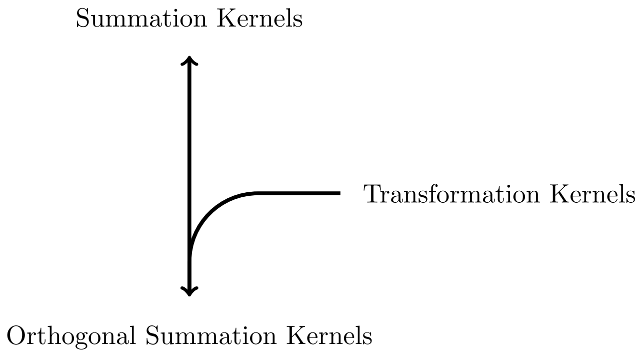

Figure 6.1: Schematic relation between summation, transformation, and orthogonal summation kernels.

The approach of *orthogonal summation kernels* aims to expand the variety of kernels available for interpolation problems. This method provides additional means of tailoring kernels to the underlying interpolation problem. The distinctive feature of these kernels, as well as the anisotropic tensor product kernels discussed in the subsequent Chapter 7, is their operation on the Cartesian product of subdomains, allowing different component kernels for each subdomain. This is particularly interesting because it is possible to impose different properties in different directions, thus adapting the kernel to the anisotropic structure of the target function or the interpolation points.

Currently, there is limited literature on this topic. In fact, [GM16] investigates positive semi-definite kernels operating on Cartesian products, introducing the term 'distinct component (DC)-strictly positive definite'. However, the focus in that work is on isotropic kernels, which we deliberately choose not to consider.

As far as we know, no one has yet explored an anisotropic sum of kernels. This is the focus of our investigation in this chapter. We combine the standard summation kernels of Chapter 3 with the transformation kernels of Chapter 5, as visualized in Fig. 6.1, to obtain an orthogonal summation kernel K on Ω,

$$K(x, y) = \sum K_{\ell,p_\ell}(x, y) = \sum K_\ell(p_\ell(x), p_\ell(y)),$$

where each component kernel K_ℓ is defined on Ω_ℓ, Ω is given by the Cartesian product of all Ω_ℓ, and p_ℓ denotes the projection form Ω to Ω_ℓ which can be viewed as a transformation. We call this kernel an orthogonal summation kernel, hinting at its underlying algebraic structure. The native space of such a kernel equals the orthogonal sum of its components' native spaces, see Theorem 6.7. Orthogonal summation kernels hold promise for good approximation of certain target functions. For instance, [WMP24] examines transformation kernels that reduce dimensionality. It is theoretically and numerically demonstrated that with such a dimension-reducing transformation and a target function that is invariant with respect to a subdomain Ω_2, i.e.,

$$f(x) = f_1(p_1(x)) + f_2(p_2(x)) \text{ with } f_2 \equiv 0,$$

the convergence rate of the error improves. Indeed, we can consider such a dimension-reducing transformation kernel K_T as an orthogonal summation kernel:

$$K_T(x, y) = K_1(p_1(x), p_1(y)) = K_1(p_1(x), p_1(y)) + K_2(p_2(x), p_2(y)),$$

where $K_2 \equiv 0$ and K_1 is a kernel operating on Ω_1, the subdomain of Ω where the target function varies. Thus, a detailed examination of orthogonal summation kernels is also warranted in terms of error improvement.

This chapter builds on the previous sections and explores the theoretical foundations and practical benefits of orthogonal summation kernels in detail. The key contributions of our work are:

- Examination of the Orthogonal Summation Kernels' Native Space: We describe the native space as an orthogonal sum of the component kernels' native spaces in Theorem 6.7.
- Detailed Analysis of Interpolation with Orthogonal Summation Kernels: We provide a thorough analysis of interpolation using orthogonal summation kernels, focusing on positive definiteness, approximation error (Section 6.3.1), and numerical stability (Section 6.3.2).
- Outstanding Performance in Anisotropic Sum Structures: We demonstrate the superior performance of interpolation with orthogonal summation kernels when the target function has an anisotropic sum structure (Section 6.4), supporting the results of Theorem 6.13 and Theorem 6.11.

We hope that this detailed analysis of orthogonal summation kernels paves the way for further research in this promising area.

This chapter is structured in the following way: In Section 6.1, we provide the definition of orthogonal summation kernels and address the issue of positive definiteness for kernels operating on a Cartesian product. We examine the native space of orthogonal summation kernels in Section 6.2, recognizing that the standard summation case can be considered a special case of the orthogonal summation kernel. The requirements for achieving a positive definite interpolation matrix and the interpolation with orthogonal summation kernels in general is discussed in Section 6.3, focusing on approximation error and numerical stability. Finally, in Section 6.4, we investigate the performance of orthogonal summation kernels using two different target functions.

6.1 Definition and Basic Properties

Subsequently, we provide a precise definition of the orthogonal summation kernel, visualized in Fig. 6.2, and draw the connection to transformation and summation kernels, that are discussed in Chapter 3 and Chapter 5. Furthermore, in Ex. 6.2, we illustrate why an orthogonal summation kernel with positive definite components is not automatically positive definite and close this section by briefly addressing translation-invariance of orthogonal summation kernels.

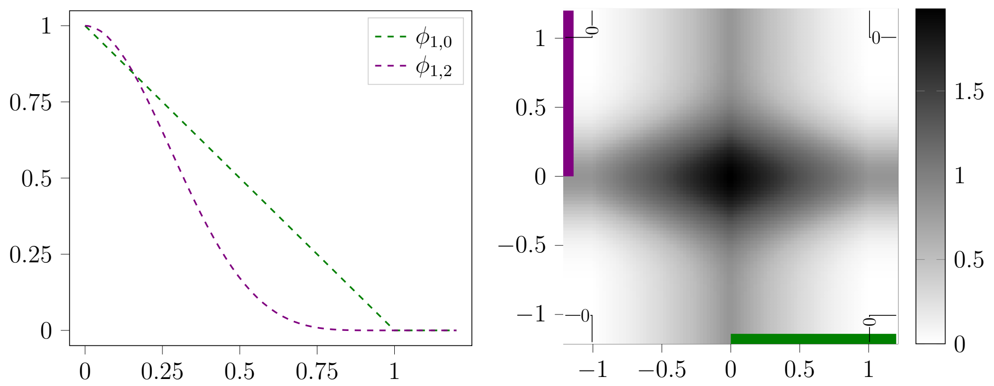

Figure 6.2: Visualization of two Wendland RBFs $\phi_{1,0}$ and $\phi_{1,2}$ (left), and their corresponding two-dimensional orthogonal summation kernel (right) with its support (black line). The green and purple bar emphasize that $\phi_{1,0}$ acts in x-direction and $\phi_{1,2}$ in y-direction.

Definition 6.1. Let K_ℓ be positive semi-definite kernels on $\Omega_\ell \subseteq \mathbb{R}^{d_\ell}$ for $\ell = 1, \dots, M$, and $\Omega = \times_{\ell=1}^{M} \Omega_\ell \subseteq \mathbb{R}^d$ the Cartesian product of Ω_ℓ, where $d = \sum_{\ell=1}^{M} d_\ell$. Then

$$K : \Omega \times \Omega \longrightarrow \mathbb{R},$$
$$K(x, y) = \sum_{\ell=1}^{M} K_\ell \left(p_\ell(x), p_\ell(y)\right) \text{ for } x, y \in \Omega$$

is called an *orthogonal summation kernel*, where $p_\ell : \Omega \longrightarrow \Omega_\ell$ denotes the projection from Ω onto Ω_ℓ for all $\ell = 1, \dots, M$.

The Definition 5.1 of transformation kernels gives the notation

$$K\left(p_\ell(x), p_\ell(y)\right) = K_{p_\ell}(x, y)$$

for all kernels K and all projections p_ℓ, $\ell = 1, \ldots, M$. Hence, the orthogonal summation kernel can be viewed as a summation kernel of transformation kernels, i.e.,

$$K(x, y) = \sum_{\ell=1}^{M} K_\ell\left(p_\ell(x), p_\ell(y)\right) = \sum_{\ell=1}^{M} K_{\ell,p_\ell}(x, y), \tag{6.1}$$

where the transformation is given by a projection. Let $X \subseteq \Omega = \times_{\ell=1}^{M}$ be a finite and pairwise distinct point set and $p_\ell(X)$ its projection onto Ω_ℓ for all $\ell = 1, \ldots, M$. By (3.1) and (5.1), the orthogonal summation kernel's interpolation matrix $\mathbf{A}_{K,X}$ is given by the sum of the components' interpolation matrices, i.e.

$$\mathbf{A}_{K,X} = \sum_{\ell=1}^{M} \mathbf{A}_{K_{\ell,p_\ell},X} = \sum_{\ell=1}^{M} \mathbf{A}_{K_\ell,p_\ell(X)}. \tag{6.2}$$

We cannot deduce that the projections $p_\ell(X) \subset \Omega_\ell$ of an arbitrary pairwise distinct point set $X \subset \Omega$ are also pairwise distinct, as Fig. 6.3 visualizes. Henceforth, we cannot conclude positive definiteness for the orthogonal summation kernel K from positive definite component kernels K_ℓ. The following Ex. 6.2 provides a setting where the orthogonal summation kernel of positive definite component kernels is not positive definite.

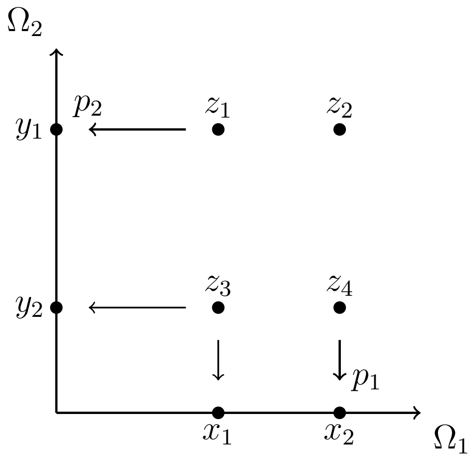

Figure 6.3: Visualization of the point set $X = \{z_1, z_2, z_3, z_4\} \in \Omega_1 \times \Omega_2$, and their projections p_1 and p_2 onto Ω_1 and Ω_2 respectively.

Example 6.2. Let K_1 and K_2 be symmetric and positive definite kernels on Ω_1 and Ω_2 respectively, and K be their orthogonal summation kernel on $\Omega = \Omega_1 \times \Omega_2$. We consider the point set $X = \{z_1, z_2, z_3, z_4\}$ visualized in Fig. 6.3. By (6.2) and Theorem 3.2 the orthogonal summation kernel's interpolation matrix is positive definite if at least one of its component matrices $\mathbf{A}_{K_1,p_1(X)}$ or $\mathbf{A}_{K_2,p_2(X)}$ are positive definite. However, in this example the projected point sets $p_1(X) = \{x_1, x_2, x_1, x_2\}$ and $p_2(X) = \{y_1, y_1, y_2, y_2\}$ are not pairwise distinct. Therefore, the interpolation matrices $\mathbf{A}_{K_1,p_1(X)}$ and $\mathbf{A}_{K_2,p_2(X)}$ are not positive definite by Remark 2.11.

Lemma 2.10 shows that the weaker requirement of K being positive semi-definite, does not depend on the point set X being pairwise distinct. This finding provides positive semi-definiteness of the interpolation matrices $\mathbf{A}_{K,p_\ell(X)}$ for all $X \subset \Omega$ and positive semi-definite kernels K. As sums of positive semi-definite matrices are again positive semi-definite, we can conclude positive semi-definiteness of the orthogonal summation kernel K. We return to the question of positive definiteness in Section 6.3.

Lemma 6.3. *If K_ℓ are positive semi-definite kernels on Ω_ℓ for $\ell = 1, \ldots, M$, their corresponding orthogonal summation kernel K is positive semi-definite on $\Omega = \times_{\ell=1}^{M} \Omega_\ell$.*

Proof. Let $X \in \Omega$ be a finite point set and $p_\ell(X)$ its projection onto Ω_ℓ, then

$$c^T \, \mathbf{A}_{K,X} \, c = \sum_{\ell=1}^{M} c^T \, \mathbf{A}_{K_\ell,p_\ell(X)} \, c \geqslant 0.$$

■

Lemma 6.4. *If K_ℓ are translation-invariant kernels with uni-variate functions Φ_ℓ on Ω_ℓ for $\ell = 1, \ldots, M$, their orthogonal summation kernel K is translation-invariant with the uni-variate function $\Phi := \sum_{\ell=1}^{M} \Phi_\ell \circ p_\ell$.*

Proof. Note, that the projection p_ℓ is linear for all $\ell = 1, \ldots, M$, and use Lemma 5.3 (*ii*) to obtain

$$K(x, y) = \sum_{\ell=1}^{M} K_{\ell,p_\ell}(x, y) = \sum_{\ell=1}^{M} (\Phi_\ell \circ p_\ell)\,(x - y).$$

■

6.2 Native Spaces

In this section we examine the orthogonal summation kernel's native space, combining results regarding the transformation kernel's native space (Section 5.2) with findings regarding the summation kernel's native space (Section 3.2.1). Our analysis culminates in Theorem 6.7, which states that the native space can orthogonally be decomposed into the component's native spaces, satisfying the name 'orthogonal summation kernel'.

Since the orthogonal summation kernel K is a summation kernel of the transformation kernels K_{ℓ,p_ℓ} for $\ell = 1, \dots, M$, see (6.1), we can use Theorem 3.9 to deduce

$$\mathcal{H}_{K,\Omega} = \sum_{\ell=1}^{M} \mathcal{H}_{K_{\ell,p_\ell},\Omega} = \left\{ \sum_{\ell=1}^{M} f_\ell \ : \ f_\ell \in \mathcal{H}_{K_{\ell,p_\ell},\Omega} \right\}, \tag{6.3}$$

equipped with the norm

$$\|f\|_{K,\Omega}^2 = \min \left\{ \sum_{\ell=1}^{M} \|f_\ell\|_{K_{\ell,p_\ell},\Omega}^2 \right\}, \tag{6.4}$$

where the minimum is taken over all decompositions $f = \sum_{\ell=1}^{M} f_\ell$ with $f_\ell \in \mathcal{H}_{K_{\ell,p_\ell},\Omega}$ for $\ell = 1, \dots, M$.

As the projections p_ℓ are surjective, we can find $x \in \Omega = \times_{\ell=1}^{M} \Omega_\ell$ for every $x^\ell \in \Omega_\ell$ so that $p_\ell(x) = x^\ell$. This shows that a function $f \in \mathcal{H}_{K_{p_\ell},\Omega}$ exists for every $\tilde{f} \in \mathcal{H}_{K,\Omega_\ell}$ so that $f(x) = \tilde{f} \circ p_\ell(x) = \tilde{f}(x^\ell)$ for all $\ell = 1, \dots, M$ and $x \in \Omega$, and all reproducing kernels K. Even if the domain of f is larger, the function $f \in \mathcal{H}_{K_{p_\ell},\Omega}$ itself does not 'see' anything outside Ω_ℓ. We refer to Theorem 5.6 for more details and apply it on each $\mathcal{H}_{K_{\ell,p_\ell},\Omega}$ to obtain

$$\mathcal{H}_{K,\Omega} = \sum_{\ell=1}^{M} \mathcal{H}_{K_{\ell,p_\ell},\Omega} = \sum_{\ell=1}^{M} \mathcal{H}_{K_\ell,\Omega_\ell} \circ p_\ell$$

and

$$\|f\|_{K,\Omega}^2 = \min \left\{ \sum_{\ell=1}^{M} \|f_\ell\|_{K_\ell,\Omega_\ell}^2 \right\},$$

from (6.3) and (6.4), where the minimum is taken over all decompositions $f = \sum_{\ell=1}^{M} f_\ell \circ p_\ell$ with $f_\ell \in \mathcal{H}_{K_\ell,\Omega_\ell}$ for $\ell = 1, \dots, M$.

Theorem 6.5. *Let K_ℓ be positive semi-definite kernels on Ω_ℓ for $\ell = 1, \dots, M$. Furthermore, let $\mathcal{H}_{K_\ell,\Omega_\ell}$ only contain the zero function as constant function for all $\ell = 1, \dots, M$. Then*

$$\mathcal{H}_{K,\Omega} = \bigoplus_{\ell=1}^{M} \mathcal{H}_{K_\ell,\Omega_\ell} \circ p_\ell,$$

where the inner product is given by

$$\langle f, g \rangle_{\mathcal{H}_{K,\Omega}} = \sum_{\ell=1}^{M} \langle f_\ell, g_\ell \rangle_{\mathcal{H}_{K_\ell,\Omega_\ell}},$$

where $f = \sum_{\ell=1}^{M} f_\ell \circ p_\ell$, $g = \sum_{\ell=1}^{M} g_\ell \circ p_\ell$, with $f_\ell, g_\ell \in \mathcal{H}_{K_\ell,\Omega_\ell}$ for $\ell = 1, \dots, M$, are the unique decompositions of $f, g \in \mathcal{H}_{K,\Omega}$.

Proof. Theorem 3.9 implies, that the representation of a function of the space $\mathcal{H}_{K,\Omega}$ is unique if and only if the component native spaces $\mathcal{H}_{K_{\ell,p_\ell},\Omega}$ are complemented (Definition 3.5). To analyze this, let

$$f \in \mathcal{H}_{K_{m,p_m},\Omega} \cap \sum_{\ell=1,\ell\neq m}^{M} \mathcal{H}_{K_{\ell,p_\ell},\Omega} \qquad \text{for } m \in \{1,\dots,M\} \tag{6.5}$$

then f can be represented in two ways

$$g_m \circ p_m(x) = f(x) = \sum_{\ell=1,\ell\neq m}^{M} g_\ell \circ p_\ell(x) \qquad \text{for all } x \in \Omega \text{ and } m \in \{1,\dots,M\}, \tag{6.6}$$

where $g_\ell \in \mathcal{H}_{K_\ell,\Omega_\ell}$ for all $\ell = 1,\dots,M$. For two arbitrary values x, y in Ω that only differ in the m-th component, i.e., $x - y = p_m(x) - p_m(y)$, the representation of f on the right-hand side of (6.6) does not change. Therefore, $g_m \circ p_m$ is a constant function on Ω, which implies that $g_m \in \mathcal{H}_{K_m,\Omega_m}$ is constant function on Ω_m. Since the only constant function of $\mathcal{H}_{K_m,\Omega_m}$ is the zero function for all $m = 1,\dots,M$, it is $0 \equiv g_m \circ p_m = f$. As $m \in \{1,\dots,M\}$ was chosen arbitrarily, the additional assumption of Theorem 3.9 is satisfied, which finishes the proof. ■

Remark 6.6. We want to state two examples of native spaces, that only contain the zero function as constant function.

1. For a translation-invariant kernel K on the whole space $\mathbb{R}^d$, it is $\mathcal{H}_{K,\mathbb{R}^d} \subset L^2(\mathbb{R}^d)$, by Theorem 2.30, and $L^2(\mathbb{R}^d)$ only contains the zero function as a constant function. Consequently, $\mathcal{H}_{K,\mathbb{R}^d}$ only contains the zero function as a constant function.

2. Remark 2.36 and the reference therein state that the only constant function of the Gaussian's native space, for a set Ω with non-empty interior, is the zero function.

In the following, we explain how the kernel derives its name. If the native spaces of all component kernels contain only the zero function as a constant function, no further demonstration is needed, and we refer to Theorem 6.5. Therefore, we focus on the case where at least one component kernel has a native space that contains nonzero constant functions. Let K_ℓ be positive semi-definite kernels on Ω_ℓ for $\ell = 1,\dots,M$ and

$$K = \sum_{\ell=1}^{M} K_{\ell,p_\ell}$$

their orthogonal summation kernel on $\Omega = \times_{\ell=1}^{M} \Omega_\ell$. Without loss of generality, we assume $\mathcal{H}_{K_{M,p_M},\Omega}$ to contain nonzero constant functions. Furthermore, let $m \in \{2,\dots,M\}$ and

$$\tilde{K}_m = \sum_{\ell=m}^{M} K_{\ell,p_\ell}$$

on $\tilde{\Omega}_m = \times_{\ell=m}^{M} \Omega_\ell$ be the orthogonal summation kernel of K_ℓ for $\ell = m,\dots,M$ and $\tilde{p}_m$ denote the projection from Ω onto $\tilde{\Omega}_m$.

1. Step: If $\mathcal{H}_{K_{1,p_1},\Omega}$ only contains the zero function as constant function, no further action is needed. In this case, we directly obtain (6.8) with $k_1 = K_1$ and continue with the second step.

Otherwise, we proceed as follows. From the above analysis, we deduce that the intersection $\mathcal{H}_{K_{1,p_1},\Omega} \cap \mathcal{H}_{\tilde{K}_{2,\tilde{p}_2},\Omega}$ equals the set of constant functions on Ω. Hence, the corresponding intersection kernel is a constant positive semi-definite kernel. Let

$$\mathbf{c_1}(x, y) = c_1 \qquad \text{for all } x, y$$

denote the intersection kernel for a constant $c_1 \in \mathbb{R}$. Furthermore, Theorem 3.26 provides the existence of positive semi-definite kernels κ_1 and κ_2 on Ω so that

$$K_{1,p_1} = \mathbf{c_1} + \kappa_1, \qquad \tilde{K}_{2,\tilde{p}_2} = \mathbf{c_1} + \kappa_2 \qquad \text{and} \qquad \mathcal{H}_{\kappa_1,\Omega} \cap \mathcal{H}_{\kappa_2,\Omega} = \{0\}. \tag{6.7}$$

Without loss of generality, we deduce that $\mathcal{H}_{\kappa_1,\Omega}$ contains only the zero function as a constant function. We can make this conclusion because if both of the native spaces $\mathcal{H}_{\kappa_1,\Omega}$ and $\mathcal{H}_{\kappa_2,\Omega}$ contain a nonzero constant function, then each must contain the entire set of constant functions on Ω, as they are vector spaces. This would contradict their trivial intersection. With this understanding, we obtain the following equation

$$\mathcal{H}_{K_{1,p_1},\Omega} = \mathcal{H}_{\mathbf{c_1},\Omega} \oplus \mathcal{H}_{\kappa_{1,p_1},\Omega}.$$

Since $\mathcal{H}_{\kappa_1,\Omega} \subset \mathcal{H}_{K_{1,p_1},\Omega} = \mathcal{H}_{K_1,\Omega_1} \circ p_1$, the native space of κ_1 contains only functions depending on Ω_1. By Lemma 5.7, there exists a kernel k_1 on Ω_1 so that $\kappa_1 = k_{1,p_1}$. This finding, the representation of K_{1,p_1} in (6.7), and the definition of K yield

$$K = K_{1,p_1} + \tilde{K}_{2,\tilde{p}_2} = \mathbf{c_1} + k_{1,p_1} + \tilde{K}_{2,\tilde{p}_2}.$$

Because of (6.7), the relation $\mathbf{c_1} \lesssim \tilde{K}_{2,\tilde{p}_2}$ is satisfied and Theorem 3.25 implies

$$K \sim k_{1,p_1} + \tilde{K}_{2,\tilde{p}_2}.$$

Since $\mathcal{H}_{\kappa_1,\Omega} = \mathcal{H}_{k_{1,p_1},\Omega}$ only contains the zero function as constant function the intersection $\mathcal{H}_{k_{1,p_1},\Omega} \cap \mathcal{H}_{\tilde{K}_{2,\tilde{p}_2},\Omega}$ is trivial. Consequently, Theorem 3.9 and Lemma 3.22 provide

$$\mathcal{H}_{K,\Omega} \sim \mathcal{H}_{k_{1,p_1},\Omega} \oplus \mathcal{H}_{\tilde{K}_{2,\tilde{p}_2},\Omega}. \tag{6.8}$$

2. Step: We repeat the first step for K_{2,p_2} and $\tilde{K}_{3,\tilde{p}_3}$ to obtain

$$\mathcal{H}_{\tilde{K}_{2,\tilde{p}_2},\Omega} \sim \mathcal{H}_{k_{2,p_2},\Omega} \oplus \mathcal{H}_{\tilde{K}_{3,\tilde{p}_3},\Omega}, \tag{6.9}$$

where $\mathcal{H}_{k_{2,p_2},\Omega}$ only contains the zero function as constant function and k_{2,p_2} is given by the orthogonal decomposition $K_{2,p_2} = \mathbf{c_2} + k_{2,p_2}$ for a constant kernel $\mathbf{c_2}$ that is possibly zero. The equation (6.9) applied on (6.8) yields

$$\mathcal{H}_{K,\Omega} \sim \mathcal{H}_{k_{1,p_1},\Omega} \oplus \mathcal{H}_{k_{2,p_2},\Omega} \oplus \mathcal{H}_{\tilde{K}_{3,\tilde{p}_3},\Omega}.$$

3. Step: We repeat the above two steps until $K_{M-1,p_{M-1}}$ and $\tilde{K}_{M,\tilde{p}_M} = K_{M,p_M}$ to obtain the following result.

Theorem 6.7. *Let K_ℓ be positive semi-definite kernels on Ω_ℓ for $\ell = 1, \dots, M$ and K their orthogonal summation kernel on $\Omega = \times_{\ell=1}^{M} \Omega_\ell$. Then the orthogonal summation kernel's native spaces satisfies*

$$\mathcal{H}_{K,\Omega} \sim \bigoplus_{\ell=1}^{M} \mathcal{H}_{k_{\ell,p_\ell},\Omega},$$

where k_{ℓ,p_ℓ} either equals the component K_{ℓ,p_ℓ} or is given by the orthogonal decomposition $K_{\ell,p_\ell} = \mathbf{c}_\ell + k_{\ell,p_\ell}$ for a constant kernel $\mathbf{c}_\ell$ for every $\ell = 1, \dots, M$.

In summary, this demonstrates that every function $f \in \mathcal{H}_{K,\Omega}$ has a unique representation $f = \sum_{\ell=1}^{M} f_\ell \circ p_\ell$, where $f_\ell \in \mathcal{H}_{k_\ell,\Omega_\ell}$ and the kernel k_ℓ equals the component K_ℓ of K, up to a constant (possibly zero) addend for all $\ell = 1, \dots, M$. With the construction of the kernels k_ℓ as outlined above, we decompose the function f so that its constant part is not distributed across multiple components f_ℓ, but appears solely in the last component f_M. Furthermore, from Theorem 6.7, we obtain that the norm of the orthogonal summation kernel's native space $\|f\|_{K,\Omega}$ is equivalent to $\sum_{\ell=1}^{M} \|f_\ell\|_{k_\ell,\Omega_\ell}$, and

$$\|f_\ell\|_{k_\ell,\Omega_\ell} = \|f_\ell\|_{K_\ell,\Omega_\ell} \qquad \text{for all } f_\ell \in \mathcal{H}_{k_\ell,\Omega_\ell},$$

because of the fact that the decomposition $K_{\ell,p_\ell} = \mathbf{c}_\ell + k_{\ell,p_\ell}$ is orthogonal and Theorem 3.12. Thus, we conclude that $\|f\|_{K,\Omega}$ is equivalent to $\sum_{\ell=1}^{M} \|f_\ell\|_{K_\ell,\Omega_\ell}$.

Even though this is a slight abuse of notation, it justifies the name of the kernel. Particularly because kernels belonging to the same equivalence class have similar characteristics with respect to interpolation, as shown in Chapter 3. In the following, we assume that the native space of the orthogonal summation kernel is the orthogonal sum of the native spaces of its components.

At the end of this section, we want to emphasize that not only does the orthogonal summation case follow from the standard summation case, but there is also a reciprocal relationship between them, schematically visualized in Fig. 6.1. The summation kernel can be considered as a special case of the orthogonal summation kernel. If $\Omega_\ell = \Omega$ for every ℓ, the summation kernel $\tilde{K}(x, y) = \sum_\ell K_\ell(x, y)$ equals the restriction of the orthogonal summation kernel $K(x, y)$ to the diagonal set $\{(x, \dots, x) : x \in \Omega\} \subset \times_\ell \Omega$, i.e.,

$$\begin{aligned} K\big((x, \dots, x), (y, \dots, y)\big) &= \sum_\ell K_\ell\big(p_\ell(x, \dots, x), p_\ell(y, \dots, y)\big) \\ &= \sum_\ell K_\ell(x, y) = \tilde{K}(x, y) \qquad \text{for all } x, y \in \Omega. \end{aligned}$$

In view of the native space, we loose orthogonality by taking the restriction, resulting in the findings of Theorem 3.9. For more details on the restriction of reproducing kernels we refer to [Aro50, Part 1, §5] and [Wen05, Chapter 10.7].

6.3 Interpolation

Here, we examine the interpolation using orthogonal summation kernels. First, we discuss the conditions under which the interpolation matrix $\mathbf{A}_{K,X}$ is positive definite. We then examine the approximation error in Section 6.3.1 and the numerical stability in Section 6.3.2.

We aim for a positive definite interpolation matrix $\mathbf{A}_{K,X}$, as described in Section 2.1. In Ex. 6.2, we observed that in regard to the orthogonal summation kernel, it is not sufficient to demand positive definiteness of the component kernels alone. This leads us to the concept of DC-strictly positive definite kernels, which was introduced by J.C. Guella and V.A. Menegatto in [GM16].

Definition 6.8. Let K be a kernel on the Cartesian product $\times_{\ell=1}^{M} \Omega_\ell$. The kernel K is called *DC-strictly positive definite* (distinct component) if its interpolation matrix $\mathbf{A}_{K,X}$ is positive definite for every point set $X \subset \times_{\ell=1}^{M} \Omega_\ell$ with pairwise distinct projections $p_\ell(X) \subset \Omega_\ell$ for $\ell = 1, \ldots, M$, where p_ℓ denotes the projection from Ω to Ω_ℓ.

We directly observe that an orthogonal summation kernel with positive definite components is DC-strictly positive definite. In the following, we minimize the requirements for a positive definite interpolation matrix, noting that positive definiteness of at least one component kernel K_ℓ and a corresponding pairwise distinct projection $p_\ell(X)$ of X is sufficient.

Lemma 6.9. *Let K_ℓ be positive semi-definite kernels on Ω_ℓ for $\ell = 1, \ldots, M$ and K their orthogonal summation kernel on $\Omega = \times_{\ell=1}^{M} \Omega_\ell$. Furthermore, let $X \subset \Omega$ be a finite point set. If there exist an index $\ell_0 \in \{1, \ldots, M\}$ satisfying that*

1. *the projection $p_{\ell_0}(X)$ of X onto Ω_{ℓ_0} is pairwise distinct and*
2. *K_{ℓ_0} is a positive definite kernel on Ω_{ℓ_0},*

then the interpolation matrix $\mathbf{A}_{K,X}$ regarding the orthogonal summation kernel K is positive definite.

Proof. By the assumptions 1. and 2., we have $c^T \mathbf{A}_{K_{\ell_0}, p_{\ell_0}(X)} c > 0$. This yields

$$c^T \mathbf{A}_{K,X} c = c^T \mathbf{A}_{K_{\ell_0}, p_{\ell_0}(X)} c + \sum_{\ell=1, \ell \neq \ell_0}^{M} c^T \mathbf{A}_{K_\ell, p_\ell(X)} c > 0.$$

■

Next, we state the anisotropic version of Lemma 3.37.

Lemma 6.10. *Let K_ℓ be positive definite kernels on Ω_ℓ for $\ell = 1, \ldots, M$ and K their orthogonal summation kernel on $\Omega = \times_{\ell=1}^{M} \Omega_\ell$. Furthermore, let $X \subset \Omega$ be a finite point set, such that its projections $p_\ell(X)$ onto Ω_ℓ are pairwise distinct for all $\ell = 1, \ldots, M$. Let the target function $f \in \mathcal{H}_{K,\Omega}$ have the unique representation $f = \sum_{\ell=1}^{M} f_\ell \circ p_\ell$ for $f_\ell \in \mathcal{H}_{K_\ell,\Omega_\ell}$. Then,*

$$\left\| s_{f,K,X} \right\|_{K,\Omega}^2 \leqslant \sum_{\ell=1}^{M} \left\| s_{f_\ell, K_\ell, p_\ell(X)} \right\|_{K_\ell,\Omega_\ell}^2 .$$

Proof. For the proof, we view the orthogonal summation kernel K as the summation kernel of the transformation kernel K_{ℓ,p_ℓ} acting on Ω, see (6.1). We apply Lemma 3.37 to obtain

$$\left\|s_{f,K,X}\right\|_{K,\Omega}^2 \leqslant \sum_{\ell=1}^{M} \left\|s_{f_\ell \circ p_\ell, K_{\ell,p_\ell}, X}\right\|_{K_{\ell,p_\ell},\Omega}^2 .$$

By Lemma 5.8, the right-hand side can be written as

$$\left\|s_{f,K,X}\right\|_{K,\Omega}^2 \leqslant \sum_{\ell=1}^{M} \left\|s_{f_\ell, K_\ell, p_\ell(X)}\right\|_{K_\ell,p_\ell(\Omega)}^2 .$$

■

6.3.1 Approximation Error

This section is concerned with the approximation error when using orthogonal summation kernels. We state the anisotropic version of Theorem 3.38 in Theorem 6.11, which demonstrates that an improved error can be expected by performing interpolation component-wise first and then summing the results in a second step. For further insights, see the numerical example in Section 6.4. Additionally, we outline the relation between the power function of an orthogonal summation kernel and that of its components, and we consider translation-invariant component kernels in Lemma 6.12.

Theorem 6.11. *Let K_ℓ be positive definite kernels on Ω_ℓ for $\ell = 1, \dots, M$ and K their orthogonal summation kernel on $\Omega = \times_{\ell=1}^{M} \Omega_\ell$. Furthermore, let $X \subset \Omega$ be a finite point set, such that its projections $p_\ell(X)$ onto Ω_ℓ are pairwise distinct for all $\ell = 1, \dots, M$, and let the target function $f \in \mathcal{H}_{K,\Omega}$ have the unique representation $f = \sum_{\ell=1}^{M} f_\ell \circ p_\ell$, with $f_\ell \in \mathcal{H}_{K_\ell,\Omega_\ell}$ for $\ell = 1, \dots, M$. Then,*

$$\sum_{\ell=1}^{M} \left\|f_\ell - s_{f_\ell, K_\ell, p_\ell(X)}\right\|_{K_\ell,\Omega_\ell}^2 \leqslant \left\|f - s_{f,K,X}\right\|_{K,\Omega}^2 .$$

Proof. Because of (6.1) and Lemma 5.8, it is

$$\sum_{\ell=1}^{M} \left\|f_\ell - s_{f_\ell, K_\ell, p_\ell(X)}\right\|_{K_\ell,\Omega_\ell}^2 = \sum_{\ell=1}^{M} \left\|f_\ell \circ p_\ell - s_{f_\ell \circ p_\ell, K_{\ell,p_\ell}, X}\right\|_{K_{\ell,p_\ell},\Omega}^2 .$$

Since the component native spaces $\mathcal{H}_{K_{\ell,p_\ell},\Omega}$ are complemented, we are in the position to apply Theorem 3.38 on the right-hand side, to obtain

$$\sum_{\ell=1}^{M} \left\|f_\ell - s_{f_\ell, K_\ell, p_\ell(X)}\right\|_{K_\ell,\Omega_\ell}^2 \leqslant \left\|f - s_{f,K,X}\right\|_{K,\Omega}^2 .$$

■

The above combined with Definition 2.27 of the power function yields

$$\sum_{\ell=1}^{M} P_{K_\ell, p_\ell(X)}^2 \leqslant P_{K,X}^2 .$$

Lemma 6.12. *Let K_ℓ be positive definite translation-invariant kernels on Ω_ℓ with univariate functions Φ_ℓ for $\ell = 1, \dots, M$. Furthermore, let Ω_ℓ fulfill an ICC and $p_\ell(X)$ be pairwise distinct and satisfy $h_{p_\ell(X),\Omega_\ell} \leqslant h_{\ell,0}$ for $\ell = 1, \dots, M$. If K is the orthogonal summation kernel of K_ℓ for $\ell = 1, \dots, M$, its power function can be bounded by*

$$P_{K,X}^2(x) \leqslant \sum_{\ell=1}^{M} F_{\Phi_{\ell,p_\ell}\Omega}(h_{p_\ell(X),\Omega_\ell}) \qquad \text{for all } x \in \Omega = \bigtimes_{\ell=1}^{M} \Omega_\ell,$$

where the functions $F_{\Phi_{\ell,p_\ell}\Omega}$ come from Remark 2.44.

Proof. The statement is a direct consequence of (6.1) and Lemma 3.39. ∎

6.3.2 Numerical Stability

Subsequently, we present the anisotropic version of Theorem 3.40, which demonstrates that the numerical stability of an orthogonal summation kernel aligns with that of its most stable component kernel. For a numerical example, see Section 6.4.

Theorem 6.13. *Let K_ℓ be positive definite kernels on Ω_ℓ for $\ell = 1, \dots, M$, K be their orthogonal summation kernel on $\Omega = \bigtimes_{\ell=1}^{M} \Omega_\ell$, and $X \subset \Omega$ be a point set such that its projections $p_\ell(X) \subset \Omega_\ell$ are pairwise distinct. Then the following statements hold.*

(i)

$$\lambda_{\min}\left(\mathbf{A}_{K,X}\right) \geqslant \max_{\ell \in \{1,\dots,M\}} \left\{\lambda_{\min}\left(\mathbf{A}_{K_\ell,p_\ell(X)}\right)\right\}$$

(ii)

$$\mathrm{cond}_2(\mathbf{A}_{K,X}) \leqslant \frac{M \max_{\ell \in \{1,\dots,M\}} \left\{\lambda_{\max}(\mathbf{A}_{K_\ell,p_\ell(X)})\right\}}{\max_{\ell \in \{1,\dots,M\}} \left\{\lambda_{\min}(\mathbf{A}_{K_\ell,p_\ell(X)})\right\}}.$$

(iii) Let $\ell_0 \in \{1, \dots, M\}$ satisfy

$$\lambda_{\min}\left(\mathbf{A}_{K_{\ell_0},p_\ell(X)}\right) = \max_{\ell \in \{1,\dots,M\}} \left\{\lambda_{\min}\left(\mathbf{A}_{K_\ell,p_\ell(X)}\right)\right\}$$

and K_{ℓ_0} be a translation-invariant kernel on Ω_{ℓ_0} with univariate function Φ_{ℓ_0} such that $\widehat{\Phi_{\ell_0}} \in C(\mathbb{R}^{d_{\ell_0}} \backslash \{0\})$. Then

$$\lambda_{\min}\left(\mathbf{A}_{K,X}\right) \geqslant G_{\Phi_{\ell_0}}(q_{p_\ell(X)}),$$

where the function $G_{\Phi_{\ell_0}}$ comes from Remark 2.49.

(iv) Let K be translation-invariant and its univariate functions Φ satisfy $\widehat{\Phi} \in C(\mathbb{R}^{d_\ell} \backslash \{0\})$. Then

$$\lambda_{min}(\mathbf{A}_{K,X}) \geqslant G_\Phi \left(\min_\ell \{q_{p_\ell(X)}\}\right),$$

where the functions G_Φ comes from Remark 2.49.

Proof. For the statements $(i) - (iii)$, we view the orthogonal summation kernel as a summation kernel of transformation kernels, see (6.1). We combine the results of Theorem 3.40 on summation kernels with those from Theorem 5.13 and (5.1) on transformation kernels. For statement (iv), we use the fact that the function G_Φ is monotonously increasing and

$$\begin{aligned} q_X^2 &= \frac{1}{4} \min_{i \neq j} \left\| x_i - x_j \right\|_2^2 \\ &= \frac{1}{4} \min_{i \neq j} \sum_{\ell=1}^{M} \left\| p_\ell(x_i - x_j) \right\|_2^2 \\ &\geqslant \frac{1}{4} \min_{i \neq j} \min_{\ell} \left\| p_\ell(x_i - x_j) \right\|_2^2 \\ &= \min_{\ell} q_{p_\ell(X)}^2. \end{aligned}$$

■

6.4 Numerical Tests

In the following, we present numerical examples that support the theoretical results discussed in the preceding sections. We compare interpolation using orthogonal summation kernels with interpolation using radially symmetric kernels. Our findings demonstrate that an orthogonal summation kernel, when adapted to a target function with an anisotropic sum structure, yields outstanding results in terms of both approximation error and numerical stability.

For the numerical tests, we consider

- the domain $\Omega = \Omega_1 \times \Omega_2$, with $\Omega_\ell = [0, 1]$, $\ell = 1, 2$ and
- the developing point sets X_n, consisting of 2^n random points in Ω for $n = 4, \ldots, 10$, and satisfying $X_m \subseteq X_n$ for $m \leqslant n$. Additionally, the projections $p_\ell(X_n)$ are pairwise distinct for all $\ell = 1, 2$ and $n = 4, \ldots, 10$.

To emphasize the benefits and challenges of using an orthogonal summation kernel, we reconstruct the following two target functions, which are visualized in Fig. 6.4 along with the point set X_7:

1. the C^∞ target function f_{franke} defined in (3.24) and visualized in Fig. 6.4 (left).
2. the C^0 target function

$$f_{\text{aniso-s}}(x^1, x^2) := f_1(x^1) + f_2(x^2) = 3\left\| x^1 - 0.5 \right\| + \sin(2\pi x^2) \qquad (6.10)$$

   for $(x^1, x^2) \in \Omega$, with anisotropic sum structure, visualized in Fig. 6.4 (right).

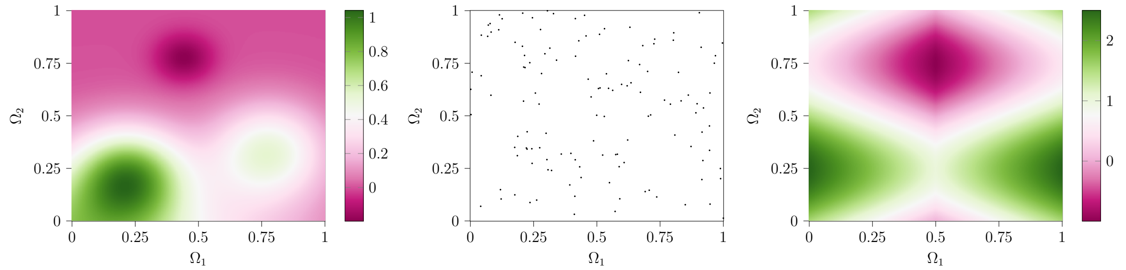

Figure 6.4: Top view of the target functions f_{franke} (left) and $f_{\text{aniso-s}}$ (right), and the scattered point set X_7 in Ω (middle).

We compare the performance of

- the two-dimensional Wendland kernel K_1 with RBF $\phi_{3,0}$ and
- the two-dimensional Wendland kernel K_2 with RBF $\phi_{3,3}$ with
- the two-dimensional orthogonal summation kernel

$$K\left((x^1, x^2), (y^1, y^2)\right) = \kappa_1(x^1, y^1) + \kappa_2(x^2, y^2) \qquad \text{for } (x^1, x^2), (y^1, y^2) \in \Omega,$$

visualized in Fig. 6.5, of the one-dimensional Wendland kernels κ_1 and κ_2 corresponding to $\phi_{1,0}$ and $\phi_{1,3}$.

When reconstructing the target $f_{\text{aniso-s}}$, we add another approach to gain an interpolant to our comparison, namely

- the anisotropic summation of interpolants (ASI). Here, we use the projections $p_\ell(X_n)$ and the function values $f_{\ell,p_\ell(X_n)}$ of the one-dimensional functions f_ℓ on $p_\ell(X_n)$ for $\ell = 1, 2$. The interpolation is carried out separately with κ_1 on Ω_1 and κ_2 on Ω_2 first, and summed up afterwards. This procedure results in the interpolant

$$s(x) = s(x^1, x^2) = s_{f_1,\kappa_1,p_1(X)}(x^1) + s_{f_2,\kappa_2,p_2(X)}(x^2) \quad \text{for all } x = (x^1, x^2) \in \Omega.$$

We remark that, we need more precise information to perform this approach, as two function values, $f_1(x^1)$ and $f_2(x^2)$ are needed for one interpolation point $x = (x^1, x^2) \in \Omega$.

Note, that we chose κ_1 to be a non-differentiable kernel mirroring f_1 of (6.10) and κ_2 to be C^3 capturing the smoothens of f_2 of (6.10).

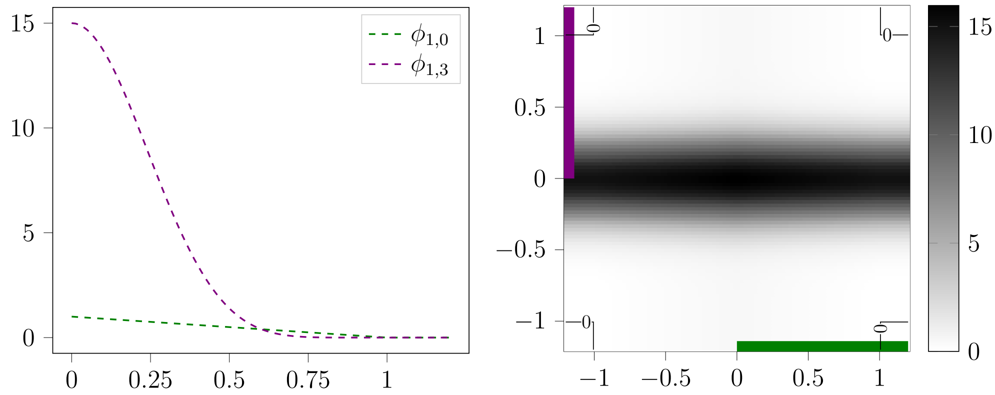

Figure 6.5: Visualization of two Wendland RBFs $\phi_{1,0}$ and $\phi_{1,3}$ (left), and their corresponding two-dimensional orthogonal summation kernel (right) with its support (black line). The green and purple bar emphasize that $\phi_{1,0}$ acts in x-direction and $\phi_{1,3}$ in y-direction.

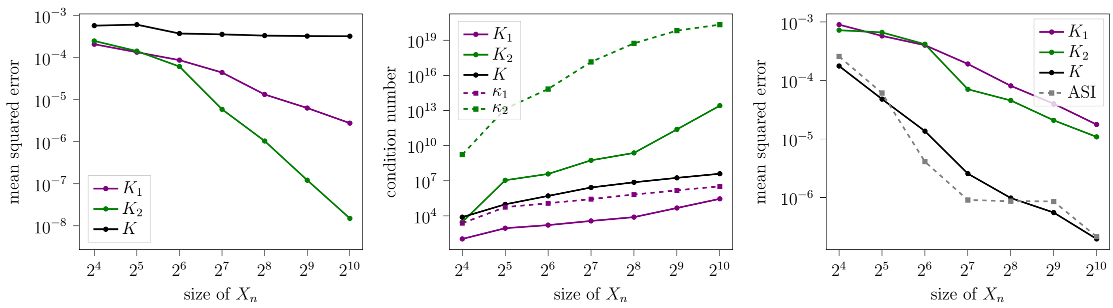

Figure 6.6: Comparison of the kernels K_j, $j = 1, 2$, and their orthogonal summation kernel K regarding the condition number (middle) and the mean squared error corresponding to the target function f_{franke} (left) and $f_{\text{aniso-s}}$ (right) on developing point sets X_n for $n = 4, \ldots, 10$.

Let us devote our attention to the interpolation with $f_{\text{aniso-s}}$, first. In Fig. 6.6 (right), we observe that the two summation approaches (K and ASI) are able to improve the approximation by two digits, compared to the radial approaches (K_1 and K_2). For the ASI approach this supports Theorem 6.11. Furthermore, we see that the interpolant corresponding to the orthogonal summation kernel K approximates the target function just as fine as the ASI approach, even though the latter requires more detailed information. This finding favors the orthogonal summation kernel.

In view of the numerical condition number in Fig. 6.6 (middle), we see that the one of the orthogonal summation kernel K rises with the same rate as its most stable component kernel κ_1, supporting the first statement of Theorem 6.13. These observations reflect well on the interpolation with the orthogonal summation kernel K.

However, looking at Fig. 6.6 (left), the interpolation of the Franke function challenges the orthogonal summation kernel. The error is significantly worse than the one corresponding to the radial symmetric approaches K_1 and K_2. Furthermore, it hardly improves with the amount of interpolation points increasing.

For more details, we visualize the absolute error

$$|\text{interpolant}(x) - f(x)| \qquad \text{for } x \in [0,1]^2$$

for the interpolants of the Franke function f_{franke} corresponding to the point set X_7 and the interpolation kernels K_1, K_2 and the orthogonal summation kernel K in Figure 6.7. The plotted error supports the result of Fig. 6.6 (left) as the radially symmetric approaches (left and middle of Figure 6.7) perform much better than the orthogonal summation kernel K (right). The bad error and its top-bottom-line structure, is due to the shape of K, particularly its unbounded support, see Fig. 6.5.

Looking at the absolute error between the corresponding interpolants and the target function $f_{\text{aniso-s}}$ with anisotropic sum structure in Fig. 6.8, the superiority of the summation approaches (K and ASI) is salient, as already observed in Fig. 6.6 (right). We see that the worst error in the summation cases does not occur at the biggest hole of the data, as it is the case for the radial approaches (Fig. 6.8 – top row), but along the kink (Fig. 6.8 – bottom row).

In summary, it can be said that the orthogonal summation kernel K outperforms radial approaches (K_1 and K_2) when it comes to target functions with an anisotropic sum structure and known properties of each target summand, allowing the adaptation of the generic kernels. In this case, the approximation error rate is significantly improved, and the numerical condition number grows with a rate depending on the most stable component kernel. Furthermore, it can be observed that even though the ASI approach does contain more detailed information, it does not improve the approximation compared to the interpolant corresponding to the orthogonal summation kernel.

We recommend the use of orthogonal summation kernels for targets with an anisotropic summation structure.

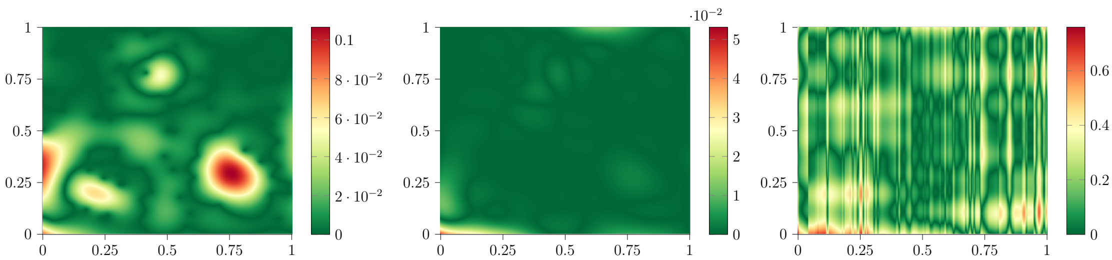

Figure 6.7: Absolute error between the target function f_{franke} and the interpolants regarding the point set X_7 corresponding to K_1 (left), K_2 (middle) and the orthogonal summation kernel K (right).

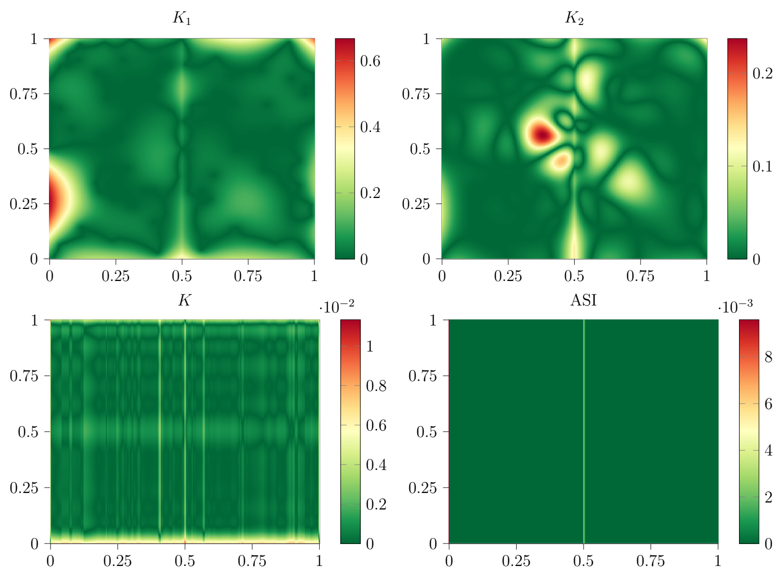

Figure 6.8: Absolute error between the target function $f_{\text{aniso-s}}$ and the interpolants regarding the point set X_7 corresponding to K_1 (top left), K_2 (top right), the orthogonal summation kernel K (bottom left) and the ASI approach (bottom right).

Chapter 7

Tensor Product Kernels

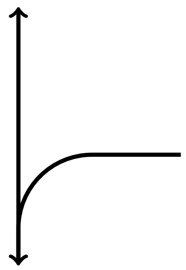

Figure 7.1: Schematic relation between product, transformation, and tensor product kernels.

In this section we naturally combine the advantages of interpolation methods by exploiting an underlying structure of the interpolation kernel. As an initial approach to tackle the lack of adaptive anisotropic kernels, transformation kernels are considered in Chapter 5. In particular, it is discussed in Section 5.1 that anisotropic kernels can be constructed by replacing the standard Euclidean norm with an anisotropic norm

$$\|x\|_A^2 := x^T A x \qquad \text{for } x \in \mathbb{R}^d,$$

in the argument of a fixed radial kernel function, where $A \in \mathbb{R}^{d \times d}$ is a symmetric positive definite matrix. For a standard Gaussian kernel and a positive definite diagonal matrix $A = \operatorname{diag}(\alpha_1, ..., \alpha_d)$, this construction yields the kernel

$$\mathrm{e}^{-\|x\|_A^2} = \mathrm{e}^{-x^T A x} = \prod_{\ell=1}^{d} \mathrm{e}^{-\alpha_\ell x_\ell^2} \qquad \text{for } x = (x_1, ..., x_d)^T \in \mathbb{R}^d,$$

see [Fas11]. Hence, this anisotropic version of the standard Gaussian kernel is given by a product of d kernels acting on one dimension, and each of them is equipped with its own shape parameter $\alpha_\ell > 0$ for $\ell = 1, \ldots, d$. This observation gives reason to study kernels that are products of positive definite component kernels defined on different lower-dimensional spaces. These kernels are referred to as *tensor product kernels*. The main intention of this approach is to further improve the flexibility of kernel-based reconstruction methods, as the initial domain can be split into several subdomains and each of them can be equipped with an individual kernel tailored to the application.

In [Saa11], it is demonstrated that tensor product kernels hold another advantage. Many real-world applications are structured as a grid. Climate data sets serve as an example, as they record measurements of variables such as ocean surface temperatures and CO_2 concentrations across a grid of locations spanning the Earth's surface. The specific example of [RBP+06] involves sea surface temperature data organized as a grid of geographical locations over discretized time-series observations. Hence, the entire dataset can be viewed as a three-dimensional Cartesian grid. When interpolation is conducted on a gridded point set, the interpolation matrix of the tensor product kernel can be expressed as the Kronecker product of its component's interpolation matrices. Y. Saatci, showed in [Saa11] that this speeds up the Gaussian process in applications where the number of interpolation points approaches millions, while the quantity of data points within a subdomain remains comparatively modest.

It turns out that tensor product kernels exhibit tensor product structure, satisfying the name. A significant part of this finding draws from the work [Aro50] of N. Aronszajn, who demonstrated that the native space of product kernels aligns with restrictions of the tensor product of the corresponding component native spaces with a well-defined inner product (Theorem 4.5). Presently, this type of tensor product is known as the Hilbert tensor product, as discussed in [KR83]. It enables us to present the result of J. Neveu in [Nev71] with this term: The Hilbert tensor product of the component kernel's native spaces is given by the native space of the corresponding tensor product kernel. Unlike the standard tensor product, the Hilbert tensor product of Hilbert spaces results in a Hilbert space. Moreover, the inner product of tensors is given by the multiplication of the inner products of the component spaces. This amounts to the fact that the product kernel of Chapter 4 can be seen as a special case of the tensor product kernel, schematically visualized in Fig. 7.1.

Our work [AEI23b], conducted in collaboration with K. Albrecht, extends existing research by providing a more detailed exploration of tensor product kernels and emphasizing their computational advantages over standard kernels. Our main contributions to this field are:

- Elucidating the Structure of Native Spaces: We provide an alternative proof for the structure of the tensor product kernel's native space in Section 7.2.2, drawing a connection from Neveu's work to Hilbert tensor spaces. This connection paves the ground for further insights, such as the tensor construction of the Newton basis of (7.6).

- Detailed Analysis of Interpolation with Tensor Product Kernels: We conduct a thorough examination of interpolation using tensor product kernels in Section 7.3, including an analysis of convergence (Theorem 7.23) and stability (Section 7.3.3). This analysis provides valuable insights into the performance and robustness of tensor product kernel employing reconstruction methods.

- Proof of Positive Definiteness: We present a rigorous proof regarding the positive definiteness of tensor product kernels, depending on the concept of grid-like structured point sets (Definition 7.12). To the best of our knowledge, this proof is the first of its kind and shows that the tensor product kernel of positive definite component kernels remains positive definite (Theorem 7.25).

These contributions collectively advance the theoretical understanding and practical applications of tensor product kernels in reconstruction methods.

This chapter proceeds as follows. Section 7.1 introduces tensor product kernels as a special type of positive semi-definite kernels, providing a comprehensive overview of their properties and characteristics. We delve into theoretical fundamentals and characterize native spaces of tensor product kernels in Section 7.2, drawing from seminal works in the field. Section 7.3 is concerned with the interpolation method using tensor product kernels. We first emphasize the computational advantages inherent in the tensor product structure, particularly when the interpolation point set is grid-like structured (Section 7.3.1). Furthermore, we discuss the efficient computation of orthonormal bases, such as the Newton basis, and its impact on convergence rates. We explore the question of positive definiteness of tensor product kernels in Section 7.3.2. Using the concepts of grid-like structure and Kronecker product, we demonstrate how tensor product kernels inherit the positive definiteness of their component kernels. Additionally, we examine the numerical stability of these kernels in Section 7.3.3, which is crucial for practical applications. Numerical examples are presented in Section 7.4 to underscore the efficiency of computing the Newton basis and the adaptive nature of tensor product kernels. These examples offer valuable insights into the potential performance enhancements achievable with tensor product kernels.

7.1 Definition and Basic Properties

This section provides a detailed definition of tensor product kernels. We present a way of viewing these kernels as a product kernel of transformation kernels and state first basic findings regarding positive semi-definiteness and translation-invariance.

Definition 7.1. Let K_ℓ be positive semi-definite kernels on $\Omega_\ell \subseteq \mathbb{R}^{d_\ell}$ for $\ell = 1, \ldots, M$, and $\Omega = \times_{\ell=1}^{M} \Omega_\ell \subseteq \mathbb{R}^d$, where $d = \sum_{\ell=1}^{M} d_\ell$. Then

$$\begin{aligned} K : \Omega \times \Omega &\longrightarrow \mathbb{R}, \\ K(x, y) &:= \prod_{\ell=1}^{M} K_\ell(p_\ell(x), p_\ell(y)) \ \text{ for } x, y \in \Omega \end{aligned}$$

is called a *tensor product kernel*, where $p_\ell : \Omega \longrightarrow \Omega_\ell$ denotes the projection from Ω onto Ω_ℓ for all $\ell = 1, \ldots, M$.

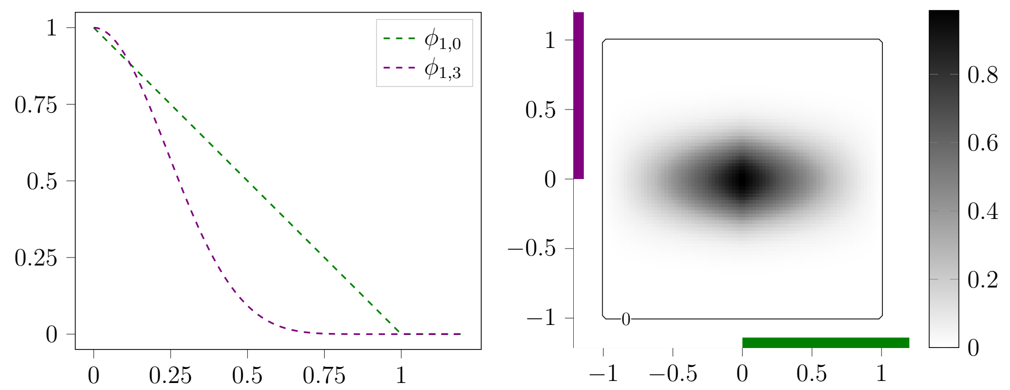

Figure 7.2: Visualization of two normalized Wendland RBFs $\phi_{1,0}$ and $\phi_{1,3}$ (left), and their corresponding two-dimensional tensor product kernel (right) with its support (black line). The green and purple bar emphasize that $\phi_{1,0}$ acts in x-direction and $\phi_{1,3}$ in y-direction.

Figure 7.2 visualizes the anisotropy and construction of a two-dimensional tensor product kernel. As done for the orthogonal summation kernel (Chapter 6), we can write any kernel K acting on a projection p_ℓ as a transformation kernel K_{p_ℓ}, where the transformation is given by projection p_ℓ. This satisfies the schematic relation visualized in Fig. 7.1, i.e.

$$K(p_\ell(x), p_\ell(y)) = K_{p_\ell}(x, y) \qquad \text{for } \ell = 1, \dots, M.$$

Hence, the tensor product kernel can be viewed as a product kernel with transformation kernels as components

$$K(x, y) = \prod_{\ell=1}^{M} K_\ell(p_\ell(x), p_\ell(y)) = \prod_{\ell=1}^{M} K_{\ell p_\ell}(x, y). \tag{7.1}$$

This approach helps to state some basic properties. Given a point set $X \subset \mathbb{R}^d$ and a tensor product kernel K with components K_ℓ for $\ell = 1, \dots, M$, the interpolation matrix $\mathbf{A}_{K,X}$ can be written as the Hadamard product

$$\mathbf{A}_{K,X} = \bigodot_{\ell=1}^{M} \mathbf{A}_{K_{\ell p_\ell},X} = \bigodot_{\ell=1}^{M} \mathbf{A}_{K_\ell, p_\ell(X)}.$$

This is a direct consequence of the interpolation matrix representation of product kernels (4.1) and transformation kernels (5.1). We can proceed as done for the product kernel. The Schur product theorem provides positive semi-definiteness of the product kernels (Theorem 4.3) as well as the tensor product kernels.

Theorem 7.2. *Let K_ℓ be positive semi-definite kernels on Ω_ℓ for $\ell = 1, \dots, M$. Then, the tensor product kernel K is positive semi-definite on $\Omega = \bigtimes_{\ell=1}^{M} \Omega_\ell$.*

It cannot be guaranteed that the projections $p_\ell(X) \subset \mathbb{R}^{d_\ell}$ of a given finite pairwise distinct point set $X \subset \mathbb{R}^d$, where $d = \sum_{d=1}^{M} d_\ell$, are pairwise distinct for every $\ell = 1, \dots, M$. A simple counterexample is given by a grid in $\mathbb{R}^d$, visualized in Fig. 6.3. Hence, the interpolation matrix $\mathbf{A}_{K_\ell, p_\ell(X)}$ of a positive definite component K_ℓ, is not necessarily positive definite. Consequently, the Hadamard product cannot provide positive definiteness for a tensor product kernel with positive definite components, even though the Hermitian of positive definite matrices is positive definite again by the Schur product theorem. This problem is solved by a workaround using a special type of data sets in Section 7.3.2.

However, we can use Bochner's theorem when restricting the component kernels to be translation-invariant. This characterization draws a link between translation-invariant kernels and non-negative Fourier transforms. Subsequently, we state an extension of [Wen05, Proposition 6.25] that provides a positive definite tensor product kernel for positive definite translation-invariant components.

Theorem 7.3. *Let K_ℓ be positive definite kernels on $\mathbb{R}^{d_\ell}$ of the form*

$$K_\ell(x_\ell, y_\ell) = \Phi_\ell(x_\ell - y_\ell) \qquad \text{for } x_\ell, y_\ell \in \mathbb{R}^{d_\ell},$$

for every $\ell = 1, \dots, M$, where $\Phi_\ell \in L_1(\mathbb{R}^{d_\ell}) \cap \mathcal{C}(\mathbb{R}^{d_\ell})$. Then the corresponding tensor product kernel K is positive definite on $\mathbb{R}^d$, with $d = \sum_{\ell=1}^{M} d_\ell$.

We close this section with the statement, that translation-invariant component kernels provide a translation-invariant tensor product kernel.

Lemma 7.4. *If K_ℓ are translation-invariant kernels with uni-variate functions Φ_ℓ on Ω_ℓ for $\ell = 1, \dots, M$, their tensor product kernel K is translation-invariant with the uni-variate function $\Phi = \prod_{\ell=1}^{M} \Phi_\ell \circ p_\ell$.*

Proof. We note that the projection p_ℓ is linear for all $\ell = 1, \dots, M$, and use Lemma 5.3 (ii) to obtain

$$K(x, y) = \prod_{\ell=1}^{M} K_{\ell, p_\ell}(x, y) = \prod_{\ell=1}^{M} \Phi_\ell \circ p_\ell(x - y).$$

■

7.2 Native Spaces

From now on, we assume the components K_ℓ to be positive semi-definite kernels for $\ell = 1, ..., M$, so that their tensor product kernel K is positive semi-definite according to Theorem 7.2. In Section 2.3, we demonstrated that any symmetric positive semi-definite kernel generates a reproducing kernel Hilbert space (RKHS) $\mathcal{H}_{K,\Omega}$ of functions, called native space. As the tensor product kernel K is composed of the different component kernels, we aim to derive a similar relation between the associated native spaces. To do so, we introduce the Hilbert tensor product in Section 7.2.1 and show that the tensor product kernel's native space is such a product in Section 7.2.2.

7.2.1 Hilbert Tensor Product

We introduce and elucidate the Hilbert tensor product here to find that the native space of the tensor product kernel is such in the subsequent Section 7.2.2. This section builds upon the foundations presented in [KR83] and [Hac12], to which we refer for more detailed information.

In general, a tensor product is a mapping that shares the same structure as the algebraic tensor product '$\otimes_a$', which is defined by a quotient vector space and therefore consists of equivalence classes. For a formal definition, we refer to [Hac12, Chapter 3.2], and characterize it as done in [Hac12, Proposition 3.12].

Definition 7.5. Let V, W and U be vector spaces over $\mathbb{R}$. The mapping

$$\otimes : V \times W \longrightarrow U$$

is a tensor product and U a tensor space (i.e., it is isomorphic to $V \otimes_a W$), if the following properties hold:

1. Span property: $U = V \otimes W = \operatorname{span}\{v \otimes w \; : \; v \in V, \; w \in W\}$
2. Bilinearity:

$$\begin{aligned}
&(av) \otimes w = v \otimes aw = a(v \otimes w) && \text{for } a \in \mathbb{R}, \; v \in V, \; w \in W \\
&(v' + v'') \otimes w = v' \otimes w + v'' \otimes w && \text{for } v', v'' \in V, \; w \in W \\
&v \otimes (w' + w'') = v \otimes w' + v \otimes w'' && \text{for } v \in V, \; w', w'' \in W \\
&v \otimes 0 = 0 = 0 \otimes w && \text{for } v \in V, \; w \in W
\end{aligned}$$

3. Linearly independent vectors $\{v_i\}_i \subset V$ and $\{w_j\}_j \subset W$ lead to linearly independent vectors $\{v_i \otimes w_j\}_{i,j} \subset U$.

De facto, we use the same symbol '$\otimes$' for two different purposes here. First, in tensor space notation, the symbol connects vector spaces, and second, it combines vectors v and w of the respective vector spaces V and W into the quantity $v \otimes w$. This is analogous to the summation '+' of vectors and vector spaces from Definition 3.4.

In order to obtain a RKHS as a tensor product, there are two important points to note with respect to Definition 7.5. First, the tensor space $V \otimes_a W$ is not complete for infinite-dimensional vector spaces V and W. Second, the mapping $\otimes_a$ initially maps to the abstract elements of equivalence classes, which need to be converted into actual functions in the RKHS setting. Regarding completion, [Hac12] introduces the topological tensor product $\otimes_{\|\cdot\|}$, which satisfies the relation

$$V \otimes_{\|\cdot\|} W = \overline{V \otimes_a W}^{\|\cdot\|},$$

for Banach spaces V and W. The steps of completion and conversion were accomplished by J. Neveu, who identified the native space's structure of tensor product kernels in

[Nev71, Chapter VI]. Today, this structure is known as Hilbert tensor product.

In the following, we introduce the concept of Hilbert tensor products and the closely related Hilbert-Schmidt mapping, which we know maps to a (complete) Hilbert space. Thus, we can skip the step of completion and still obtain J. Neveu's result in Theorem 7.10. First, we fix definitions and notations as in [KR83, Chapter 2].

Definition 7.6. Let $\mathcal{H}_1, ..., \mathcal{H}_M, Z$ be real Hilbert spaces and

$$\varphi : \bigtimes_{\ell=1}^{M} \mathcal{H}_\ell \longrightarrow Z$$

be a function on the Cartesian product of $\mathcal{H}_1, ..., \mathcal{H}_M$.

(i) The function φ is called a *bounded multilinear mapping* if it is linear in each of its variables (while the other variables remain fixed), and there exists a constant $c \in \mathbb{R}_+$ such that

$$\|\varphi(x_1, ..., x_M)\|_Z \leqslant c \cdot \prod_{\ell=1}^{M} \|x_\ell\|_{\mathcal{H}_\ell}$$

holds for any $(x_1, ..., x_M) \in \bigtimes_{\ell=1}^{M} \mathcal{H}_\ell$.

(ii) We call φ a *weak Hilbert-Schmidt mapping* if it is a bounded multilinear mapping and there exists a constant $d \in \mathbb{R}_+$ such that the estimate

$$\sum_{b_1 \in B_1} \cdots \sum_{b_M \in B_M} |\langle \varphi(b_1, ..., b_M), z \rangle_Z|^2 \leqslant d^2 \|z\|_Z^2$$

holds for any orthonormal bases $B_1 \subset \mathcal{H}_1, ..., B_M \subset \mathcal{H}_M$ and $z \in Z$.

The following theorem is a slight modification of [KR83, Theorem 2.6.4], which details the main properties of Hilbert tensor products. In contrast to the standard tensor product, the Hilbert tensor product of Hilbert spaces is again a Hilbert space. This is a crucial detail, as kernel-based approximation theory mainly works with RKHSs.

Theorem 7.7. *Let $\mathcal{H}_1, ..., \mathcal{H}_M$ be real Hilbert spaces. Then there exists a tuple $(\mathcal{H}, \varphi)$ of a Hilbert space $\mathcal{H}$ and a multilinear mapping*

$$\varphi : \bigtimes_{\ell=1}^{M} \mathcal{H}_\ell \longrightarrow \mathcal{H}$$

satisfying

$$\langle \varphi(x_1, ..., x_M), \varphi(y_1, ..., y_M) \rangle_{\mathcal{H}} = \prod_{\ell=1}^{M} \langle x_\ell, y_\ell \rangle_{\mathcal{H}_\ell}$$

for any $(x_1, ..., x_M), (y_1, ..., y_M) \in \bigtimes_{\ell=1}^{M} \mathcal{H}_\ell$. Additionally,

$$\mathcal{H}_0 = \operatorname{span}\left\{\varphi(x_1, ..., x_M) \ : \ (x_1, ..., x_M) \in \mathcal{H}_1 \times ... \times \mathcal{H}_M\right\}$$

is a dense subset of $\mathcal{H}$.

Furthermore,

a) Let $B_1 \subset \mathcal{H}_1, ..., B_M \subset \mathcal{H}_M$ *be orthonormal bases. Then*

$$B = \left\{ \varphi(b_1, ..., b_M) \ : \ (b_1, ..., b_M) \in B_1 \times ... \times B_M \right\}$$

is a orthonormal basis of $\mathcal{H}$.

The following uniqueness statements hold:

b) The map φ *is a weak Hilbert-Schmidt mapping and satisfies the universal property: If* Z *is another Hilbert space and* $\psi : \times_{\ell=1}^{M} \mathcal{H}_\ell \longrightarrow Z$ *is a weak Hilbert-Schmidt mapping, there is a unique bounded linear map* $T : \mathcal{H} \longrightarrow Z$ *such that* $\psi = T \circ \varphi$.

c) Let $\tilde{\mathcal{H}}$ *be a Hilbert space and* $\tilde{\varphi} : \times_{\ell=1}^{M} \mathcal{H}_\ell \longrightarrow \tilde{\mathcal{H}}$ *be a weak Hilbert-Schmidt mapping that satisfy the universal property from part b). Then, there exists an isometric isomorphism* $U : \mathcal{H} \longrightarrow \tilde{\mathcal{H}}$ *with* $U \circ \varphi = \tilde{\varphi}$. *Hence,* $(\tilde{\mathcal{H}}, \tilde{\varphi})$ *satisfies the above properties of* $(\mathcal{H}, \varphi)$.

d) If $(\tilde{\mathcal{H}}, \tilde{\varphi})$ *satisfies the above properties of* $(\mathcal{H}, \varphi)$, *there exists an isometric isomorphism* $U : \mathcal{H} \longrightarrow \tilde{\mathcal{H}}$ *with* $U \circ \varphi = \tilde{\varphi}$.

Note that φ, being the identity, does not satisfy the required properties, as it is no multilinear mapping, but only a linear one. Consequently, the Cartesian product '$\times$' is not a tensor product.

Finally, we can provide a precise definition of the Hilbert tensor product.

Definition 7.8. Let $\mathcal{H}_1, ..., \mathcal{H}_M$ be Hilbert spaces and the tuple $(\mathcal{H}, \varphi)$ be given by Theorem 7.7. Then $(\mathcal{H}, \varphi)$ is called the *Hilbert tensor product* of $\mathcal{H}_1, ..., \mathcal{H}_M$, denoted by

$$\bigotimes_{\ell=1}^{M} \mathcal{H}_\ell := \overline{\operatorname{span}\left\{ \bigotimes_{\ell=1}^{M} x_\ell \ : \ x_\ell \in \mathcal{H}_\ell \text{ for } \ell = 1, \dots, M \right\}} = (\mathcal{H}, \varphi),$$

where

$$\bigotimes_{\ell=1}^{M} x_\ell = x_1 \otimes ... \otimes x_M := \varphi(x_1, ..., x_M) \qquad \text{for } (x_1, ..., x_M) \in \bigtimes_{\ell=1}^{M} \mathcal{H}_\ell$$

Shortly, we write $\mathcal{H} = \bigotimes_{\ell=1}^{M} \mathcal{H}_\ell$.

Note, that due to Theorem 7.7 c) and d) the Hilbert tensor product is, up to isometric isomorphy, uniquely characterized.

7.2.2 Native Space as Hilbert Tensor Product

We identify the native space $\mathcal{H}_{K,\Omega}$ of a tensor product kernel K with components K_ℓ as the Hilbert tensor product of its components' native spaces $\mathcal{H}_{K_\ell,\Omega_\ell}$ in the following. To do so, we define a mapping φ from the Cartesian product of the components' native spaces to the tensor product kernel's native space and show that the tuple $(\mathcal{H}_{K,\Omega}, \varphi)$ satisfies the properties of Theorem 7.7. This means that, unlike J. Neveu's approach, we directly assume that φ maps onto a set of functions, thereby allowing us to omit the conversion step. Note that the underlying algebraic structure of the kernel justifies its designation as 'tensor product kernel'.

Theorem 7.9. *Let K_ℓ be positive semi-definite kernels on Ω_ℓ for $\ell = 1, ..., M$, and K be the corresponding tensor product kernel on $\Omega = \times_{\ell=1}^M \Omega_\ell$. Then the mapping*

$$\begin{aligned} \varphi : \bigtimes_{\ell=1}^{M} \mathcal{H}_{K_\ell,\Omega_\ell} &\longrightarrow \mathcal{H}_{K,\Omega}, \\ (f_1, ..., f_M) &\longmapsto \prod_{\ell=1}^{M} f_\ell \circ p_\ell \end{aligned} \tag{7.2}$$

is well-defined, multilinear, and satisfies the equation

$$\big\langle \varphi(f_1, ..., f_M), \varphi(g_1, ..., g_M) \big\rangle_{K,\Omega} = \prod_{\ell=1}^{M} \langle f_\ell, g_\ell \rangle_{K_\ell,\Omega_\ell} \tag{7.3}$$

for all $(f_1, ..., f_M), (g_1, ..., g_M) \in \times_{\ell=1}^M \mathcal{H}_{K_\ell,\Omega_\ell}$. Moreover, it is

$$S_{K,\Omega} \subset \operatorname{span}\left\{\varphi(f_1, ..., f_M) \ : \ f_\ell \in \mathcal{H}_{K_\ell,\Omega_\ell} \text{ for all } \ell = 1, ..., M\right\}. \tag{7.4}$$

Proof. The proof is carried out via induction on M.
$\underline{M = 2}$: We divide the initial step into three parts. In (i), we show (7.2) and (7.3) for the pre Hilbert spaces $S_{K,\Omega}$. In the second part (ii), we extend the results to $\mathcal{H}_{K,\Omega}$, and in (iii), we conclude the initial step by by proving (7.4).

(i) Let $f_1 \in S_{K_1,\Omega_1}$ and $f_2 \in S_{K_2,\Omega_2}$ be given by

$$f_1 = \sum_{i=1}^{N_1} a_i K_1(\cdot, y_i), \quad f_2 = \sum_{j=1}^{N_2} b_j K_2(\cdot, z_j)$$

with $\{y_1, ..., y_{N_1}\} \subset \Omega_1$, $\{z_1, ..., z_{N_2}\} \subset \Omega_2$. Applying φ on (f_1, f_2) leads to

$$\begin{aligned} \varphi(f_1, f_2) &= (f_1 \circ p_1) \cdot (f_2 \circ p_2) \\ &= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} a_i b_j \, K_1\left(p_1(\cdot), y_i\right) \, K_2\left(p_2(\cdot), z_j\right) \\ &= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} a_i b_j \, K_1\left(p_1(\cdot), p_1(x_{i,j})\right) \, K_2\left(p_2(\cdot), p_2(x_{i,j})\right) \\ &= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} a_i b_j \, K(\cdot, x_{i,j}) \qquad \in S_{K,\Omega} \subseteq \mathcal{H}_{K,\Omega}, \end{aligned}$$

where K denotes the tensor product kernel of K_1 and K_2 and $x_{i,j} = (y_i, z_j)$ is an element of $\Omega_1 \times \Omega_2 = \Omega$. As a consequence, φ is well-defined on the dense subsets. Clearly, the mapping is multilinear as well. Given the additional elements

$$g_1 = \sum_{k=1}^{M_1} c_k K_1(\cdot, \tilde{y}_k) \in S_{K_1,\Omega_1}, \quad g_2 = \sum_{\ell=1}^{M_2} d_\ell K_2(\cdot, \tilde{z}_\ell) \in S_{K_2,\Omega_2}$$

with $\{\tilde{y}_1, ..., \tilde{y}_{M_1}\} \subset \Omega_1, \{\tilde{z}_1, ..., \tilde{z}_{M_2}\} \subset \Omega_2$, the computation

$$\begin{aligned}
\big\langle \varphi(f_1, f_2), \varphi(g_1, g_2) \big\rangle_{K,\Omega} &= \Bigg\langle \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} a_i b_j \; K_1\big(p_1(\cdot), y_i\big) \; K_2\big(p_2(\cdot), z_j\big), \\
&\qquad \sum_{k=1}^{M_1} \sum_{\ell=1}^{M_2} c_k d_\ell \; K_1\big(p_1(\cdot), \tilde{y}_k\big) \; K_2\big(p_2(\cdot), \tilde{z}_\ell\big) \Bigg\rangle_{K,\Omega} \\
&= \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \; \sum_{k=1}^{M_1} \sum_{\ell=1}^{M_2} a_i b_j \; c_k d_\ell \; K_1(y_i, \tilde{y}_k) \, K_2(z_j, \tilde{z}_\ell) \\
&= \sum_{j=1}^{N_1} \sum_{k=1}^{M_1} a_i c_k \; K_1(y_i, \tilde{y}_k) \; \cdot \; \sum_{j=1}^{N_2} \sum_{\ell=1}^{M_2} b_j d_\ell \; K_2(z_j, \tilde{z}_\ell) \\
&= \langle f_1, \; g_1 \rangle_{K_1,\Omega_1} \cdot \langle f_2, \; g_2 \rangle_{K_2,\Omega_2}
\end{aligned}$$

yields (7.3) using Theorem 5.6 for the last equation.

(ii) For all $f_1 \in \mathcal{H}_{K_1,\Omega_1}$, $f_2 \in \mathcal{H}_{K_2,\Omega_2}$, there exist a normwise convergent sequences $\big(s_n^1\big)_{n\in\mathbb{N}} \subset S_{K_1,\Omega_1}$ and $\big(s_n^2\big)_{n\in\mathbb{N}} \subset S_{K_2,\Omega_2}$ with

$$\lim_{n\to\infty} s_n^1 = f_1 \qquad \lim_{n\to\infty} s_n^2 = f_2.$$

Due to part (i), we can state the estimate

$$\begin{aligned}
&\big\|\varphi(s_m^1, s_m^2) - \varphi(s_n^1, s_n^2)\big\|_{K,\Omega} \\
&\quad = \big\|(s_m^1 \circ p_1) \cdot (s_m^2 \circ p_2) - (s_n^1 \circ p_1) \cdot (s_n^2 \circ p_2)\big\|_{K,\Omega} \\
&\quad \leqslant \big\|(s_m^1 - s_n^1) \circ p_1\big\|_{K_{1,p_1},\Omega} \cdot \big\|s_m^2 \circ p_2\big\|_{K_{2,p_2},\Omega} \\
&\qquad\qquad + \big\|s_n^1 \circ p_1\big\|_{K_{1,p_1},\Omega} \cdot \big\|(s_m^2 - s_n^2) \circ p_2\big\|_{K_{2,p_2},\Omega} \\
&\quad = \big\|s_m^1 - s_n^1\big\|_{K_1,\Omega_1} \cdot \big\|s_m^2\big\|_{K_2,\Omega_2} + \big\|s_n^1\big\|_{K_1,\Omega_1} \cdot \big\|s_m^2 - s_n^2\big\|_{K_2,\Omega_2}
\end{aligned}$$

for all $n, m \in \mathbb{N}$. Hence, the sequence $\big(\varphi(s_n^1, s_n^2)\big)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}_{K,\Omega}$ and therefore approaches a normwise limit in $\mathcal{H}_{K,\Omega}$, here denoted with g, i.e.,

$$g = \lim_{n\to\infty} \varphi(s_n^1, s_n^2).$$

Since norm convergence implies pointwise convergence in RKHSs, see Theorem 2.21, we obtain

$$g(x) = \lim_{n\to\infty} s_n^1(y) \cdot s_n^2(z) \qquad \text{for all } x = (y, z) \in \Omega,$$

where the right-hand side can be written as

$$\lim_{n\to\infty} s_n^1(y)\cdot s_n^2(z) = \lim_{n\to\infty} s_n^1(y)\cdot \lim_{n\to\infty} s_n^2(z) = f_1(y)\cdot f_2(z) \qquad \text{for all } x=(y,z)\in\Omega,$$

by applying the same argument with respect to the norm convergence in the spaces $\mathcal{H}_{K_1,\Omega_1}$ and $\mathcal{H}_{K_2,\Omega_2}$. Consequently,

$$\varphi(f_1,f_2)\in\mathcal{H}_{K,\Omega} \qquad \text{for all } f_1\in\mathcal{H}_{K_1,\Omega_1}, f_2\in\mathcal{H}_{K_2,\Omega_2},$$

showing that φ is a well-defined function. Since taking the limit is a linear operator, φ maintains multilinear, when extending its domain from $S_{K_1,\Omega_1}\times S_{K_2,\Omega_2}$ to $\mathcal{H}_{K_1,\Omega_1}\times\mathcal{H}_{K_2,\Omega_2}$.

Given additional elements $g_1\in\mathcal{H}_{K_1,\Omega_1}$ and $g_2\in\mathcal{H}_{K_2,\Omega_1}$, we can approximate them with convergent sequences $\left(\tilde{s}_n^1\right)_{n\in\mathbb{N}}\subset S_{K_1,\Omega_1}$ and $\left(\tilde{s}_n^2\right)_{n\in\mathbb{N}}\subset S_{K_2,\Omega_2}$ as well. The continuity of inner products and part (i) results in (7.3), as

$$\begin{aligned}\left\langle\varphi(f_1,f_2),\ \varphi(g_1,g_2)\right\rangle_{K,\Omega} &= \lim_{n\to\infty}\left\langle\varphi(s_n^1,s_n^2),\ \varphi(\tilde{s}_n^1,\tilde{s}_n^2)\right\rangle_{K,\Omega}\\ &= \lim_{n\to\infty}\langle s_n^1,\ \tilde{s}_n^1\rangle_{K_1,\Omega_1}\cdot\langle s_n^2,\ \tilde{s}_n^2\rangle_{K_2,\Omega_2}\\ &= \langle f_1,\ g_1\rangle_{K_1}\cdot\langle f_2,\ g_2\rangle_{K_2}.\end{aligned}$$

(iii) Let $x=(y,z)\in\Omega_1\times\Omega_2=\Omega$, then

$$K(\cdot,x)=K_1\left(p_1(\cdot),y\right)\ K_2\left(p_2(\cdot),z\right)=\varphi\left(K_1(\cdot,y),K_2(\cdot,z)\right).$$

With this we obtain the relation (7.4), since

$$\begin{aligned}S_{K,\Omega} &= \operatorname{span}\left\{K(\cdot,x)\ :\ x\in\Omega\right\}\\ &= \operatorname{span}\left\{\varphi\left(K_1(\cdot,y),K_2(\cdot,z)\right)\ :\ y\in\Omega_1, z\in\Omega_2\right\}\\ &\subset \operatorname{span}\left\{\varphi(f_1,...,f_M)\ :\ f_\ell\in\mathcal{H}_{K_\ell,\Omega_\ell} \text{ for all } \ell=1,...,M\right\}.\end{aligned}$$

$\underline{M\to M+1}$: Let $\tilde{\Omega}=\bigtimes_{\ell=1}^M\Omega_\ell$ and $\tilde{K}$ the tensor product kernel of K_ℓ for $\ell=1,\dots,M$. Due to the induction basis and hypothesis, the mappings

$$\tilde{\varphi}:\bigtimes_{\ell=1}^{M}\mathcal{H}_{K_\ell,\Omega_\ell}\longrightarrow\mathcal{H}_{\tilde{K},\tilde{\Omega}},\quad (f_1,...,f_M)\longmapsto\prod_{\ell=1}^{M}f_\ell\circ p_\ell$$

and

$$\bar{\varphi}:\mathcal{H}_{\tilde{K},\tilde{\Omega}}\times\mathcal{H}_{K_{M+1},\Omega_{M+1}}\longrightarrow\mathcal{H}_{K,\Omega},\quad \left(\tilde{f},f_{M+1}\right)\longmapsto(\tilde{f}\circ\tilde{p})\cdot(f_{M+1}\circ p_{M+1}),$$

where $\tilde{p}$ defines the projection from $\tilde{\Omega}\times\Omega_{M+1}$ to $\tilde{\Omega}$, are well-defined and satisfy (7.3) and (7.4) on their respective domains. Since

$$\prod_{\ell=1}^{M+1}f_\ell\circ p_\ell=\bar{\varphi}\left(\tilde{\varphi}(f_1,...,f_M),\ f_{M+1}\right)$$

holds for all $(f_1, ..., f_{M+1}) \in \bigtimes_{\ell=1}^{M+1} \mathcal{H}_{K_\ell,\Omega_\ell}$, the mapping

$$\varphi : \bigtimes_{\ell=1}^{M+1} \mathcal{H}_{K_\ell,\Omega_\ell} \longrightarrow \mathcal{H}_{K,\Omega}, \quad (f_1, ..., f_{M+1}) \longmapsto \prod_{\ell=1}^{M+1} f_\ell \circ p_\ell$$

is well-defined and multilinear. In order to show (7.3), let

$$f = (f_1, ..., f_{M+1}),\ g = (g_1, ..., g_{M+1}) \in \bigtimes_{\ell=1}^{M+1} \mathcal{H}_{K_\ell,\Omega_\ell}.$$

Then we obtain

$$\begin{aligned}
\big\langle \varphi(f), \varphi(g) \big\rangle_{K,\Omega} &= \Big\langle \bar{\varphi}\left(\tilde{\varphi}(f_1, ..., f_M), f_{M+1}\right), \bar{\varphi}\left(\tilde{\varphi}(g_1, ..., g_M), g_{M+1}\right) \Big\rangle_{K,\Omega} \\
&= \big\langle \tilde{\varphi}(f_1, ..., f_M), \tilde{\varphi}(g_1, ..., g_M) \big\rangle_{\tilde{K},\tilde{\Omega}} \cdot \langle f_{M+1}, g_{M+1} \rangle_{K_{M+1},\Omega_{M+1}} \\
&= \prod_{\ell=1}^{M} \langle f_\ell, g_\ell \rangle_{K_\ell,\Omega_\ell} \cdot \langle f_{M+1}, g_{M+1} \rangle_{K_{M+1},\Omega_{M+1}} \\
&= \prod_{\ell=1}^{M+1} \langle f_\ell, g_\ell \rangle_{K_\ell,\Omega_\ell}.
\end{aligned}$$

A similar argument as in the case $M = 2$ (iii) validates (7.4). ■

The next theorem concludes the above analysis by identifying the tensor product kernel's native space $\mathcal{H}_{K,\Omega}$ as the Hilbert tensor product of its components' native spaces $\mathcal{H}_{K_\ell,\Omega_\ell}$.

Theorem 7.10. *Let K_ℓ be positive semi-definite kernels on Ω_ℓ for $\ell = 1, ..., M$ and K their corresponding tensor product kernel on $\Omega = \bigtimes_{\ell=1}^{M} \Omega_\ell$. Then*

$$\mathcal{H}_{K,\Omega} = \bigotimes_{\ell=1}^{M} \mathcal{H}_{K_\ell,\Omega_\ell},$$

where

$$f_1 \otimes \cdots \otimes f_M = \prod_{\ell=1}^{M} f_\ell \circ p_\ell \qquad \text{for all } f_\ell \in \mathcal{H}_{K_\ell,\Omega_\ell},\ \ell = 1, \ldots, M.$$

Proof. By Theorem 7.2 and Section 2.3.2 there exist the native space $\mathcal{H}_{K,\Omega}$ of the kernel K. Consider the mapping φ of (7.2) in Theorem 7.9. Since φ satisfies (7.3) and (7.4) and $S_{K,\Omega}$ is a dense subset of $\mathcal{H}_{K,\Omega}$, the tuple $(\mathcal{H}_{K,\Omega}, \varphi)$ satisfies all properties from Theorem 7.7. Definition 7.8 finishes the proof. ■

Let us approach the matter from a different perspective, not starting from the tensor product kernel, but rather from the tensor product of the native spaces. To do so, let $(\mathcal{H}, \otimes)$ be a Hilbert tensor product of the native spaces $\mathcal{H}_{K_\ell,\Omega_\ell}$, defined as in Theorem 7.7. Then the tensor of the reproducing kernels K_ℓ lies in $\mathcal{H}$, i.e.,

$$\bigotimes_{\ell=1}^{M} K_\ell(\cdot, x_\ell) \in \mathcal{H} \qquad \text{for all } x_\ell \in \Omega_\ell,\ \ell = 1, \ldots, M.$$

Furthermore, for an element $f \in \mathcal{H}_0$, lying in the dense subset $\mathcal{H}_0 \subset \mathcal{H}$, of the form

$$f = \sum_{i=1}^{N} \alpha_i \bigotimes_{\ell=1}^{M} f_\ell^i, \qquad \text{where } f_\ell^i \in \mathcal{H}_{K_\ell,\Omega_\ell} \text{ for } i = 1,\dots,N \text{ and } \ell = 1,\dots,M,$$

it holds that

$$\begin{aligned}\left\langle f,\ \bigotimes_{\ell=1}^{M} K_\ell(\cdot,x_\ell)\right\rangle_{\mathcal{H}} &= \sum_{i=1}^{N} \alpha_i \left\langle \bigotimes_{\ell=1}^{M} f_\ell^i,\ \bigotimes_{\ell=1}^{M} K_\ell(\cdot,x_\ell)\right\rangle_{\mathcal{H}} \\ &= \sum_{i=1}^{N} \alpha_i \prod_{\ell=1}^{M} \left\langle f_\ell^i,\ K_\ell(\cdot,x_\ell)\right\rangle_{\mathcal{H}_{K_\ell,\Omega_\ell}} \\ &= \sum_{i=1}^{N} \alpha_i \prod_{\ell=1}^{M} f_\ell^i(x_\ell),\end{aligned}$$

where the right hand side is a function on $\times_{\ell=1}^{M} \Omega_\ell$. Consequently, if

$$\bigotimes_{\ell=1}^{M} f_\ell = \prod_{\ell=1}^{M} f_\ell \circ p_\ell,$$

the space $\mathcal{H}$ is a RKHS with the tensor product kernel as its reproducing kernel. Thanks to the uniqueness of reproducing kernels of Theorem 2.22, the tensor product kernel is the unique reproducing kernel corresponding to this mapping. We summarize the above analysis in the following lemma.

Lemma 7.11. *Let $(\mathcal{H},\otimes)$ be a Hilbert tensor product of the native spaces $\mathcal{H}_{K_\ell,\Omega_\ell}$, i.e.,*

$$\mathcal{H} = \bigotimes_{\ell=1}^{M} \mathcal{H}_{K_\ell,\Omega_\ell}.$$

Then $(\mathcal{H},\otimes) = (\mathcal{H}_{K,\Omega}, \|\cdot\|_K)$ if and only if

$$\bigotimes_{\ell=1}^{M} f_\ell = \prod_{\ell=1}^{M} f_\ell \circ p_\ell, \qquad \text{where } f_\ell \in \mathcal{H}_{K_\ell,\Omega_\ell} \text{ for } \ell = 1,\dots,M.$$

7.3 Interpolation

For interpolation with tensor product kernels, we first examine two important concepts: point sets of grid-like structure and the Kronecker product. In Section 7.3.1, we combine these two concepts to demonstrate that unique interpolation with the tensor product kernel on grid-like structured data sets is feasible (Corollary 7.16). Furthermore, we delve into the convergence of the reconstruction method resulting in Theorem 7.23 as well as the construction of orthonormal Newton basis given by (7.6).

Despite the practical advantages of interpolation on grid-like point sets, it remains to be determined whether a tensor product kernel is positive definite, that is, whether the interpolation matrix is positive definite for all pairwise distinct data sets (not only of grid-like structure). We affirm this in Section 7.3.2, Theorem 7.25, and subsequently address the numerical stability of the interpolation method with tensor product kernel for both scattered and grid-like point sets in Section 7.3.3.

Definition 7.12. If a point set $X \subset \mathbb{R}^d$ can be written as a Cartesian product

$$X = X^1 \times \ldots \times X^M = \bigtimes_{\ell=1}^{M} X^\ell$$

of finite pairwise distinct sets $X^\ell \subset \mathbb{R}^{d_\ell}$, where $\ell = 1, \ldots, M$ and $d = \sum_{\ell=1}^M d_\ell$, we say that it has *grid-like* structure.

Definition 7.13. The *Kronecker product*

$$A \otimes B$$

of two matrices $A = (a_{i,j})_{1 \leqslant i \leqslant m,\, 1 \leqslant j \leqslant n} \in \mathbb{R}^{m \times n}$ and $B = (a_{i,j})_{1 \leqslant i \leqslant p,\, 1 \leqslant j \leqslant q} \in \mathbb{R}^{p \times q}$ is defined as the block matrix

$$A \otimes B = \begin{pmatrix} a_{1,1}B & \cdots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{m,1}B & \cdots & a_{m,n}B \end{pmatrix} \in \mathbb{R}^{mp \times nq}.$$

The entries of $A \otimes B$ are given by

$$(A \otimes B)_{j,k} = a_{\lceil j/p \rceil, \lceil k/q \rceil} \cdot b_{(j-1) \bmod p+1,\, (k-1) \bmod q+1}$$

for $1 \leqslant j \leqslant mp$ and $1 \leqslant k \leqslant nq$, where $\lceil \cdot \rceil$ denotes the ceiling function and 'mod' is the remainder after division.

The Kronecker product is an algebraic tensor product of matrices (cf. [Hac12, Chapter 1.1.2]). Thus, not surprisingly, this product aligns closely with the tensor product kernel. We state some basic properties of the Kronecker product, which will be needed throughout this section. For proofs, we refer to [HJ91, Chapter 4.2].

Lemma 7.14. *The following statements hold:*

(i) *Let* $\alpha \in \mathbb{R}, A \in \mathbb{R}^{m \times n},\ B \in \mathbb{R}^{p \times q}, C \in \mathbb{R}^{n \times k}$ *and* $D \in \mathbb{R}^{q \times r}$*, then*

$$\begin{aligned} (\alpha A) \otimes B &= A \otimes (\alpha B) \\ (A \otimes B)^T &= A^T \otimes B^T \\ (A\,B) \otimes (C\,D) &= (A \otimes C)\,(B \otimes D). \end{aligned}$$

(ii) *Let* $A \in \mathbb{R}^{n \times n}$ *and* $B \in \mathbb{R}^{m \times m}$ *be regular matrices, then* $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

(iii) *For positive (semi-)definite square matrices* $A \in \mathbb{R}^{n \times n}$ *and* $B \in \mathbb{R}^{m \times m}$*, the corresponding Kronecker product* $A \otimes B$ *is positive (semi-)definite.*

(iv) *For square matrices* $A \in \mathbb{R}^{n \times n}$ *and* $B \in \mathbb{R}^{m \times m}$*, every eigenvalue* $\lambda(A \otimes B)$ *of* $A \otimes B$ *arises as a product of eigenvalues* $\lambda(A), \lambda(B)$ *of* A, B *respectively, i.e.,*

$$\lambda(A \otimes B) = \lambda(A)\lambda(B).$$

This directly implies

$$\lambda_{\max}(A \otimes B) = \lambda_{\max}(A)\lambda_{\max}(B) \qquad \text{and} \qquad \lambda_{\min}(A \otimes B) = \lambda_{\min}(A)\lambda_{\min}(B).$$

Subsequently, we apply these properties to a product of arbitrary but finitely many matrices. This is valid, as the aforementioned statements can be extended by induction.

7.3.1 Interpolation on Grid-Like Point Sets

The following combines the concept of grid-like structured point sets with that of the Kronecker product. Specifically, the interpolation matrix $\mathbf{A}_{K,X}$ corresponding to the tensor product kernel K and grid-like data X can be expressed as a Kronecker product. While [Saa11] identified this relationship, they did not provide a rigorous proof. Here, we present a formal one.

Theorem 7.15. *Let K_ℓ be positive semi-definite kernels on Ω_ℓ and $X^\ell \subset \Omega_\ell$ be finite and pairwise distinct point sets for $\ell = 1, ..., M$. Moreover, let K be the corresponding tensor product kernel and $X = \times_{\ell=1}^{M} X^\ell$. Then there exists an ordering of X, such that the corresponding interpolation matrix $\mathbf{A}_{K,X}$ can be written as*

$$\mathbf{A}_{K,X} = \bigotimes_{\ell=1}^{M} \mathbf{A}_{K_\ell, X^\ell}.$$

Proof. The proof is carried out via induction on M.
$\underline{M = 2}$: Let $X^1 = \{x_1^1, ..., x_{N_1}^1\} \subset \Omega_1$ and $X^2 = \{x_1^2, ..., x_{N_2}^2\} \subset \Omega_2$ be pairwise distinct point sets, and $N = N_1 N_2$. For $k \in \{1, ..., N\}$, we set

$$x_k = (x^1_{\lceil k/N_2 \rceil},\ x^2_{(k-1) \bmod N_2+1}).$$

This leads to an ordering $X = X^1 \times X^2 = \{x_k \ : \ k = 1, ..., N\}$, which results in

$$\begin{aligned}
\left(\mathbf{A}_{K,X}\right)_{j,k} &= K\left(\left(x^1_{\lceil j/N_2 \rceil},\ x^2_{(j-1) \bmod N_2+1}\right),\ \left(x^1_{\lceil k/N_2 \rceil},\ x^2_{(k-1) \bmod N_2+1}\right)\right) \\
&= K^1\left(x^1_{\lceil j/N_2 \rceil},\ x^1_{\lceil k/N_2 \rceil}\right) \cdot K^2\left(x^2_{(j-1) \bmod N_2+1},\ x^2_{(k-1) \bmod N_2+1}\right) \\
&= \left(\mathbf{A}_{K_1,X^1}\right)_{\lceil j/N_2 \rceil,\, \lceil k/N_2 \rceil} \cdot \left(\mathbf{A}_{K_2,X^2}\right)_{(j-1) \bmod N_2+1,\, (k-1) \bmod N_2+1}
\end{aligned}$$

for any $j, k \in \{1, \dots, N\}$. Hence, the equation $\mathbf{A}_{K,X} = \mathbf{A}_{K_1,X^1} \otimes \mathbf{A}_{K_2,X^2}$ holds.

$\underline{M \to M+1}$: Let K_ℓ be positive semi-definite kernels on Ω_ℓ and $X^\ell \subset \Omega_\ell$ be finite and pairwise distinct point sets for $\ell = 1, ..., M+1$. Furthermore, let $\tilde{X} = \times_{\ell=1}^{M} X^\ell$, $N_\ell = |X^\ell|$ for $\ell = 1, ..., M$, and $\tilde{N} = \prod_{\ell=1}^{M} N_\ell$. We denote by $\tilde{K}$ the tensor product kernel corresponding to K_ℓ for $\ell = 1, \dots, M$. Due to the induction hypothesis, there is an ordering $\tilde{X} = \{\tilde{x}_\ell \ : \ \ell = 1, ..., \tilde{N}\}$ so that the equation

$$\mathbf{A}_{\tilde{K},\tilde{X}} = \bigotimes_{\ell=1}^{M} \mathbf{A}_{K_\ell, X^\ell}$$

holds. As in the initial case, we can order $X = \tilde{X} \times X^{M+1}$ to get

$$\mathbf{A}_{K,X} = \mathbf{A}_{\tilde{K},\tilde{X}} \otimes \mathbf{A}_{K_{M+1}, X^{M+1}},$$

where K is the tensor product kernel corresponding to K_ℓ for $\ell = 1, \dots, M+1$. In total, we conclude

$$\mathbf{A}_{K,X} = \bigotimes_{\ell=1}^{M+1} \mathbf{A}_{K_\ell, X^\ell}.$$

■

Combining the Theorem 7.15 with Lemma 7.14 (iii), yields positive definiteness of the interpolation matrix $\mathbf{A}_{K,X}$ in the setting of grid-like structured point sets X.

Corollary 7.16. *Let K_ℓ be positive definite kernels on Ω_ℓ for $\ell = 1, \dots, M$, and K their corresponding tensor product kernel on $\Omega = \times_{\ell=1}^M \Omega_\ell$. Furthermore, let $X \subset \Omega$ be grid-like structured. Then the interpolation matrix $\mathbf{A}_{K,X}$ of the tensor product kernel K is positive definite.*

Remark 7.17. Despite the considerable benefits due to interpolation on grid-like structured point sets, this assumption simultaneously imposes constraints. Naturally, not all applications feature interpolation point sets of grid-like structure, and even when they do, data may frequently be missing due to water, governmental boundaries, missing pixels, and the like. Particularly in the latter scenario, it becomes imperative to preserve the structural advantages while relaxing the grid assumption. For more information and a detailed process to do so, we refer to [WGNC14].

Now that we provided unique interpolation on grid-like point sets, we turn our attention to interpolation with tensor product kernels in this setting. The upcoming theorem demonstrates that the tensor of the individual component interpolants coincides with the interpolant of the tensor product kernel. Consequently, there are two methods of achieving the same interpolant in a grid-like setting. However, if we decide upon computing the component interpolants s_{f_ℓ,K_ℓ,X^ℓ} initially and subsequently multiply them, we require knowledge of the functions f_ℓ on X^ℓ. Such detailed information about the problem is unlikely to be available. In contrast, the tensor product kernel does not need this level of detail, but only requires the function values of the target function f_X, thereby yielding the same interpolant.

Theorem 7.18. *Let K_ℓ be positive definite kernels on Ω_ℓ for $\ell = 1, \dots, M$, and K their tensor product kernel acting on $\Omega = \times_{\ell=1}^M \Omega_\ell$. Furthermore, let the target function be of the tensor form $f = \prod_{\ell=1}^M f_\ell \circ p_\ell$ with f_ℓ acting on Ω_ℓ and p_ℓ being the projection from Ω to Ω_ℓ for $\ell = 1, \dots, M$, and $X = \times_{\ell=1}^M X^\ell$ be of grid-like structure. Then,*

$$s_{f,K,X} = \prod_{\ell=1}^M s_{f_\ell,K_\ell,X^\ell} \circ p_\ell.$$

Proof. The function $\prod\limits_{\ell=1}^M s_{f_\ell,K_\ell,X^\ell} \circ p_\ell$ lies in $S_{K,X}$ by the proof of Theorem 7.9. Moreover, it satisfies the interpolation condition, as

$$f(x_n) = \prod_{\ell=1}^M f_\ell \circ p_\ell(x_n) = \prod_{\ell=1}^M f_\ell(x_n^\ell) = \prod_{\ell=1}^M s_{f_\ell,K_\ell,X^\ell}(x_n^\ell) = \prod_{\ell=1}^M s_{f_\ell,K_\ell,X^\ell} \circ p_\ell(x_n),$$

for all $x_n = (x_n^1, \dots, x_n^M) \in \times_{\ell=1}^M X^\ell = X$. Section 2.1 combined with Corollary 7.16 provides the uniqueness of an interpolant in $S_{K,X}$. This finishes the proof. ∎

Newton Basis In [MS09], the orthonormal Newton basis of $S_{K,X}$ was introduced to improve stability and efficiency of kernel-based reconstruction methods. The construction of this orthonormal basis relies on the Cholesky decomposition of the interpolation matrix $\mathbf{A}_{K,X}$ and is rather costly compared to the standard basis $K(\cdot, x_i)$. But, the property of a tensor product kernel to write its interpolation matrix as a Kronecker product, does notably decrease the computational expenses associated with computing the Newton basis, as shown in the numerical tests of Section 7.4.2.

We start by introducing and constructing the Newton basis. For details and proofs we refer to [MS09].

Definition 7.19. The set of functions $\mathcal{N} = \{\mathfrak{n}_1, \ldots, \mathfrak{n}_N\}$ acting on Ω regarding the point set $\{x_1, \ldots, x_N\} = X \subset \Omega$ and the kernel K, defined by

$$\begin{aligned} \mathfrak{n}_i(x_j) &= 0 \qquad \text{for } 0 \leqslant j < i \leqslant N \\ \mathfrak{n}_i(x_i) &= 1 \qquad \text{for } 0 \leqslant i \leqslant N \end{aligned}$$

and the requirement

$$\mathfrak{n}_i \in S_{K,X,i} := \operatorname{span}\left\{K(\cdot, x_1), \ldots, K(\cdot, x_i)\right\} \qquad \text{for } 0 \leqslant i \leqslant N,$$

is called the *Newton basis* of $S_{K,X}$.

Theorem 7.20. *Let K be a positive definite kernel on Ω and $\{x_1, \ldots, x_N\} = X \subset \Omega$ be a pairwise distinct point set. Furthermore, let $\mathbf{A}_{K,X} = LL^T$ be the Cholesky decomposition of the interpolation matrix $\mathbf{A}_{K,X}$, where L is a lower triangular matrix, and let $(L^T)^{-1} = \left(u_{ij}\right)_{1 \leqslant i,j \leqslant N}$ denote the upper triangular inverse of L^T. Then the functions*

$$\mathfrak{n}_i = \sum_{j=1}^{i} u_{ij}\, K(\cdot, x_j) \qquad i = 1, \ldots, N$$

form an orthonormal Newton basis of $S_{K,X}$.

In matrix notation, Theorem 7.20 provides

$$V_{\mathcal{N},X} = \mathbf{A}_{K,X} \left(L^T\right)^{-1} = L, \tag{7.5}$$

where $V_{\mathcal{N},X}$ denotes the Vandermonde matrix (Definition 2.1) of the Newton basis $\mathcal{N}$ evaluated at the point set X. We observe that the Vandermonde matrix of a Newton basis must be a lower triangular matrix with diagonal entries equal to one.

Below, we outline how we can leverage a tensor structure for computing the Newton basis. Let K_ℓ be positive definite kernels on Ω_ℓ for $\ell = 1, \ldots, M$ and K their tensor product kernel on $\Omega = \times_{\ell=1}^{M} \Omega_\ell$. Furthermore, let $X = \times_{\ell=1}^{M} X^\ell$ have grid-like structure, with $X^\ell \subset \Omega_\ell$. We combine the individual Newton bases of the component spaces

S_{K_ℓ,X^ℓ} to compute the Newton basis of $S_{K,X}$. Due to the grid-like structure of the point set X, we can write the approximation space $S_{K,X}$ as the Hilbert tensor product

$$S_{K,X} = \bigotimes_{\ell=1}^{M} S_{K_\ell,X^\ell},$$

where

$$\bigotimes_{\ell=1}^{M} f_\ell = \prod_{\ell=1}^{M} f_\ell \circ p_\ell \qquad \text{for all } f_\ell \in S_{K_\ell,X^\ell},\ \ell = 1, \dots, M.$$

Let $\mathcal{N}_\ell$ denote the Newton basis of S_{K_ℓ,X^ℓ}, then Theorem 7.7 a) in combination with Theorem 7.10 provides that

$$\mathcal{N} := \left\{ \bigotimes_{\ell=1}^{M} \mathfrak{n}_\ell \ :\ (\mathfrak{n}_1, ..., \mathfrak{n}_M) \in \mathcal{N}_1 \times ... \times \mathcal{N}_M \right\} \tag{7.6}$$

is an orthonormal basis of $S_{K,X}$. The Vandermonde matrix $V_{\mathcal{N},X}$ of $\mathcal{N}$ on X can be computed by the Kronecker product of the component Vandermonde matrices $V_{\mathcal{N}_\ell,X^\ell}$, resulting in

$$V_{\mathcal{N},X} = \bigotimes_{\ell=1}^{M} V_{\mathcal{N}_\ell,X^\ell} = \bigotimes_{\ell=1}^{M} L_\ell = \bigotimes_{\ell=1}^{M} \left(\mathbf{A}_{K_\ell,X^\ell} (L_\ell^T)^{-1} \right) \tag{7.7}$$

by (7.5). To show that the above representation of $V_{\mathcal{N},X}$ is the Choleski factor L of $\mathbf{A}_{K,X}$, we state the subsequent lemma.

Lemma 7.21. *Let K be a tensor product kernel of positive definite components K_ℓ acting on Ω_ℓ and $X = \times_{\ell=1}^{M} X^\ell$ have grid-like structure, with $X^\ell \subset \Omega_\ell$ and $\Omega = \times_{\ell=1}^{M} \Omega_\ell$. Additionally, let*

$$\mathbf{A}_{K_\ell,X^\ell} = L_\ell L_\ell^T \qquad \text{for } \ell = 1, ..., M$$

be the Cholesky factorizations of the component interpolation matrices $\mathbf{A}_{K_\ell,X^\ell}$ and X be ordered such that

$$\mathbf{A}_{K,X} = \bigotimes_{\ell=1}^{M} \mathbf{A}_{K_\ell,X^\ell}.$$

The Cholesky factor L of $\mathbf{A}_{K,X}$ is then given by the Kronecker product

$$L = \bigotimes_{\ell=1}^{M} L_\ell \qquad \text{and} \qquad \left(L^T\right)^{-1} = \bigotimes_{\ell=1}^{M} \left(L_\ell^T\right)^{-1}.$$

Proof. The statements directly follow from the properties of the Kronecker product provided in Lemma 7.14 (*i*) and the fact that the Kronecker product of lower triangular matrices is also a lower triangular matrix. ■

We apply Lemma 7.14 *(i)* and Lemma 7.21 on (7.7) to obtain

$$V_{\mathcal{N},X} = \bigotimes_{\ell=1}^{M} \mathbf{A}_{K_\ell,X^\ell} \bigotimes_{\ell=1}^{M} (L_\ell^T)^{-1} = \mathbf{A}_{K,X} \left(L^T\right)^{-1} = L,$$

showing that $\mathcal{N}$ is the Newton basis of $S_{K,X}$. This provides us with two distinct approaches for computing the Newton basis $\mathcal{N}$ of $S_{K,X}$. Fist, the common way proposed in Theorem 7.20 and second the tensor structure exploiting way of combining the component Newton basis $\mathcal{N}_\ell$ to $\mathcal{N}$ by (7.6).

Convergence We focus on the power function introduced in Definition 2.27, i.e.,

$$P_{K,X}(x) = \|K(\cdot,x) - s_{K(\cdot,x),K,X}\|_K \qquad \text{for all } x \in \Omega.$$

Hence, the power functions of the component kernels can be written as

$$P_{K_\ell,X^\ell}(x) = \|K_\ell(\cdot,x) - s_{K_\ell(\cdot,x),K_\ell,X^\ell}\|_{K_\ell} \qquad \text{for } x \in \Omega_\ell,\ \ell = 1, ..., M.$$

We note that $\mathfrak{n}_\ell(x) = P_{K_\ell,X^\ell}(x)$ for $x \in \Omega_\ell$ holds by [PS11]. In order to prove convergence, we state a representation of the tensor product kernel's power function.

Lemma 7.22. *Let K_ℓ be positive definite kernels on Ω_ℓ for $\ell = 1, \dots, M$ and K be their tensor product kernel on $\Omega = \times_{\ell=1}^M \Omega_\ell$ and $X = X^1 \times ... \times X^M \subset \Omega$ be a grid-like point set. Then its power function can be written as*

$$P_{K,X}(x)^2 = K(x,x) - \prod_{\ell=1}^{M} \left(K_\ell\left(p_\ell(x), p_\ell(x)\right) - P_{K_\ell,X^\ell}\left(p_\ell(x)\right)^2\right) \qquad \textit{for all } x \in \Omega,$$

where p_ℓ denotes the projection form Ω to Ω_ℓ for all $\ell = 1, ..., M$.

Proof. Let $\mathcal{N}$ denote the Newton basis of $S_{K,X}$ and $\mathcal{N}_\ell$ denote the component Newton bases of S_{K_ℓ,X^ℓ} for $\ell = 1, ..., M$. Due to the orthonormality, we have

$$\begin{aligned} P_{K,X}(x)^2 &= K(x,x) - \sum_{\mathfrak{n} \in \mathcal{N}} \mathfrak{n}(x)^2 \qquad \text{and} \\ P_{K_\ell,X^\ell}\left(p_\ell(x)\right)^2 &= K_\ell\left(p_\ell(x), p_\ell(x)\right) - \sum_{\mathfrak{n}_\ell \in \mathcal{N}_\ell} \mathfrak{n}_\ell\left(p_\ell(x)\right)^2 \end{aligned}$$

for all $x \in \Omega$ and $\ell = 1, ..., M$. Combining these results with (7.6), we obtain

$$\begin{aligned} P_{K,X}(x)^2 &= K(x,x) - \sum_{\mathfrak{n} \in \mathcal{N}} \mathfrak{n}(x)^2 \\ &= K(x,x) - \sum_{\mathfrak{n}_1 \in \mathcal{N}} \cdots \sum_{\mathfrak{n}_M \in \mathcal{N}_M} \prod_{\ell=1}^{M} \mathfrak{n}_\ell\left(p_\ell(x)\right)^2 \\ &= K(x,x) - \prod_{\ell=1}^{M} \left(\sum_{\mathfrak{n}_\ell \in \mathcal{N}_\ell} \mathfrak{n}_\ell\left(p_\ell(x)\right)^2 \right) \\ &= K(x,x) - \prod_{\ell=1}^{M} \left(K_\ell\left(p_\ell(x), p_\ell(x)\right) - P_{K_\ell,X^\ell}\left(p_\ell(x)\right)^2\right) \quad \text{for all } x \in \Omega. \end{aligned}$$

∎

With the previous results, we can show that the interpolation method converges if the power functions of the component kernels decay to zero.

Theorem 7.23. *Let* K_ℓ *be positive definite kernels on* Ω_ℓ *for* $\ell = 1, \ldots, M$ *and* K *be their tensor product kernel on* $\Omega = \times_{\ell=1}^M \Omega_\ell$*. If* $(X_n)_{n\in\mathbb{N}}$ *is a sequence of grid-like subsets in* Ω*, i.e.,*

$$X_n = X_n^1 \times \ldots \times X_n^M \qquad \text{where } X_n^\ell \subset \Omega_\ell \text{ for } \ell = 1, \ldots, M,$$

that satisfies the condition

$$P_{K_\ell, X_n^\ell}\left(p_\ell(x)\right) \xrightarrow{n\to\infty} 0 \qquad \text{for } x \in \Omega,\ \ell = 1, \ldots, M, \tag{7.8}$$

then the interpolant converges

$$\left\| f - s_{f,K,X_n} \right\|_K \xrightarrow{n\to\infty} 0 \qquad \text{for all } f \in \mathcal{H}_{K,\Omega}.$$

Proof. Let the targe function be given by the linear combination

$$f = \sum_{i=1}^N c_i K(\cdot, y_i) \in S_{K,\Omega}.$$

Due to the linearity of the interpolation operator, we obtain

$$\left\| f - s_{f,K,X_n} \right\|_K \leqslant \sum_{i=1}^N |c_i| \left\| K(\cdot, y_i) - s_{K(\cdot,y_i),K,X_n} \right\|_K = \sum_{i=1}^N |c_i|\ P_{K,X_n}(y_i).$$

The power function's representation of Lemma 7.22 then implies

$$\left\| f - s_{f,K,X_n} \right\|_K \leqslant \sum_{i=1}^N |c_i| \left(K(y_i, y_i) - \prod_{\ell=1}^M \left(K_\ell\left(p_\ell(y_i), p_\ell(y_i)\right) - P_{K_\ell, X_n^\ell}\left(p_\ell(y_i)\right)^2 \right) \right).$$

The right-hand side converges to zero for $n \longrightarrow \infty$ by (7.8) and the definition of K. This proves the desired result for all functions $f \in S_{K,\Omega}$. Since $S_{K,\Omega}$ is dense in $\mathcal{H}_{K,\Omega}$, the convergence extends from target functions in $S_{K,\Omega}$ to target functions in $\mathcal{H}_{K,\Omega}$. ∎

Remark 7.24. Let $(X_n)_n$ be a sequence of grid like structured point sets, which satisfies

$$P_{K_\ell, X_n^\ell}(x^j) \xrightarrow{n\to\infty} 0 \qquad \text{for all } \ell = 1, \ldots, M,\ l \neq j \text{ and}$$
$$X^j = X_n^j \qquad \text{for all } n \in \mathbb{N}.$$

Then

$$P_{K,X_n}(x)^2 \xrightarrow{n\to\infty} P_{K_j,X^j}\left(p_j(x)\right)^2 \prod_{j\neq i} K_j\left(p_j(x), p_j(x)\right).$$

This shows that the convergence of every but one of the components' power functions $P_{K_j,X_n^j}(p_j(x))$ is not only sufficient but also necessary for the convergence of the power function $P_{K,X_n}(x)$. We remark that the power function can be bounded by the fill distance $h_{X_n,\Omega}$ by Theorem 2.43, and conclude that we cannot expect the error to converge to zero in a setting, where $(h_{X_n^j,\Omega_j})_{n\in\mathbb{N}}$ does not converge to 0.

For a detailed discussion on the relationship between the power function and convergence, we refer to [AI24].

7.3.2 Positive Definiteness

Recall from Section 7.2 that the Hadamard product is not suitable to show positive definiteness of a tensor product kernel with positive definite components. De facto, these kernels are positive definite. We demonstrate this in Theorem 7.25, and hence generalize Theorem 7.3, by omitting the assumption of the kernel's translation-invariance. To this end, we use the concepts of grid-like structure and Kronecker product, which have already led to intriguing insights on structure and efficiency earlier in this chapter.

Theorem 7.25. *Let K_ℓ be positive definite kernels on Ω_ℓ for $\ell = 1, \dots, M$. Then their corresponding tensor product kernel K is positive definite on $\Omega = \times_{\ell=1}^{M} \Omega_\ell$.*

Proof. Let $X = \{x_1, \dots, x_n\} \subset \Omega$ be a finite and pairwise distinct point set. Let Y^ℓ be the projection of X onto Ω_ℓ, such that any point that occurs multiple times in $p_\ell(X)$ appears only once in Y^ℓ. These sets give rise to a grid-like point set

$$Y = \bigtimes_{\ell=1}^{M} Y^\ell,$$

such that $X \subseteq Y$ and $Y^\ell \subset \Omega_\ell$ is pairwise distinct for all $\ell = 1, \dots, M$. See Fig. 7.3 for a visualization. Due to Corollary 7.16, the interpolation matrix $\mathbf{A}_{K,Y}$ is positive definite. Since $\mathbf{A}_{K,X}$ is a submatrix of $\mathbf{A}_{K,Y}$, it is positive definite as well. ■

This is a crucial result, as we are now able to use tensor product kernels for multivariate scattered data reconstruction methods. From previous results, we can further conclude the following norm relation.

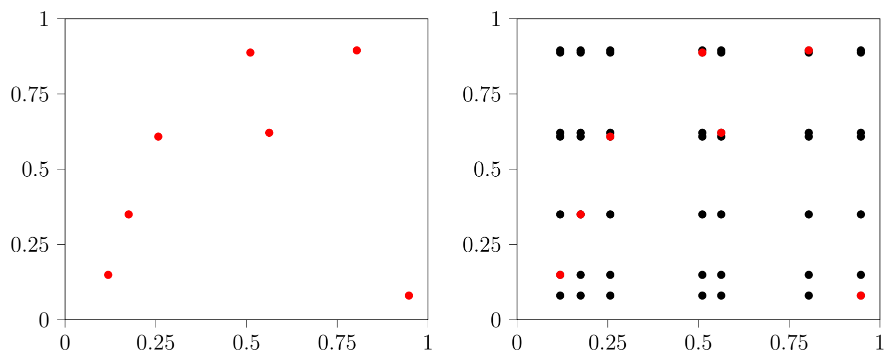

Figure 7.3: Visualization of a scattered point set (left) and the corresponding grid-like structured superset (right).

Lemma 7.26. *Let K_ℓ be positive definite kernels on Ω_ℓ for $\ell = 1, \ldots, M$, K be their tensor product kernel on $\Omega = \times_{\ell=1}^{M} \Omega_\ell$, and $X \subset \Omega$ be pairwise distinct. Furthermore, let the target function be of the tensor form*

$$f = \prod_{\ell=1}^{M} f_\ell \circ p_\ell$$

with f_ℓ acting on Ω_ℓ and p_ℓ being the projection from Ω to Ω_ℓ. Let $X \subset \Omega$ have pairwise distinct projections $p_\ell(X) \subset \Omega_\ell$ and the function values $f_{\ell_{p_\ell(X)}}$ be known for all $\ell = 1, \ldots, M$.

Then,

$$\left\| s_{f,K,X} \right\|_{K,\Omega} \leqslant \prod_{\ell=1}^{M} \left\| s_{f_\ell,K_\ell,p_\ell(X)} \right\|_{K_\ell,\Omega_\ell}.$$

Proof. Let us view the tensor product kernel K as the product kernel of transformed component kernels as in (7.1), i.e.,

$$K(x,y) = \prod_{\ell=1}^{M} K_{\ell_{p_\ell}}(x,y) \qquad \text{for all} x, y \in \Omega.$$

Applying Lemma 4.6 of the product kernel's chapter yields

$$\left\| s_{f,K,X} \right\|_{K,\Omega} \leqslant \prod_{\ell=1}^{M} \left\| s_{f_\ell \circ p_\ell, K_{\ell_{p_\ell}}, X} \right\|_{K_{\ell_{p_\ell}},\Omega}.$$

By Theorem 5.6 and Lemma 5.8 of the transformation kernel's chapter, we further deduce

$$\left\| s_{f,K,X} \right\|_{K,\Omega} \leqslant \prod_{\ell=1}^{M} \left\| s_{f_\ell,K_\ell,p_\ell(X)} \right\|_{K_\ell,\Omega_\ell}.$$

■

7.3.3 Numerical Stability

The main indicator for the stability of kernel-based reconstruction methods is the spectral condition number of the interpolation matrix $\mathbf{A}_{K,X}$, which is given by

$$\operatorname{cond}_2(\mathbf{A}_{K,X}) = \frac{\lambda_{\max}(\mathbf{A}_{K,X})}{\lambda_{\min}(\mathbf{A}_{K,X})},$$

where $\lambda_{\max}(\mathbf{A}_{K,X})$ and $\lambda_{\min}(\mathbf{A}_{K,X})$ denote the largest and smallest (positive) eigenvalue of $\mathbf{A}_{K,X}$, see Section 2.4.2.

In accordance with the previous sections, we will demonstrate that the condition number corresponding to the tensor product kernel can be bounded by the product of the condition numbers of its components. The following results are based on the product property of the eigenvalues of the Kronecker product of Lemma 7.14 *(iv)*.

Theorem 7.27. *Let K_ℓ be positive definite kernels on Ω_ℓ for $\ell = 1, \dots, M$ and K be their tensor product kernel on $\Omega = \times_{\ell=1}^M \Omega_\ell$. Furthermore, let $X \subset \times_{\ell=1}^M \Omega_\ell$ be a pairwise distinct point set and $Y^1 \times \cdots \times Y^M = Y \in \times_{\ell=1}^M \Omega_\ell$ its grid-like structured superset. Then, the following statements hold:*

(i)
$$\mathrm{cond}_2(\mathbf{A}_{K,Y}) = \prod_{\ell=1}^M \mathrm{cond}_2(\mathbf{A}_{K_\ell,Y^\ell}).$$

(ii)
$$\mathrm{cond}_2(\mathbf{A}_{K,X}) \leqslant \mathrm{cond}_2(\mathbf{A}_{K,Y}).$$

(iii)
$$\lambda_{\min}(\mathbf{A}_{K,X}) \geqslant \prod_{\ell=1}^M \lambda_{\min}(\mathbf{A}_{K_\ell,Y^\ell}).$$

(iv) Let K_ℓ be translation-invariant kernels and their univariate functions Φ_ℓ satisfy $\widehat{\Phi_\ell} \in C(\mathbb{R}^{d_\ell} \backslash \{0\})$. Then,

$$\lambda_{min}(\mathbf{A}_{K,X}) \geqslant \prod_{\ell=1}^M G_{\Phi_\ell}(q_{Y^\ell}),$$

where the functions G_{Φ_ℓ} come from Remark 2.49.

(v) Let K be a translation-invariant kernel and its univariate functions Φ satisfy $\widehat{\Phi} \in C(\mathbb{R}^{d_\ell} \backslash \{0\})$. Then

$$\lambda_{min}(\mathbf{A}_{K,X}) \geqslant \lambda_{min}(\mathbf{A}_{K,Y}) \geqslant G_\Phi\left(\min_\ell \{q_{Y^\ell}\}\right),$$

where the functions G_Φ comes from Remark 2.49.

Proof. According to Lemma 7.14 *(iv)* and Theorem 7.15, the spectrum $\lambda(\mathbf{A}_{K,Y})$ of $\mathbf{A}_{K,Y}$ can be written as

$$\lambda(\mathbf{A}_{K,Y}) = \left\{ \prod_{\ell=1}^M \lambda^\ell \ : \ \lambda^\ell \in \lambda(\mathbf{A}_{K_\ell,Y^\ell}) \text{ for } \ell = 1, ..., M \right\}. \tag{7.9}$$

This directly implies (i), since

$$\mathrm{cond}_2(\mathbf{A}_{K,Y}) = \frac{\lambda_{\max}(\mathbf{A}_{K,Y})}{\lambda_{\min}(\mathbf{A}_{K,Y})} = \frac{\prod_{\ell=1}^M \lambda_{\max}(\mathbf{A}_{K_\ell,Y^\ell})}{\prod_{\ell=1}^M \lambda_{\min}(\mathbf{A}_{K_\ell,Y^\ell})} = \prod_{\ell=1}^M \mathrm{cond}_2(\mathbf{A}_{K_\ell,Y^\ell}).$$

We combine Cauchy's interlacing theorem, given in Lemma 4.7, with the fact that $\mathbf{A}_{K,X}$ is a principal submatrix of $\mathbf{A}_{K,Y}$ to get

$$\lambda_{\min}(\mathbf{A}_{K,Y}) \leqslant \lambda_{\min}(\mathbf{A}_{K,X}) \text{ and } \lambda_{\max}(\mathbf{A}_{K,X}) \leqslant \lambda_{\max}(\mathbf{A}_{K,Y}). \tag{7.10}$$

The statement of (ii) follows, as

$$\mathrm{cond}_2(\mathbf{A}_{K,X}) = \frac{\lambda_{\max}(\mathbf{A}_{K,X})}{\lambda_{\min}(\mathbf{A}_{K,X})} \leqslant \frac{\lambda_{\max}(\mathbf{A}_{K,Y})}{\lambda_{\min}(\mathbf{A}_{K,Y})} = \mathrm{cond}_2(\mathbf{A}_{K,Y}).$$

A combination of (7.9) and (7.10) implies (iii). The statement (iv) follows by (iii) and the additional assumption, as

$$\lambda_{\min}(\mathbf{A}_{K,X}) = \lambda_{\min}(\mathbf{A}_{\Phi,X}) \geqslant \prod_{\ell=1}^{M} \lambda_{\min}(\mathbf{A}_{\Phi_\ell,Y^\ell}) \geqslant \prod_{\ell=1}^{M} G_\ell(q_{Y^\ell}).$$

The first inequality of (v) is a consequence of (7.10). By the assumptions and Theorem 2.48, it is

$$\lambda_{\min}(\mathbf{A}_{K,Y}) \geqslant G(q_Y) \geqslant G\left(\min_{\ell}\{q_{Y^\ell}\}\right).$$

■

We immediately conclude that, for the sake of stability, the component kernels K_ℓ should be chosen, such that they generate sufficiently small condition numbers on the corresponding data sets. A supporting numerical result is provided in Section 7.4.1.

For the kernels used in this thesis, estimates on the eigenvalues of their interpolation matrices can be found in Section 2.4.2. These can be combined with Theorem 7.27 (iv) to derive condition number estimates for tensor product kernels.

7.4 Numerical Tests

In the following, we discuss how a tensor product kernel can be adapted to the underlying problem and the advantages of its use for interpolation problems. First, in Section 7.4.1, we provide an example of how adapting the tensor product kernel to the point set can improve stability and error. Then, in Section 7.4.2, we demonstrate how and when its tensor structure can be used to accelerate the computation of the interpolant depending on the Newton basis.

7.4.1 Error vs. Stability

The following numerical example illustrates how the tensor product kernel can be used to achieve a stable approximation method with high accuracy. We use the setting described in Section 5.4.2, specifically the domain $\Omega = \Omega_1 \times \Omega_2 = [0, 0.5] \times [0, 2]$, the point sets $\mathbf{X_n} = X_n^1 \times X^2$ for $n = 2, \ldots, 8$, where X_n^1 is given by $2^n + 1$ equally distributed points in Ω_1 and X^2 stays unchanged for developing n, and the target function $f_{\text{kink-t}}$ (defined in Section 5.4.2 and visualized in Fig. 5.8). Recall that the kink of this target makes it a difficult function to approximate for most standard kernels. This choice is intentional to demonstrate the enhancement of problem-adapted kernel methods, such as the transformation kernel and the tensor product kernel, even with such a problematic target function. For other, smoother target functions, we can expect similar improvements. Since we are working within the same setting as in Section 5.4.2, we subsequently incorporate the results obtained there.

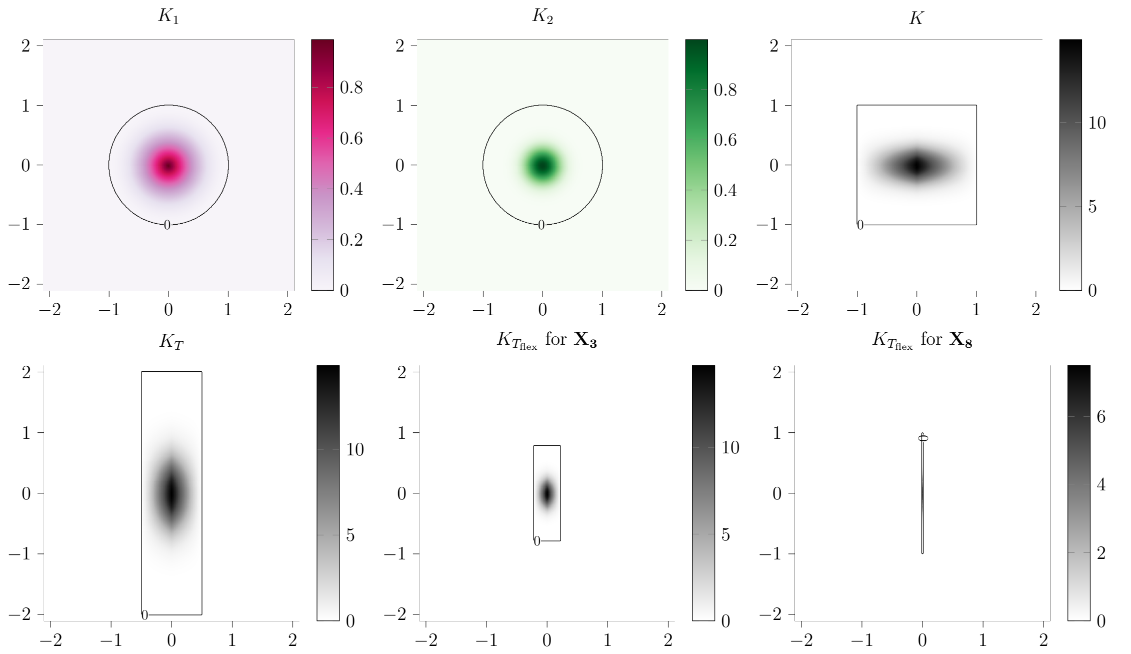

Figure 7.4: Top view of two-dimensional kernels and their support (black line). Top row: Standard kernels K_1 (left) and K_2 (middle), and the tensor product kernel K (right). Bottom row: The transformed versions of K, namely K_T (left), $K_{T_{\text{flex}}}$ for $\mathbf{X_3}$ (middle), and $K_{T_{\text{flex}}}$ for $\mathbf{X_8}$ (right).

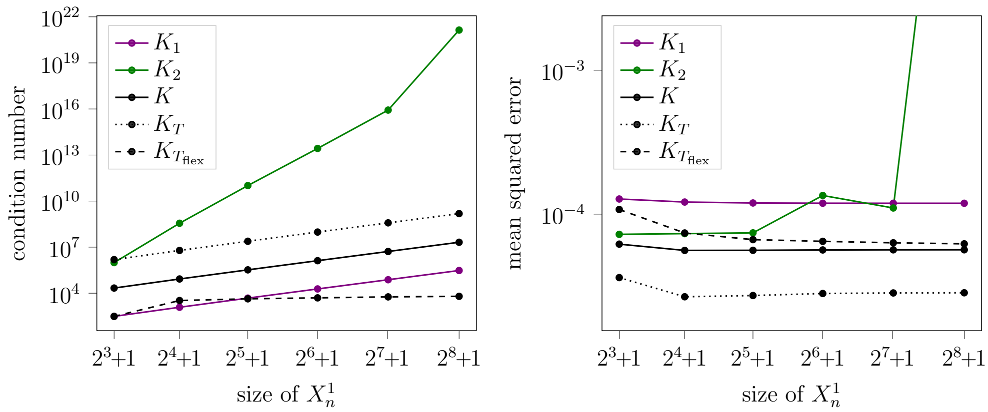

Figure 7.5: Comparison of the standard kernels K_1 and K_2, the tensor product kernel K, and its transformed versions K_T adapted to the domain Ω and $K_{T_{\text{flex}}}$ flexibly adapted to the point sets $\mathbf{X_n} = X_n^1 \times X^2$ for $n = 3, \ldots, 8$ regarding the condition number (left) and mean squared error (right) regarding.

From previous results, we know that the Wendland kernel of the RBF $\phi_{d,0}$ exhibits good stability but poor error performance, whereas the Wendland kernel of $\phi_{d,3}$ has the opposite characteristics. Additionally, we understand that closely spaced points lead to poor numerical condition number but good approximation error, while widely spaced points result in the opposite properties for the interpolant. This analysis leads us to the following choice of the tensor product kernels components in the setting of Section 5.4.2.

1. In the Ω_1-direction, with closely spaced points X_n^1, we use the Wendland kernel κ_1 of the RBF $\phi_{1,0}$. Due to the proximity of the points in X_n^1, we can expect a reasonably small error, while the choice of the kernel should maintain a low condition number despite the narrow point spacing.

2. In the Ω_2-direction, with comparatively wide spaced points X^2, we use the Wendland kernel κ_2 of the RBF $\phi_{1,3}$. This choice is intended to control the error despite wide point spacing of X^2, while the condition number should not rise excessively due to the points distance.

In the following, we compare the tensor product kernel K with its components κ_1 and κ_2, which we convert into two-dimensional kernels for comparison purposes, namely K_1 and K_2 resulting from the Wendland RBFs $\phi_{3,0}$ and $\phi_{3,3}$. We want to emphasize the possibility to combine previous adaptation methods. To this end, we apply the transformations T (domain adapted) and T_{flex} (point set adapted) developed in Section 5.4.2 to the tensor product kernel K and include these transformed tensor product kernels K_T and $K_{T_{\text{flex}}}$ to our comparison. Fig. 7.4 visualizes the different kernels.

First, we look at the relationship between K and the comparison kernels K_1 and K_2. In regard to the spectral condition number visualized in Fig. 7.5 (left), we can observe the following. As expected from Theorem 7.27, the curve for K runs parallel to that of K_1. Specifically, this theorem provides $\operatorname{cond}_2(\mathbf{A}_{K,X}) = \operatorname{cond}_2(\mathbf{A}_{K_1,X^1}) \operatorname{cond}_2(\mathbf{A}_{K_2,X^2})$. The fact that $\operatorname{cond}_2(\mathbf{A}_{K_2,X^2})$ remains constant as X^2 does not alter, explains the numerical result.

In Fig. 7.5 (right), we see that the tensor product kernel K yields a better approximation than the comparison kernels K_1 and K_2. This is also due to the fact that the poor condition number of K_2 adversely affects its corresponding error. Additionally, it is observed that the error for all other kernels stabilizes for X_n^1, $n \geqslant 5$. This confirms the expectation from Remark 7.24, which states that the interpolant does not converge to the target function if a component of the grid-like sets $\mathbf{X_n}$ has a fill distance that does not approach zero, which is the case for X^2 in this example.

Let us now turn to the transformed kernels. As expected from Section 5.4.2, the transformations T and T_{flex} have the same effect on the tensor product kernel K as on K_1 and K_2 (cf. Fig. 5.10). The transformation T improves the error while the condition number shifts upwards by a factor, and T_{flex} improves the condition number while maintaining comparable approximation quality.

We summarize the following observations:

1. A tensor product kernel improves the error while maintaining a controllable condition number.

2. The combinations of anisotropic kernels presented in this thesis, namely the transformation and tensor product kernels, yield the best results in this example. The condition number of K_T grows only as fast as that of K_1, and its approximation error even drops below the one of K. Meanwhile, $K_{T_{\text{flex}}}$ keeps the condition number nearly constant while achieving an error comparable to K.

The tensor product kernel significantly enhances interpolation methods in a grid-like setting. If further improvement is desired and more effort can be invested, a combination of transformation and tensor product kernels serves as the final touch for improved performance.

7.4.2 Efficiency in Time

In the following, we demonstrate that interpolation methods leveraging the tensor structure of the underlying problem are more efficient, meaning that the time needed to compute an interpolant on the underlying domain can be reduced.

The setup is as follows:

- The dimension d and domain $\Omega = [0,1]^d = \times_{l=1}^{d}[0,1]$.
- The point sets

$$\mathbf{X}_\mathbf{n}^\mathbf{d} = \bigtimes_{\ell=1}^{d} X_n \subset \Omega, \qquad \text{where } X_n = \left\{ \frac{i}{2^n} \ : \ i = 0, 1, \ldots, 2^n \right\} \subset [0,1],$$

  resulting in the sizes $|X_n| = 2^n + 1$ and $|\mathbf{X}_\mathbf{n}^\mathbf{d}| = (2^n+1)^d$.
- The kernel K acting on Ω is the tensor product kernel of d times κ, where κ is the kernel of the Wendland RBF $\phi_{1,0}$, given in Theorem 2.34, acting on $[0,1]$.
- The target function has no effect on the results, as execution times do not vary for different function values $f_{\mathbf{X}_\mathbf{n}^\mathbf{d}}$.

In this setting we compare four interpolation methods, two regular methods (`standard` and `newton`) and its corresponding tensor structure exploiting methods (`std.tensor` and `new.tensor`):

- `standard`:
  1. compute interpolation matrix $\mathbf{A}_{K,\mathbf{X}_\mathbf{n}^\mathbf{d}}$
  2. solve $\mathbf{A}_{K,\mathbf{X}_\mathbf{n}^\mathbf{d}} c = f_{\mathbf{X}_\mathbf{n}^\mathbf{d}}$ for c
  3. compute the interpolant on an Ω-grid by $V_{\mathcal{B}_{K,\mathbf{X}_\mathbf{n}^\mathbf{d}},\Omega\text{-grid}}\, c$.

- `std.tensor`: exploiting the tensor structure of $\mathbf{A}_{K,\mathbf{X_n^d}}$ (Theorem 7.15)
  1. compute components interpolation matrices $\mathbf{A}_{\kappa,X_n}$
  2. compute interpolation matrix $\mathbf{A}_{K,\mathbf{X_n^d}} = \bigotimes_{\ell=1}^{d} \mathbf{A}_{\kappa,X_n}$
  3. solve $\mathbf{A}_{K,\mathbf{X_n^d}} c = f_{\mathbf{X_n^d}}$ for c
  4. evaluate the interpolant on an Ω-grid by $\left(\bigotimes_{\ell=1}^{d} V_{\mathcal{B}_{\kappa,X_n},[0,1]\text{-grid}}\right) c$
- `newton`:
  1. compute Choleski factor L of $\mathbf{A}_{K,\mathbf{X_n^d}}$
  2. solve $Lc = f_{\mathbf{X_n^d}}$ for c
  3. compute interpolant on an Ω-grid by $V_{\mathcal{N},\Omega\text{-grid}}\, c$
- `new.tensor`: exploiting tensor structure of the Newton basis $\mathcal{N}$ of K (Section 7.3.1)
  1. compute component Choleski factors L_ℓ of $\mathbf{A}_{\kappa,X_n}$ and compute the Choleski factor $L = \bigotimes_{\ell=1}^{d} L_\ell$ of $\mathbf{A}_{K,\mathbf{X_n^d}}$
  2. compute Vandermonde matrix $V_{\mathcal{N},\Omega\text{-grid}} = \left(\bigotimes_{\ell=1}^{d} V_{\kappa,[0,1]\text{-grid}}\right) \left(\bigotimes_{\ell=1}^{d} (L_\ell^T)^{-1}\right)$
  3. solve $Lc = f_{\mathbf{X_n^d}}$ for c
  4. evaluate interpolant on an Ω-grid by $V_{\mathcal{N},\Omega\text{-grid}}\, c$

At this point, we would like to note that significant time savings are achieved when, as in our example, there are multiple directions where points and kernels are identical. In our example, $X_\ell = X_n$ for all $\ell = 1, \dots, d$, regardless of ℓ, and the kernel used is the same for every direction, namely κ. This means that we need to compute the interpolation matrix $\mathbf{A}_{\kappa,X_n}$ or the Cholesky factor L_ℓ only once, rather than multiple times. However, this comparison does not address the flexibility of tensor product kernels and is therefore not discussed here.

To demonstrate the efficiency of the proposed tensor methods in Section 7.3.1, we analyze the execution time as the number of data points and the dimensionality increase. Since execution times can vary significantly, particularly for shorter durations, we run each method ten times, and take the average execution time.

In Fig. 7.6 (left), we present execution times for the developing point sets $\mathbf{X_n^2}$, with $n = 1, \dots, 6$, in a two-dimensional setting, $d = 2$, resulting in the sizes

$$\left\{|\mathbf{X_n^2}|\right\}_n = \left\{|X_n|^2\right\}_n = \{9, 25, 81, 289, 1089, 4225\}.$$

We observe that the `newton` and `standard` method can be accelerated to `new.tensor` and `std.tensor`.

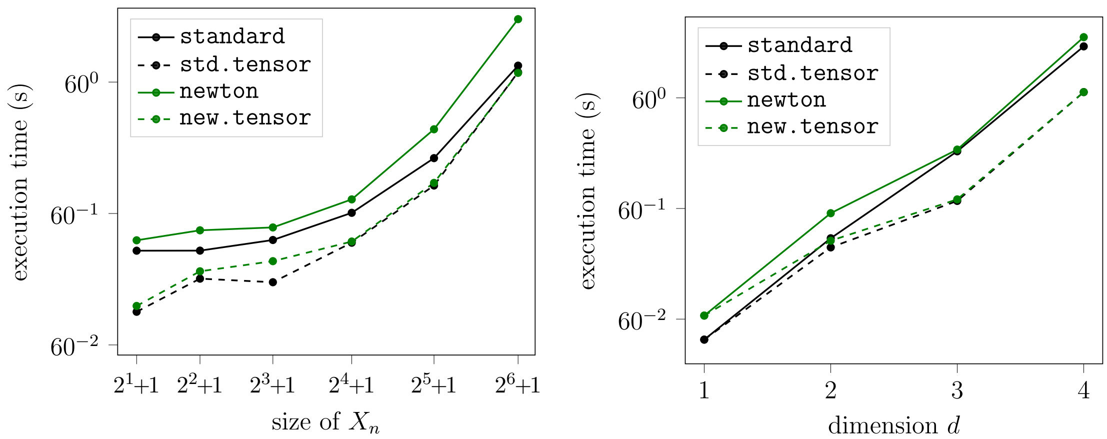

Figure 7.6: Comparison of the interpolation methods `standard` and `newton`, and their tensor structure exploiting versions `std.tensor` and `new.tensor` regarding execution times in seconds corresponding to increasing sizes of the component point sets X_n for dimension $d = 2$ (left), and increasing dimensions d for the component point set X_2 (right).

In Fig. 7.6 (right), we consider execution times for increasing dimensions and X_2. Namely, we consider the sets $\mathbf{X_2^d} = \times_{l=1}^{d} X_2$ for dimensions $d = 1, 2, 3, 4$, with sizes $5, 25, 125, 625$. For $d = 1$, i.e., $\mathbf{X_2^1}$, the approaches `standard` and `newton` do not differ from the corresponding tensor structure exploiting methods. However, as we examine higher dimensions, the `std.tensor` and `new.tensor` methods prove to be more efficient compared to the original `standard` and `newton` methods.

We conclude that tensor product kernels enable a speedup of interpolation methods in grid-like settings.

Part IV

Final Remarks

Conclusion and Outlook

This thesis set out to enhance the toolbox for adapting kernels to underlying problems, focusing on the interpolation of multivariate scattered data with an emphasis on anisotropies. By developing five novel classes of flexible kernels – summation, product, transformation, orthogonal summation, and tensor product kernels – we address significant limitations of traditional radial kernels. These classes, some entirely new and others building on existing structures, provide the capability to select and combine kernels tailored to the underlying problem, thereby extending the variety of interpolation methods. Our theoretical analysis expands the understanding of native spaces and their impact on the corresponding interpolation method. Key contributions include connecting N. Aronszajn's results to Sobolev spaces and J. Mercer's Theorem to infinite sums of kernels, as well as exploring kernel equivalence classes. Furthermore, we provide a structural analysis of the native spaces corresponding to each of the five novel kernel classes, as well as an examination of their impacts on interpolation accuracy and stability. Numerical tests confirm our theoretical findings and identify which classes of kernels are suitable for specific problem adaptations, resulting in the following suggestions:

- For an adaptation to the point set X, we recommend using transformation kernels or tensor product kernels.
- For an adaptation to the domain Ω, we recommend using transformation kernels.
- For an adaptation to the properties of the target function f, we recommend using summation kernels, transformation kernels, or orthogonal summation kernels.

As shown in Section 7.4.1, it is possible to combine different classes of kernels, extending the variety of interpolation methods even further. Fig. 7.7 demonstrates this extensive variety.

The results and methodologies presented in this thesis open several avenues for future research. The flexibility and adaptability of these kernels make them suitable for integration into other kernel methods beyond interpolation. Techniques such as Hermite-Birkhoff interpolation, kernel regression, and support vector machines could benefit from the enhanced performance and tailored adaptations offered by the developed kernels. Another exciting prospect is the exploration of greedy methods in combination with our developed kernels. Greedy algorithms, which iteratively select the most promising candidate for an additional interpolation point, could further improve the efficiency and effectiveness of our interpolation methods. Investigating the interplay between greedy methods and the proposed kernel classes could yield valuable insights and performance gains. Moreover, the anisotropic behavior of the presented kernel classes makes them promising for real-world applications. For instance, in geospatial data analysis or medical imaging, where data is often scattered and of anisotropic nature, these kernels can significantly enhance accuracy and stability.

In summary, this thesis has laid a robust foundation for the continued evolution of kernel-based methods, offering both theoretical advancements and practical tools for future research and application.

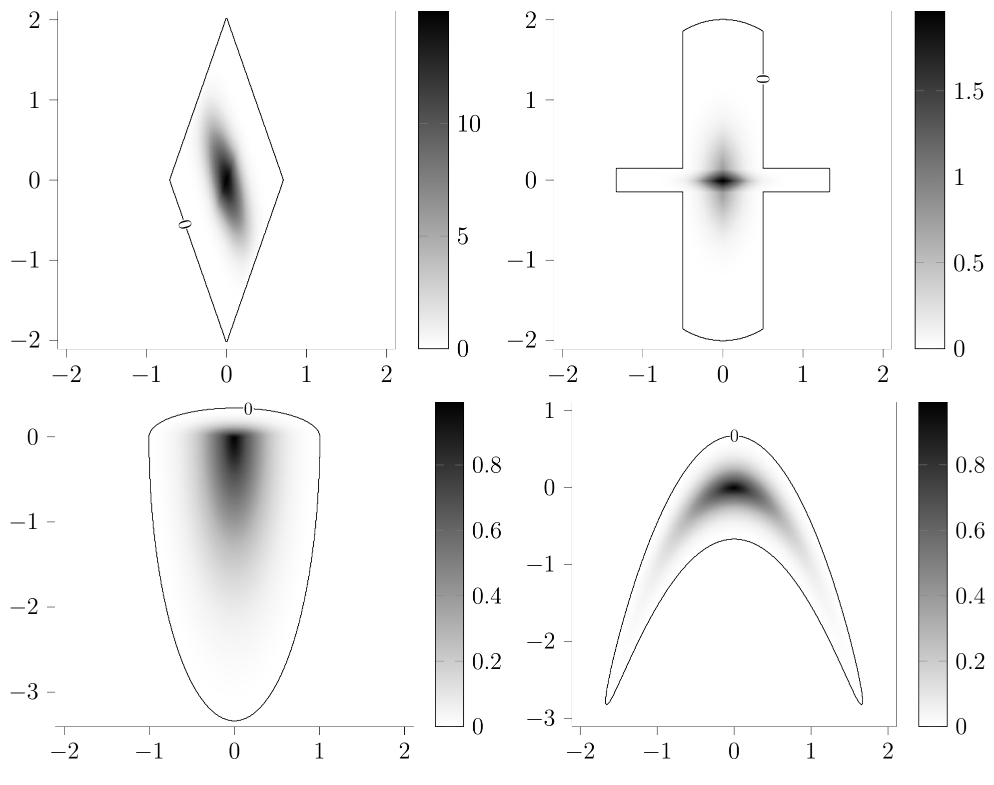

Figure 7.7: Top view of two-dimensional kernels and their support (black line). Top row: Transformed tensor product kernel (left) and product of an orthogonal summation kernel and a transformation kernel (right). Bottom row: Transformation kernels.

List of Figures

List of Symbols

$\times$	Cartesian product, Definition 2.5
$V_{\mathcal{B},X}$	Vandermonde matrix, Definition 2.1
$\mathbf{A}_{K,X}$	Interpolation matrix, Eq. (2.5)
$s,\ s_{f,K,X}$	Interpolant, Eq. (2.7)
$S_{K,X}$	Interpolation space, Eq. (2.8)
$\mathcal{H}_{K,\Omega},\ \langle\cdot,\cdot\rangle_K$	Native space with inner product, Theorem 2.24
$P_{K,X}$	Power function, Definition 2.27
$\lambda_{\min},\ \lambda_{\max}$	Smallest and largest eigenvalue of a matrix
$W^m(\mathbb{R}^d)$	Sobolev space, Definition 2.31
ψ_l	Matérn's radial basis functions, Tab. 2.1
$\phi_{d,k}$	Wendland's radial basis functions, Tab. 2.2
$h_{X,\Omega}$	Fill distance, Definition 2.41
$F_{\Phi,\Omega}$	Function for error bound, Remark 2.44, Tab. 2.3
q_X	Separation distance, Definition 2.47
G_Φ	Function for stability bound, Remark 2.49, Tab. 2.4
$\oplus, \perp$	Orthogonal direct sum, orthogonal complement Definition 3.7
$\ll$	Partial order of reproducing kernels, Definition 3.14
$\lesssim$	Preorder of Hilbert spaces and reproducing kernels, Definition 3.19
$\sim$	Equivalence relation corresponding to $\lesssim$, Remark 3.21
$\odot$	Hadamard product, Definition 4.2
$\otimes$	Hilbert tensor product of Hilbert spaces, Definition 7.8
	or Kronecker product of matrices, Definition 7.13

Bibliography

[AD14] Aiolli, F. ; Donini, M.: Learning anisotropic RBF kernels. In: *International Conference on Artificial Neural Networks* Springer, 2014

[AEI23a] Albrecht, K. ; Entzian, J. ; Iske, A.: Anisotropic kernels for particle flow simulation. In: *Modeling, Simulation and Optimization of Fluid Dynamic Applications* Springer, 2023

[AEI23b] Albrecht, K. ; Entzian, J. ; Iske, A.: Product kernels are efficient and flexible tools for high-dimensional scattered interpolation. In: *arXiv preprint arXiv:2312.09949* (2023)

[AI24] Albrecht, K. ; Iske, A.: On the convergence of generalized kernel-based interpolation by greedy data selection algorithms. In: *arXiv preprint arXiv:2407.03840* (2024)

[Alb24] Albrecht, K.: *Kernel-based generalized interpolation and its application to computerized tomography*, Universität Hamburg, Diss., 2024

[Aro50] Aronszajn, N.: Theory of reproducing kernels. In: *Transactions of the American Mathematical Society* 68 (1950)

[Ask73] Askey, R.: Radial characteristic functions. Technical Reprt Nr. 1262. University of Wisconsin. (1973)

[Axl15] Axler, S.: *Linear algebra done right.* Springer, 2015

[BDL10] Beatson, Rick ; Davydov, Oleg ; Levesley, Jeremy: Error bounds for anisotropic RBF interpolation. In: *Journal of Approximation Theory* 162 (2010)

[BLRS15] Bozzini, M. ; Lenarduzzi, L. ; Rossini, M. ; Schaback, R.: Interpolation with variably scaled kernels. In: *IMA Journal of Numerical Analysis* 35 (2015)

[Boc32] Bochner, S.: *Vorlesungen über fouriersche Integrale.* Akademische Verlagsgesellschaft, 1932

[Boc33] Bochner, S.: Monotone funktionen, stieltjessche integrale und harmonische analyse. In: *Mathematische Annalen* 108 (1933)

[BTA11] Berlinet, A. ; Thomas-Agnan, C.: *Reproducing kernel Hilbert spaces in probability and statistics.* Springer Science & Business Media, 2011

[Buh03] Buhmann, M. D.: *Radial basis functions: Theory and implementations.* Cambridge University Press, 2003

[CJ59] Curtis Jr, P. C.: N-parameter families and best approximation. In: *Pacific Journal of Mathematics* (1959)

[CLMM06] Casciola, G. ; Lazzaro, D. ; Montefusco, L. B. ; Morigi, S.: Shape preserving surface reconstruction using locally anisotropic radial basis function interpolants. In: *Computers & Mathematics with Applications* 51 (2006)

[DMSW05] De Marchi, S. ; Schaback, R. ; Wendland, H.: Near-optimal data-independent point locations for radial basis function interpolation. In: *Advances in Computational Mathematics* 23 (2005)

[DNPV12] Di Nezza, E. ; Palatucci, G. ; Valdinoci, E.: Hitchhiker's guide to the fractional Sobolev spaces. In: *Bulletin des Sciences Mathématiques* 136 (2012)

[Dri73] Driscoll, M. F.: The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 26 (1973)

[Eva10] Evans, L. C.: *Partial differential equations.* 2nd Ed. American Mathematical Society, 2010

[Fas07] Fasshauer, G. E.: *Meshfree approximation methods with MATLAB.* World Scientific, 2007

[Fas11] Fasshauer, G. E.: Positive definite kernels: Past, present and future. In: *Dolomites Research Notes on Approximation* 4 (2011)

[FM15] Fasshauer, G. E. ; McCourt, M. J.: *Kernel-based approximation methods using Matlab.* World Scientific Publishing Company, 2015

[Fra79] Franke, R.: A critical comparison of some methods for interpolation of scattered data. Technical Report. Naval Postgraduate School. (1979)

[Fra82] Franke, R.: Scattered data interpolation: Tests of some methods. In: *Mathematics of Computation* 38 (1982)

[FY11] Fasshauer, G. E. ; Ye, Q.: Reproducing kernels of generalized Sobolev spaces via a Green function approach with distributional operators. In: *Numerische Mathematik* 119 (2011)

[GM16] Guella, J. C. ; Menegatto, V. A.: Strictly positive definite kernels on a product of spheres. In: *Journal of Mathematical Analysis and Applications* 435 (2016)

[Haa10] Haar, A.: Zur Theorie der orthogonalen Funktionensysteme. In: *Mathematische Annalen* 69 (1910)

[Hac12] Hackbusch, W.: *Tensor spaces and numerical tensor calculus.* Springer, 2012

[HJ91] Horn, R. A. ; Johnson, C. R.: *Topics in matrix analysis.* Cambridge University Press, 1991

[Isk18] Iske, A.: *Approximation theory and algorithms for data analysis.* Springer, 2018

[KC92] Kansa, E. J. ; Carlson, R. E.: Improved accuracy of multiquadric interpolation using variable shape parameters. In: *Computers & Mathematics with Applications* 24 (1992)

[Kol92] Koldobsky, Alexander: Schoenberg's problem on positive definite functions. In: *Algebra and Analysis* 3 (1992)

[Kol09] Koldobsky, A.: A note on positive definite norm dependent functions. In: *High Dimensional Probability V: The Luminy Volume* Bd. 5. Institute of Mathematical Statistics, 2009

[KR83] Kadison, R. V. ; Ringrose, J. R.: *Fundamentals of the theory of operator algebras I: Elementary Theory.* Academic Press, 1983

[LF05] Larsson, E. ; Fornberg, B.: Theoretical and computational aspects of multivariate interpolation with increasingly flat radial basis functions. In: *Computers & Mathematics with Applications* 49 (2005)

[LMZ+24] Li, G. ; Ma, S. ; Zhang, D. ; Yang, L. ; Zhang, W. ; Wu, Z.: An efficient sequential anisotropic RBF reliability analysis method with fast cross-validation and parallelizability. In: *Reliability Engineering & System Safety* 241 (2024)

[Mai56] Mairhuber, J. C.: On Haar's theorem concerning Chebychev approximation problems having unique solutions. In: *Proceedings of the American Mathematical Society* 7 (1956)

[Mat86] Matérn, B.: *Spatial variation. 2nd Ed. Lecture notes in statistics.* Springer, 1986

[McL00] McLean, W. C. H.: *Strongly elliptic systems and boundary integral equations.* Cambridge University Press, 2000

[Mer09] Mercer, J.: Functions of positive and negative type, and their connection the theory of integral equations. In: *Philosophical Transactions of the Royal Society A – Mathematical, Physical and Engineering Sciences* 209 (1909)

[Moo39] Moore, E. H.: *General analysis, Parts I and II.* Memoirs of the American Philosophical Society, 1935 and 1939

[MS09] Müller, S. ; Schaback, R.: A Newton basis for kernel spaces. In: *Journal of Approximation Theory* 161 (2009)

[Mül09] Müller, S.: *Komplexität und Stabilität von kernbasierten Rekonstruktionsmethoden*, Universität Göttingen, Diss., 2009

[Mus14] Muscat, J.: *Functional analysis: an introduction to metric spaces, Hilbert spaces, and Banach algebras.* Springer, 2014

[MVHÖ23] Mojarrad, F. N. ; Veiga, M. H. ; Hesthaven, J. S. ; Öffner, P.: A new variable shape parameter strategy for RBF approximation using neural networks. In: *Computers & Mathematics with Applications* 143 (2023)

[Nev71] Neveu, J.: *Processus aléatoires gaussiens - Séminaire de mathématiques supérieures.* Presses de l'Université de Montréal, 1971

[OS17] Owhadi, H. ; Scovel, C.: Separability of reproducing kernel spaces. In: *Proceedings of the American Mathematical Society* 145 (2017)

[PS11] Pazouki, M. ; Schaback, R.: Bases for kernel-based spaces. In: *Journal of Computational and Applied Mathematics* 236 (2011)

[RBP+06] Rayner, N. A. ; Brohan, P. ; Parker, D. E. ; Folland, C. K. ; Kennedy, J. J. ; Vanicek, M. ; Ansell, T. J. ; Tett, S. F. B.: Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. In: *Journal of Climate* 19 (2006)

[Rea99] Reams, R.: Hadamard inverses, square roots and products of almost semidefinite matrices. In: *Linear Algebra and its Applications* 288 (1999)

[Saa11] Saatçi, Y.: *Scalable inference for structured Gaussian process models*, Universitiy of Cambridge, Diss., 2011

[SC08] Steinwart, I. ; Christmann, A.: *Support vector machines.* Springer Science & Business Media, 2008

[Sch38] Schoenberg, I. J.: Metric spaces and completely monotone functions. In: *Annals of Mathematics* 39 (1938)

[Sch64] Schwartz, L.: Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). In: *Journal d'analyse Mathématique* 13 (1964)

[Sch95] Schaback, R.: Error estimates and condition numbers for radial basis function interpolation. In: *Advances in Computational Mathematics* 3 (1995)

[SH17] Santin, G. ; Haasdonk, B.: Convergence rate of the data-independent P-greedy algorithm in kernel-based approximation. In: *Dolomites Research Notes on Approximation* 10 (2017)

[SW00] Schaback, R. ; Wendland, H.: Adaptive greedy techniques for approximate solution of large RBF systems. In: *Numerical Algorithms* 24 (2000)

[UA22] Unser, M. ; Aziznejad, S.: Convex optimization in sums of Banach spaces. In: *Applied and Computational Harmonic Analysis* 56 (2022)

[Wen05] Wendland, H.: *Scattered data approximation.* Cambridge University Press, 2005

[Wer18] Werner, D.: *Funktionalanalysis.* Springer Spektrum, 2018

[WGNC14] Wilson, A. G. ; Gilboa, E. ; Nehorai, A. ; Cunningham, J. P.: Fast kernel learning for multidimensional pattern extrapolation. In: *Advances in Neural Information Processing Systems* 27 (2014)

[WMP24] Wenzel, T. ; Marchetti, F. ; Perracchione, E.: Data-driven kernel designs for optimized greedy schemes: A machine learning perspective. In: *SIAM Journal on Scientific Computing* 46 (2024)

[WSH23] Wenzel, T. ; Santin, G. ; Haasdonk, B.: Analysis of target data-dependent greedy kernel algorithms: Convergence rates for f-, $f \cdot P$ and f/P-greedy. In: *Constructive Approximation* 57 (2023)

[Ylv62] Ylvisaker, N. D.: On linear estimation for regression problems on time series. In: *The Annals of Mathematical Statistics* 33 (1962)